HAL open science

Quelques contributions à l'extraction, l'enrichissement des connaissances et à la prédiction de données spatio-temporelles massives

Nédra Mellouli-Nauwynck

▶ To cite this version:

Nédra Mellouli-Nauwynck. Quelques contributions à l'extraction, l'enrichissement des connaissances et à la prédiction de données spatio-temporelles massives. Apprentissage [cs.LG]. Université Paris 8, 2021. tel-03982917

HAL Id: tel-03982917

https://univ-paris8.hal.science/tel-03982917v1

Submitted on 10 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

École doctorale n° 224 : Cognition, Langage, Interaction

HABILITATION A DIRIGER DES RECHERCHES

Université Paris 8 Vincennes Saint-Denis

Spécialité : Informatique

présentée et soutenue publiquement par

Nédra Mellouli-Nauwynck

le 7 Janvier 2021

Quelques contributions à l'extraction, l'enrichissement des connaissances et à la prédiction de données spatio-temporelles massives

Jury :

Pierre Gançarski,	Professeur, Université de Strasbourg	Rapporteur
Maria Rifqi,	Professeure, Université Paris II-Panthéon-Assas.	Rapporteure
Lynda Tamine-Lechani,	Professeure, Université Paul Sabatier	Rapporteure
Massih-Reza Amini,	Professeur, Université Grenoble Alpes	Examinateur
Sadok Ben Yahia,	Professeur, Faculté des Sciences de Tunis	Examinateur
Nicolas Passat,	Professeur, Université de Reims Champagne-Ardenne	Examinateur
Myriam Lamolle,	Professeure, Université Paris 8, IUT de Montreuil	Garant.e

H
D
R

Remerciements

Je tiens à remercier de nombreuses personnes : mes collègues à l'IUT, au LIASD, au MAP5, au CREM, mes étudiants en thèse, ma petite famille. Je voudrais remercier Myriam Lamolle d'avoir accepté d'être garante de mon Habilitation à Diriger des Recherches. Elle a été très disponible pour m'accompagner dans différents projets de recherche, dans le co-encadrement des thèses et dans la lecture des premières versions de ce mémoire. Je tiens à remercier Maria Rifqi, Lynda Tamine-Lechani et Pierre Gançarski d'avoir accepté d'être les rapporteurs de ce mémoire et Sadok Ben Yahya, Massih-Reza Amini et Nicolas Passat d'avoir accepté de participer au jury.

J'adresse mes remerciements tout particulièrement à Anne Ricordeau, mon binôme en recherche et en enseignement pendant 10 ans et qui m'a permis de démarrer les premiers projets de recherche et collaborations avec le laboratoire MAP5 de l'université ParisDescartes.

Je voudrais également remercier tous mes collègues du département Informatique, en particulier Jean-Hugues Réty pour avoir repris la responsabilité de la licence CSID que je dirigeais depuis douze ans et Philippe Bonnot qui a su tailler l'emploi du temps à mes disponibilités. Cela m'a permis de rédiger ce mémoire dans de bonnes conditions.

Enfin, je remercie mon époux et mes enfants d'avoir été patients pour accomplir mon projet professionnel.

Résumé

Face à un déluge des données, les tâches usuelles d'extraction de connaissances que nous avons connues il y a dix ans prennent de nouvelles directions à la fois théoriques, technologiques, méthodologiques et applicatives. C'est-à-dire, qu'il est possible par exemple de recueillir des informations variées sur les sources des données, les modalités d'usage (les habitudes) ou encore les comportements des utilisateurs qui se manifestent à travers l'émotion, les opinions, sa localisation, etc. Le thème de recherche phare qui structure mes travaux est la modélisation de la prédiction d'événements dans le futur à partir des données massives observées dans le passé. Nous partons de l'hypothèse que la finalité de ce déluge de données n'est pas seulement de les stocker mais de pouvoir en sélectionner les plus pertinentes, d'en extraire des relations sémantiques, d'en déduire des corrélations, des relations de causalité ou d'augmenter certaines données. Le but étant d'enrichir continuellement la prédiction par des connaissances collectives, ouvertes, générales et/ou spécifiques. Quand on dispose de données hétérogènes (mélange de textes et images), la problématique de construction des descripteurs se définie différemment des approches classiques. C'est d'abord se posent les questions de comment déterminer les objets signifiants dans une image et comment utiliser les annotations textuelles comme source contextuelle de données signifiées pour l'image. C'est avec une approche hybride de web sémantique et d'analyse d'image que nous avons abordé ces problématiques connexes à la recherche sémantique d'images à partir de requêtes textuelles/visuelles ou l'indexation sémantique d'images. Le web sémantique a contribué en effet à imposer ses modèles de représentation des connaissances par des ontologies et de pouvoir formaliser les relations qui peuvent exister entre les objets signifiants de l'images augmentées par les connaissances extraites des annotations associées. Ensuite se posent les questions du volume et de la temporalité des données. Pour y répondre, nos méthodes de parallélisation du processus de construction apportent davantage de modularité contribuent à la réduction du temps de calcul, en particulier dans le contexte des images satellitaires et le suivi des changements géoclimatiques ou des changements d'urbanisation. Pour ces contexte, les algorithmes de traitement sont entièrement indépendant des solutions de stockage et d'accès aux données en utilisant principalement l'écosystème Spark. nous présentons enfin nos travaux qui s'attellent à la prédiction et la prévision dans le contexte de données massives et temporelles. Une attention particulière est accordée à ces données en ce qui concerne la vérification permanente de leur cohérence avant la phase d'analyse. C'est une étape supplémentaire que nous proposons en amont de l'analyse afin de maintenir une performance de prévision interprétable dans le temps.

Abstract

Faced with a deluge of data, the usual knowledge extraction tasks that we experienced ten last years are taking new conceptual, technological, methodological and applicative directions. In other words, it is possible, for example, to collect a variety of information on data sources, patterns of use or user behaviors that are manifested through emotion, opinions, location, etc. The leading research theme that structures my work is the modeling of the forecasting of events in the future based on big data observed in the past. We suppose that the purpose of this deluge of data is not only to store them but to be able to select the most relevant ones, to extract semantic relationships, to deduce correlations, causal relationships or to increase certain data. The goal is to continuously enrich prediction with collective, open, general and/or specific knowledge. When heterogeneous data are available (mixture of texts and images), the problem of feature construction is defined differently from standard approaches. First it arises some questions as how to determine the significant objects in an image and how to use textual annotations as a part of contextual source of meaningful data for the image. It is with a hybrid approach of semantic web and image analysis that we propose to resolve these issues related to semantic image retrieval based on textual/visual queries or semantic image indexing. The semantic web has indeed contributed to establishing its models of knowledge representation with ontologies and to be able to formalize the relations that could exist between the signifying objects of the image augmented by the knowledge extracted from the associated annotations. Then the questions of the volume and temporality of the data are raised. To answer these questions, our methods of parallelization of the construction process bring more modularity and contribute to the reduction of calculation time, particularly in the context of satellite images and the monitoring of geoclimatic changes or changes in urbanization. For these contexts, the processing algorithms are completely independent of data storage and access solutions, using mainly the Spark ecosystem. Finally, we present our work that deals with forecasting problem in the context of big and temporal data. Special attention is given to these data with respect to the ongoing verification of their consistency before the analysis phase. This is an additional step that we propose earlier in the analysis leading to maintain an interpretable forecasting performance over time.

Table des matières

I Ingénierie des descripteurs visuels et sémantiques sur des données multi-modales **9**

1 État de l'art : fouille de données Complexes **11**
1.1 Introduction 11
1.2 Revue de la littérature 13
1.3 Processus d'analyse et extraction de connaissances 15
1.4 Approches d'analyse 18
1.5 Données complexes et massives 22
1.6 Défis à venir pour l'analyse des données massives 25

2 Extraction de descripteurs visuels **30**
2.1 Introduction 30
2.2 Approche fondée sur les connaissances du domaine 31
2.2.1 Extraction d'objets sémantiques 32
2.3 Vers une approche de segmentation sémantique d'images 46
2.3.1 Approche *« split and merge »* 48
2.3.2 Apport du *watershed* 50
2.3.3 Distribution avec MapReduce 53
2.3.4 Vers une segmentation sémantique 57
2.3.5 De l'image segmentée à l'image labellisée par apprentissage profond 60
2.4 Conclusions 65

3 Enrichissement sémantiques des descripteurs visuels **67**
3.1 Sémantique, texte et images : quelles correspondances ? 69
3.2 Images médicales et descripteurs sémantiques 71
3.2.1 Images 2D 72
3.2.2 Images 3D 73
3.3 Descripteurs sémantiques d'images photographiques 77
3.3.1 Relations spatiales 78
3.3.2 Formalisation des relations spatiales 79
3.3.3 Relations sémantiques 85
3.4 Enrichissement des domaines et module ontologique 91
3.4.1 Conceptualisation du lexique 92
3.4.2 Détection des relations 94
3.5 Représentation multi-dimensionnelle des connaissances 97

3.5.1 Raisonnement 99
3.6 Conclusions 100

II Apprentissage supervisé et prédiction sur des données massives, spatiales et temporelles 103

4 Analyse Prédictive et Données Temporelles 105
4.1 Introduction 106
4.2 Prédiction et apprentissage supervisé 106
4.2.1 Cadre formel de l'apprentissage supervisé 107
4.2.2 Modèles prédictifs induits par l'espace des données 111
4.3 Modèles prédictifs pour des données massives 114
4.4 Modèles prédictifs et images spatio-temporelles 118
4.4.1 Données spatio-temporelles simulées 121
4.4.2 Description du simulateur 124
4.4.3 Prédiction de labels pour images satellitaires 127
4.5 Prédiction de séries temporelles issues de capteurs 133
4.6 Prédiction à partir des données de réseaux sociaux 146
4.6.1 Word2Vec et Doc2Vec pour la représentation textuelle 148
4.6.2 Prédiction de séquences symboliques 152
4.7 Conclusions et premières perspectives 157

5 Conclusions et projets scientifiques 159
5.1 Fusion Multimodale et explicabilité des algorithmes 161
5.2 Image et recherche visuelle sémantique 163
5.3 Apprentissage par transfert 165

Liste des figures

1.1 Processus de base d'analyse des données inspiré de crisp-dm.org 18
1.2 Hiérarchie des analyses (source [Del15]) 19

2.1 Représentation multi-échelles de la sémantique des images 32
2.2 Os trabéculaire du calcanéum 34
2.3 Radiographie du calcanéum 35
2.4 Connexités : (A) pour l'objet et (B) pour le fond 37
2.5 Squelette sur la radiographie du calcanéum 39
2.6 Volume reconstruit du calcanéum 40
2.7 Diversité des objets dans des images photographiques 41
2.8 Étapes de construction des règles d'association 43
2.9 Extrait de l'ontologie de domaine 44
2.10 Extrait des règles d'association 45
2.11 Exemple d'un arbre quadtree 48
2.12 Processus de l'algorithme de segmentation SWM 53
2.13 (a) Image à segmenter ;(b) représentation encodée des quadtrees ; (c) représentation logique ; (d) fonction map() dans MapReduce 55
2.14 Meilleure segmentation obtenue pour le type « forêt » 56
2.15 Meilleure segmentation obtenue pour les routes et les forêts 57
2.16 Macro-architecture du réseau UNET inspirée de [SNIU19] [1] 61
2.17 Macro-architecture du réseau VGGNet [SZ15] 62

3.1 Images extraites de la collection Image CLEF 2008 70
3.2 Les trois cas de figures d'ILS 76
3.3 Squelette du volume initial 77
3.4 Classification des relations spatiales [HAB08] 79
3.5 Relations spatiales orientées 79
3.6 Encodage de la distance en sous-ensembles flous et résultats de recherche avec la relation *très proche* 81
3.7 Exemples de boîtes englobantes 82
3.8 Exemple de relations d'ordre 83
3.9 Exemple de relations d'ordre : l'objet A est aligné avec l'objet B. 83
3.10 Requêtes spatiales 84
3.11 Images résultats des requêtes 84
3.12 Processus de construction du module ontologique en 5 étapes. 92

3.13 Catégorisation des relations taxonomiques et sémantiques entre concepts . 94
3.14 Extraction des relations sémantiques pour des phrases sans verbes 96
3.15 Exemple de patron pour la représentation des différents niveaux conceptuels des connaissances 98
3.16 Extrait d'une visualisation d'un module ontologique 99
3.17 Extrait détaillée d'une visualisation modulaire 100

4.1 Deux étapes illustrées par les images du simulateur 3D : Résorption et formation 126
4.2 Visualisation 3D par le simulateur. zones A : résorption est plus importante que la formation ; Zone B : résorption puis formation. Zones en rouges sont les formes de la zone de remodelage (BMU). 127
4.3 Résultat du remodelage en 3D 128
4.4 Extrait des paramètres du simulateur de remodelage en 3D et dans le temps 128
4.5 Modèle de prédiction distribué d'objets à partir d'images satellitaires[BMC+19] 131
4.6 Images satellitaires 3 bandes et exemples de prédiction[Mel19b] 133
4.7 Images satellitaires 16 bandes et exemples de prédiction 134
4.8 Trois régimes : (1) état stationnaire ; (2) effacement électrique ; (3) dégivrage 137
4.9 Régime sans et avec charge : (a) sans charge, (b) avec charge[Mah20] . . . 137
4.10 Composition d'une cellule LSTM 140
4.11 Prédiction des séries temporelles par modèles LSTM sur les données de E_1 143
4.12 Prédiction de la série temporelle $Tp_{product}$ par modèles LSTM sur les données de E_2 144
4.13 Prédiction de la série temporelle $Tp_{product}$ par modèles LSTM sur les données de $Compressor_{Energy}$ sur les données de E_3 144
4.14 Prédiction des séries temporelles par modèles LSTM sur les données de E_5 144
4.15 Du texte aux vecteurs en passant par les tokens [Cho17]. 150
4.16 Exemple d'une série chronologique encodée en une séquence SAX. À chaque position de la fenêtre, la valeur moyenne est calculée puis encodée avec un symbole. 154
4.17 Représentation spectrale de certains canaux de diffusion d'offres d'emploi avec la série symbolique SAX. Chaque ligne verticale représente une séquence symbolique d'un canal. Chaque pixel de la ligne représente la quantification des clics avec la série symbolique SAX 156
4.18 Représentation spectrale de certains canaux de diffusion d'offres d'emploi avec la série symbolique PMVQ. Chaque ligne verticale représente la séquence symbolique d'un tableau d'affichage des offres d'emploi. Chaque pixel de la ligne représente la quantification des clics avec les PMVQ. . . . 156
4.19 Valeurs de précision moyenne de la prédiction avec des LSTM entraînés sur des séquences symboliques. Les résultats sont affichés pour chaque résolution (allant de 1 à 8) avec SAX et PMVQ. 157

Liste des tableaux

1.1 Une minute sur Internet (source : www.visualcapitalist.com) 13
1.2 Requêtes utilisées pour élaborer la revue de littérature 16
1.3 Total des articles retenus 17

2.1 Résumé des projets et des encadrements sur la thématique de la construction de descripteurs 33
2.2 Moyenne des résultats obtenus par VGG et UNET sur les trois collections d'images. 63

3.1 Résumé des projets et des encadrements sur la thématique de l'enrichissement sémantique des descripteurs visuels 69
3.2 Paramètres descriptifs de la structure 2D à base de segments et de points d'intersection. 73
3.3 Trois classes de structures à l'issue des valeurs propres idéales. 75
3.4 Mesures de similarité sémantique 89
3.5 Relations dérivées des ressources intégrées à BabelNet 93

4.1 Récapitulatif des travaux et collaborations sur la prédiction à partir de données massives et hétérogènes 119
4.2 Comparaison des résultats sur des bases d'images à 3 et 16 bandes. 132
4.3 Entrées/sorties du système dynamique d'une chambre froide 136
4.4 Prédiction de la Température T_p pour chaque E_i avec quatre modèles dérivés du LSTM. 143
4.5 E_i $Compressor_{Energy}$ prediction with the four derived LSTM models . . . 143

Contexte général et thématiques de recherche

Introduction

Les données multimodales sont omniprésentes dans notre vie, qu'elles soient personnelles ou professionnelles grâce à la transformation numérique. Dans des domaines comme l'imagerie médicale, la télédétection, les systèmes d'information géographique, ou encore les humanités numériques, les données se multiplient, deviennent de plus en plus volumineuses et d'origine très diverses. Elles peuvent venir de travaux sur le terrain, d'expérience de laboratoire ou de simulations. Il faut donc prendre en considération de nouvelles sources de données générées par des machines telles que les capteurs mais aussi générées par l'homme comme les données provenant des médias, des réseaux sociaux mais aussi celles relatives aux parcours de navigation issues des interactions sur la toile. De surcroît, avec l'augmentation des capacités de stockage et de traitement informatique des données, la taille des bases de données à disposition s'est élargie d'une manière exponentielle dans de multiples domaines.

Ainsi, les données massives ne sont pas volumineuses uniquement parce qu'elles sont nombreuses. Elles le sont aussi et surtout par leur diversité, puisqu'elles regroupent non seulement des données empiriques venant de nombreuses disciplines mais également des données provenant des simulations et de résultats d'autres recherches effectuées parfois dans des domaines connexes voire très différents. La finalité de ce déluge de données n'est pas de les stocker mais de permettre d'anticiper ce qui va se passer dans l'avenir, soit de prédire

les tendances futures avec précision. Ainsi une analyse prédictive révèle les modèles et les relations entre les données. Elle rend possible des prévisions à plus ou moins long terme avec une probabilité élevée sur la base de l'historique des données collectées, ainsi que les tendances qui s'y dessinent. L'analyse prédictive englobe diverses techniques statistiques, telles que l'exploration de données, la modélisation prédictive et l'apprentissage automatique. En effet, les grandes quantités de données sont trop complexes et volumineuses pour être traitées et analysées sans automatisation. Il est alors nécessaire d'utiliser des moyens de calcul algorithmique et donc des moyens informatiques idoines pour rechercher parmi ces données les informations pertinentes et les synthétiser. Ces outils doivent relever le défi d'outrepasser la complexité de ces données afin de pouvoir exploiter la totalité des informations disponibles et renforcer ainsi les connaissances extraites. Cela suppose que la machine soit capable d'intégrer des informations de nature différente par l'intermédiaire de descriptions sémantiques permettant de les lier à une même catégorie sémantique. La multitude des sources de données (ou données multi-sources), la diversité des formats de représentation d'une même donnée, l'évolution temporelle, le passage à l'échelle, l'accès, l'interrogation et l'analyse de ces nouvelles masses de données sont essentiels pour élargir les connaissances du domaine y afférent.

Avant d'exposer mes thématiques de recherche de ces dernières années, un focus sera fait sur deux notions fondamentales autour desquelles gravitent mes travaux : la notion de complexité liée aux données et la notion de passage à l'échelle des données à traiter.

Données Complexes

La gestion des données collectées à partir de différentes sources et supports est devenue un défi d'envergure. En effet, l'explosion technologique des outils d'acquisition a contribué à la mutation des données du monde réel. Elles ne se présentent plus de manière *uni-modale* et structurée. Par exemple, un dossier médical contient non seulement des données numériques pouvant être structurées de manière *tabulaire* comme des résultats d'analyses radiologiques, mais il intègre également des données textuelles comme les comptes-rendus cliniques, le suivi temporel d'observations cliniques, ou encore des graphiques tels qu'un

électrocardiogramme voire des images complexes fournies par les radiographies, échographies, etc. De même, les données intégrées par des sites web se présentent selon différents modes. Elles peuvent être sous la forme de tableaux, de textes, de graphiques, d'images, voire même de fichiers audio ou vidéo. Elles ne sont pas forcément redondantes du point de vue informationnel. Bien au contraire, elles véhiculent souvent des connaissances complémentaires et permettent quelques fois de révéler des connaissances nouvelles. Ainsi dans bien des domaines applicatifs (comme le multimédia, la télédétection, l'imagerie médicale, les systèmes d'information géographique, la bio-informatique, etc.), les données collectées présentent différents formats de codage (pour le texte, les graphiques, les images, la vidéo ou audio), contiennent des catégories variées d'information (qu'il s'agisse de descriptions factuelles ou non), etc. Leur sémantisation s'appuie sur l'usage de ressources ontologiques pour transformer l'information en connaissances contextuelles. En effet l'intégration des ontologies dans le processus d'extraction de connaissances permet de formaliser les concepts d'un domaine ainsi que les relations qu'ils entretiennent. Grâce à leur représentation formelle, des inférences peuvent être effectuées pour déduire de véritables connaissances contextuelles.

La première apparition du terme *donnée complexe* a eu lieu en France lors du premier atelier sur la *Fouille de données complexes dans un processus d'extraction des connaissances* par Gançarski et Trousse en 2004 [2]. [DBRA07] proposent une première définition exhaustive qualifiant les données de complexes si elles sont : 1) multiformats, 2) multistructures , 3) multisources, 4) multimodales, 5) multiversions (évolutives en termes de définition ou de valeur).

Au vue de l'évolution des technologies de collecte des données, et de l'évolution des besoins applicatifs, il est très difficile de dresser une définition immuable des données complexes. En effet, plus récemment, les avancées technologiques des capteurs et des objets embarqués qui se répandent à grande vitesse grâce à la démocratisation de la 5G ont fait émerger une nouvelle dimension de la complexité à savoir l'acquisition de données embarquées en temps réel. On passe alors de l'*Internet of things* (IoT) à l'*Internet of Everythings* (IoE)

2. https://www.egc.asso.fr/wp-content/uploads/egc2004_atelier_fdc.pdf

nécessitant une interopérabilité *web of things* (WoT) énoncé par le W3C [3].

Données massives

La disponibilité et l'adoption de nouveaux appareils mobiles plus puissants, ainsi que l'accès omniprésent aux réseaux mondiaux, voire à différents satellites, etc, entraînent déjà la création de nouvelles sources de données en masses. L'augmentation exponentielle des données est, en particulier, portée par les objets intelligents qui seront plus de 50 milliards dans le monde en 2020. À cette échéance, ce sont 40 000 milliards de milliards de données qui seront générées [4]. Les données massives, dites aussi *big data*, sont définies par au moins trois caractéristiques communes (appelées les **3V** dans le vocabulaire des anglo-saxons) : 1) un volume extrêmement large, 2) une vitesse d'acquisition et de traitement extrêmement élevée, 3) une très grande variété. Plus important encore est le quatrième V, la véracité. Celle-ci pose les questions fondamentales en analyse des données qui sont : dans quelle mesure ces données sont-elles exactes pour faire de bonnes prédictions ? Les résultats d'une analyse de données massives ont-ils vraiment un sens ?

Synthèse des travaux menés

Reconnaître des tendances, des corrélations dans des ensembles de données fait penser à la faculté du cerveau humain à interpréter et/ou faire des analogies, bien que les capacités de ce dernier sont très vite dépassées en terme d'analyse des données complexes et massives. Le processus de prédiction peut être vu comme un raisonnement capable d'inférer l'occurrence future d'un phénomène à partir de conditions initiales et de connaissances générales. La structure logique de ce raisonnement prédictif a considérablement évolué ces dernières décennies en s'adaptant principalement à l'évolution des propriétés des données utilisées par l'analyse. En effet, les approches prédictives traditionnelles ont été guidées

3. WOT architecture

4. D'après les informations du site planetoscope (www.planetescope.com/.../nombre-de-tweets-expedies-sur-twitter.html), l'humanité produit 183 960 000 000 tweets (350 000 par minutes), 2000 milliards de requêtes annuelles Google contre 1 204 milliards par an en 2014 (65 000 par seconde), et 2 millions d'emails. Actuellement, le trafic web mondial surpasse 2 zettabytes par an (le temps de connexion mensuel cumulé représente presque 4 000 milliards d'années)

par les modèles tant que la taille des données à analyser et leur complexité restaient raisonnable. Un modèle est une représentation abstraite du comportement d'un système après une déformation volontaire de certaines propriétés pour des raisons de simplification et un pré-établissement d'un certains nombre de descripteurs (les paramètres pertinents qui décrivent la majorité des données). Quand la taille des données et leur complexité croient, les approches à base de modèles deviennent vite obsolètes puisqu'elles ne représentent qu'une tendance spécifique d'un groupe de données. À l'inverse, les approches guidées par les données, initialement utilisées pour compléter les données manquantes dans des bases de données de taille relativement raisonnable par interpolation, s'appuient sur les liens directs et indirects entre les données et donc sur des inférences inductives différentes sans déterminer au préalable les descripteurs. Ces approches permettent de maximiser la précision des prédictions et de minimiser l'usage d'hypothèses auxiliaires. Les effets paradoxaux de ces deux avantages sont le sur-ajustement, la lenteur de l'exécution des algorithmes et l'explicabilité.

Le sur-ajustement L'analyse de gros volumes de données à des fins de recherche de corrélations sur lesquelles un système prédictif sera fondé, n'est pas dénuée de risques. On peut observer des corrélations fallacieuses [CG16] quand le jeu de données d'apprentissage n'est pas représentatif de son contexte d'exploitation (biais d'apprentissage) ou quand les données sont massives ; ce qui augmente la probabilité de découvrir des relations qui ne sont que du bruit. Le sur-ajustement consiste à utiliser des équations ayant de nombreux paramètres libres pour s'ajuster aux données sans connaissances au préalable.

L'explicapibilité L'explicabilité est l'absence de définition des chemins qui ont permis de calculer le sens du raisonnement prédictif mais aussi du pourquoi de la prédiction elle-même. Nous cherchons désormais à caractériser ce lien fort et fonctionnel entre explication et prédiction afin de dégager la trajectoire dynamique du raisonnement prédictif augmentée par des connaissances sous-jacentes et complémentaires aux données statistiques. Par exemple, on peut prédire que le tabac provoque le cancer grâce aux données statistiques mais nous ne pouvons pas expliquer comment et pourquoi fumer donne le cancer. C'est ici que les connaissances à partir d'autres domaines annexes tels que l'analyse génétique,

ou encore les données biologiques et les données radiologiques peuvent être croisées avec les données statistiques pour trouver le pourquoi. Les explications fournissent un ou plusieurs chemins cognitifs vers les prédictions qui contribuent ensuite à tester et à affiner les explications.

Le passage à l'échelle Comme cela était évoqué par Serge Abiteboul, si les algorithmes de tri des données sont bien implantés et popularisés, ils sont bien souvent trop longs à s'exécuter [5], ce qui rend utile et incontournable non seulement la mise en parallèle de nombreuses machines mais aussi l'utilisation d'outils statistiques pour filtrer ces données. Pour extraire des informations de ces grandes bases de données, on voit donc apparaître des éco-systèmes spécifiquement développés dans l'optique d'analyser des données très nombreuses et variées en des temps restreints grâce à leur distribution.

Nos travaux contribuent à lever ces verrous en proposant des raisonnements prédictifs sémantiques. Pour cela, nous nous intéressons d'une part à la représentation, la caractérisation des liens sémantiques des données complexes et massives via la construction de descripteurs sémantiques multi-dimensionnels et leur mécanisme de fusion. D'autre part, l'usage de ces descripteurs sémantiques nous permet de résumer (hors ligne), d'indexer, et de chercher efficacement les données archivées. Leur injection dans le processus d'apprentissage en ligne nous permet de réduire le sur-ajustement et de mieux expliquer la prédiction grâce à la mise en place de systèmes d'apprentissage sémantiques et ouverts. Ces recherches se déclinent principalement dans le cadre de collaborations académiques, industrielles et, ce notamment par le co-encadrement de doctorants, de post-doctorants et de stagiaires en master recherche (M1 et M2).

Organisation du manuscrit

Ce manuscrit synthétise les trois axes principaux de recherche en suivant un fil directeur détaillé comme suit. Nous débutons par le chapitre 1 d'état de l'art qui nous permet de définir le cadre central de nos travaux. Il sert de vecteur de transition aux chapitres 2, 3 et 4

5. Serge Abiteboul, "Sciences des données : de la logique du premier ordre à la Toile, Leçon inaugurale au Collège de France, 8 mars 2012" (2012), p. 21.

présentant nos contributions elles-mêmes. Après avoir introduit la notion de données massives complexes, nous présentons tout d'abord la démarche des approches existantes dans la littérature qui permettent de fusionner ces données pour en extraire des connaissances. Dans le chapitre 2, nous nous intéressons à l'extraction et la mise en relief de connaissances de bas niveau à partir de données brutes semi-structurées. Ces données collectées sont pour la plupart des images et/ou du texte. D'ailleurs, avec la démocratisation des données massives (ou *big data*), les images et le texte sont des données qui couvrent beaucoup de domaines d'application. Au stade d'extraction de ces objets, que nous nommerons objets perceptifs ou *percepts*, nous proposons de les caractériser par des descripteurs de bas niveau. Ces derniers sont, à ce moment-là, vidés de tout sens ; nous nous intéressons dans le chapitre 2 à définir les liens éventuels entre le percept et les concepts. Ces liens sont des connaissances intermédiaires entre le bas et le haut niveau. À partir de ces liens, de ces connaissances intermédiaires, nous proposons dans le chapitre 3 de définir la notion de *bagage conceptuel* associé à un percept. La notion de bagage conceptuel se construit à partir de connaissances sous-jacentes qui peuvent être d'ordre social, culturel, ethnique, etc. On s'intéresse, d'une part, à la méthodologie d'enrichissement des connaissances et, d'autre part, à la modélisation multi-niveau des percepts et multi-dimensionnelle des relations afférentes aux concepts. Les correspondances entre percepts et concepts sémantiques sont injectées par la suite dans le raisonnement prédictif en tant que bagage sémantique contribuant à l'élaboration de la trajectoire explicative du résultat de la prédiction. C'est le sujet que nous abordons dans le chapitre 4 qui présentera principalement des modèles hybrides de prédiction guidés par les données. Ces dernières pouvant varier dans le temps et dans l'espace, imposant aux modèles d'être ouvert mais surtout explicatifs. Nous finirons par le chapitre 5 où nous exposons le bilan de nos recherches et nous présentons les lignes directrices de leurs évolutions ainsi que les projets de recherche à court et à moyen termes qui sont visés.

Première partie

Ingénierie des descripteurs visuels et sémantiques sur des données multi-modales

Chapitre 1

État de l'art : fouille de données Complexes

Sommaire

1.1 Introduction 11
1.2 Revue de la littérature 13
1.3 Processus d'analyse et extraction de connaissances 15
1.4 Approches d'analyse 18
1.5 Données complexes et massives 22
1.6 Défis à venir pour l'analyse des données massives 25

1.1 Introduction

La conjoncture actuelle liée à la facilité d'acquisition des données, l'explosion des moyens techniques pour les stocker et le foisonnement de nouvelles approches d'analyse multiplie les applications possibles dans tous les domaines. Comme peuvent les illustrer les deux figures de la Table 1.1, la quantité de données qui transite sur le web par minute ne cesse d'augmenter entre 2018 et 2019. D'ailleurs, le PDG de Google a estimé que, tous les deux jours, l'humanité crée une quantité de données équivalente à la quantité totale produite depuis l'aube des temps jusqu'en 2003 [Sch10].

Si la volonté de produire et de conserver des données exhaustives existe depuis plusieurs dizaines d'années dans certains domaines (météorologie, finance, santé, etc.), la vague actuelle est sans précédent dans la diversité des activités transformées par l'essor de cette démarche. Plusieurs évolutions ont contribué de près ou de loin à cette transformation. D'abord, la popularité d'Internet qui a facilité l'interconnexion perpétuelle des données de tout type. Ensuite, la production de la donnée est aujourd'hui peu coûteuse, issue de capteurs connectés observant le fonctionnement d'un objet ou donnant des informations sur son milieu de surveillance (humidité, luminosité, température, mouvement, parole, etc.). Le déferlement des données a favorisé l'évolution des capacités de stockage, la rapidité des algorithmes d'accès et par conséquent, l'émergence de nouvelles solutions de décentralisation de traitement des données tels que la dématérialisation (ou *cloud computing*).

La baisse des coûts de télécommunication et l'augmentation des débits est le troisième facteur qui a permis de simplifier le recueil des données et d'accélérer le processus de mise à disposition pour exploitation. Enfin, le stockage des données a été favorisé par la baisse continue de son coût unitaire malgré l'augmentation des capacité, et par le développement de systèmes de fichiers distribués adéquats. Face à ce déluge de données, les tâches usuelles d'extraction de connaissances que nous avons connues il y a dix ans prennent de nouvelles directions à la fois technologiques, méthodologiques et applicatives. Dans l'objectif de situer nos travaux par rapport à l'existant (à notre connaissance), nous exposons dans la suite de ce chapitre comment les données multimodales et massives sont gérées et intégrées au processus de fouille données. Dans la littérature, le terme « données complexes » est très peu employé. Les auteurs utilisent souvent le terme données multimédia pour représenter la diversité des modalités alors que les deux notions, modalité et média, véhiculent des sens divergents. De même, en informatique, souvent on parle de *Fouille de Données*, soit *Data Mining* pour le terme anglophone (DM), et d'*analyse visuelle* alors qu'en statistique on trouve souvent le terme d'*analyse exploratoire des données*, dit *Exploratory Data Analysis* (EDA). Or, quelle que soit la discipline ou la communauté scientifique, la démarche reste identique et peut se nommer communément *analyse de données*, (DA pour *Data Analytics*). C'est un processus composé de plusieurs étapes visant à découvrir des connaissances utiles (PKDD pour *Process of Knowledge Discovery in Databases*) pour la prise de déci-

sion, voire même la suggestion de conclusions et d'actions. Afin de lever les confusions sur ces définitions, nous présentons tout d'abord dans la section 1.3 le processus d'analyse de données pour l'extraction des connaissances. Une taxonomie des méthodes d'analyse sera présentée dans la section 1.4. Nous détaillons par la suite dans la section 1.5 les caractéristiques de nos données, en particulier ce que nous entendons par données complexes. Leur gestion par les processus d'analyse sera analysée dans la section **??**. Enfin, dans la section 1.6, nous mettons en lumière certaines limites et challenges associées à ces méthodes aussi bien de gestion que d'analyse et nous positionnons les travaux décrits dans la suite de ce manuscrit.

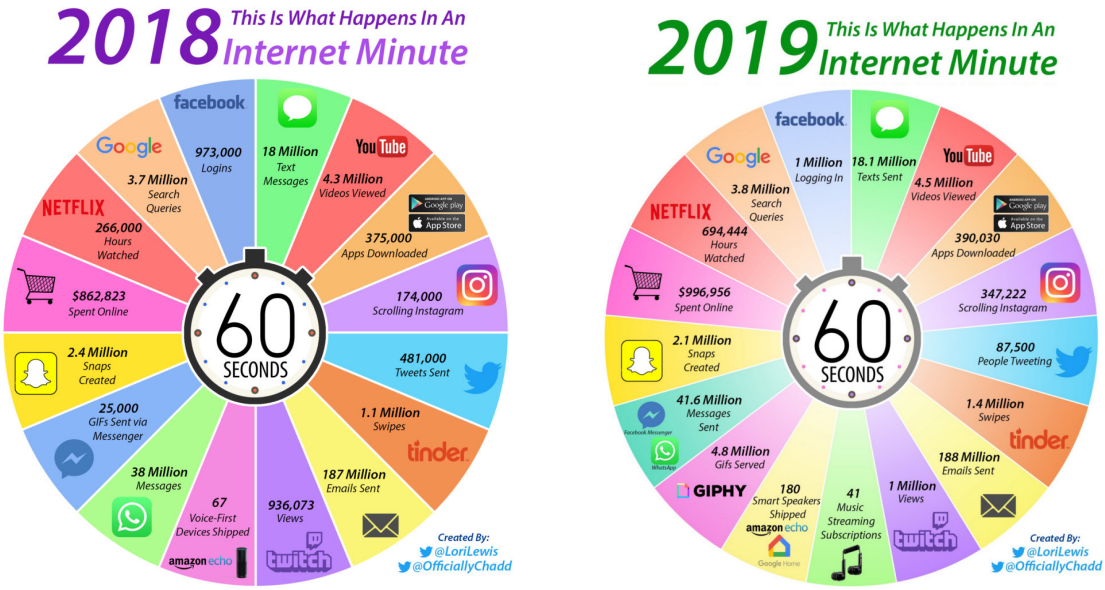


TABLE 1.1 – Une minute sur Internet (source : www.visualcapitalist.com)

1.2 Revue de la littérature

Dans les sections 1.3, 1.4, 1.5, nous réalisons une revue des articles publiés sur le thème de la fouille de données multimodales, multimédias ou données complexes. Le but de cette revue est principalement dédié à l'identification, d'une part, de la notion de données complexes et du contexte de leur usage ; d'autre part, à pouvoir identifier les caractéristiques méthodologiques de ces études par rapport à la prise en compte de ces données dans le processus d'analyse. Sans prétendre constituer une bibliographie exhaustive sur ces sujets,

nous souhaitons donner au lecteur une perception assez complète de la diversité des termes et de la richesse des méthodes d'analyse valorisant les données complexes.

Pour composer notre corpus d'articles de référence, nous avons effectué des recherches croisées essentiellement sur trois sites de recherche d'articles scientifiques à savoir 1findr [1], arXiv [2] et hal-archives [3]. Moins populaire que arXiv, la version gratuite de *1findr* est un moteur de recherche permettant d'avoir accès à un grand nombre d'articles scientifiques en ligne. Grâce à son catalogue contenant 105 542 131 articles (dont 31 771 849 en accès libre), il permet de trouver des ressources en fonction de la langue, de l'année de publication, du sujet et des auteurs. Tous les domaines sont représentés qu'il s'agisse de sciences humaines ou de sciences exactes. Nous avons cherché dans un premier temps, moyennant différentes requêtes (DMod1 à DMod3 et DMul du tableau 1.2), les articles dont le titre et le résumé contenaient au moins les deux termes *« multimodal data »* (ou *« multimedia data »*, ou *« complex data »*) et *« data mining »* (ou *« data analytics »*). Nous avons sélectionné seulement les articles qui ont été publiés entre 2000 et 2020. Sur cette sélection, nous avons retenus les premiers articles en anglais les plus cités en supposant qu'un article de référence répond au critère *« plus il est cité, plus il est pertinent »* [4]. Nous avons pu également regarder le nombre d'articles sélectionnés par thème ce qui nous a permis de mettre une attention particulière sur les articles dont la thématique est le multimedia. En effet, les documents multimedia sont considérés également multimodaux par la communauté. Nous avons privilégié en premier lieu les articles de journaux par rapport aux articles de conférence. Les résultats de nos recherches on fait apparaître que le terme *« complex data »* est très peu fréquent dans les articles en anglais. Il est souvent relié aux données temporelles ou bien aux données massives. En revanche, le terme Français « données complexes » apparaît largement sur la plateforme de recherche HAL-archives et ce depuis les années 2000. C'est un terme devenu très répandu dans la communauté de la fouille de données en France grâce à l'atelier national « Fouille de données complexes » organisé chaque année lors des journées thématiques à la conférence EGC depuis 2003. La requête de recherche DMod2 (tableau 1.2) n'a donné aucun résultat et, ce, par aucun

1. 1findr.1science.com/
2. arxiv.org
3. archives-ouvertes
4. bien qu'il puisse y avoir quelques exceptions

moteur de recherche (même par IEEE-xplore). Tandis que la requête DMod3 contraignant moins l'existence des deux termes a donné très peu d'articles. Enfin, les requêtes DMBig et MLBig (tableau 1.2) ont obtenu un grand nombre d'articles dont le pic est étalé principalement entre 2017 et 2019. Le résumé des résultats des différentes requêtes est synthétisé par le tableau 1.2. Le nombre d'articles retenus pour chaque requête est résumé par le tableau 1.3.

1.3 Processus d'analyse et extraction de connaissances

La communauté « Extraction de Connaissances » (noté KDD pour *Knowledge Discovery in Database*) définit souvent l'analyse des données comme une étape inhérente à un processus interactif et itératifs d'aide à l'extraction de nouvelles informations [HK06] [HK98]. La première version de ce processus a été présentée en 1996 par *Usama Fayyad, Gregory Piatetsky-Shapiro et Padhraic Smyth* dans [FPS96] comme, je cite, : *« a new generation of computational theories and tools to assist humans in* ***extracting useful information*** *(knowledge) from the* ***rapidly growing volumes*** *of digital data ».*

Le premier terme clé évoqué dans cette description est d'attribuer un sens aux données. Le second terme est la dimension croissante des données récoltées. C'est un processus complètement dirigé par les données par opposition aux processus dirigés par les modèles [HTF01]. Typiquement, l'analyse des données a souvent été comparée ou confondue avec l'analyse statistique des données où le problème est de trouver le plus petit ensemble de données permettant d'obtenir des estimations suffisamment fiables. Si nous synthétisons cette description, nous adoptons la définition suivante. C'est un processus qui vise à transformer des données hétérogènes de bas niveau sous d'autres formes plus compactes, plus homogènes, plus abstraites et surtout compréhensibles et réutilisables par l'humain. Les activités couvertes par ce processus peuvent prendre des formes diverses qui s'étendent de la tâche de compréhension du domaine, en passant par la préparation et l'analyse proprement dite des données, jusqu'à l'évaluation, la compréhension et l'application des résultats générés. Les principales différences entre les modèles de la littérature résident dans le nombre et la portée de leurs étapes spécifiques pouvant varier considérablement en fonction de la

Requêtes de recherche
DMod1 « mining » and « multimodal data » **Résultat par 1findr** : 50 (20 Accès Libre (AL)), répartis : IA : 7 ; computer Science : 6 ; sciences de l'information et des données : 3 , multimedia : 4. Années : 2000-2019 dont 8 en 2017 **Résultat par arXiv.org (arXiv)** : 20/287
DMod2 « multimodal data analytics » Résultats 0
DMod3 « multimodal data » and « data analytics » **Résultats par 1findr** : 5 en AL répartis : génie électrique et électronique : 1 ; automatisation et génie industriel 1 ; imagerie, reconnaissance des formes et vision : 1 ; technologie de l'éducation : 1 ; science et technologie général :1. Année : 2015-2019 dont 3 en 2019. **Résultats par arXiv** :0
DMul « multimedia data » and « data mining » **Resultats par 1findr** : 172 (62 AL), multimedia : 34 ; science de l'information et des données : 32 ; IA 11 ; systèmes informatique et ordinateur : 21 . Année 1990 - 2019, 16 en 2016 **Résultat par arXiv** : 2
DMBig « big data » and « data mining » Résultats 4 788 (2 240 AL), répartis : sciences de l'information et des données : 455 ; ordinateurs et informatique : 382 ; systèmes informatiques : 210, Réseau et télécommunications : 158 ; science et technologie, général : 153. Année 2001-2019, 914 en 2016, 2017 **Résultat par arXiv** : 90
MLBig « big data » and « machine learning » Résultats : 3 227 (1 461 AL) science de l'information et des données : 236 ; IA : 182 ; ordinateur et informatique : 155 ; réseau et télécommunication : 131 ; science et technologie : 130. Année 2010-2019, 900 en 2018, 2019 **Résultat par arXiv** : 305

Table 1.2 – Requêtes utilisées pour élaborer la revue de littérature

nature des données et des objectifs [WFHP16a] de l'application.

Le processus standard [HK13] se compose au moins des étapes suivantes (illustré par la

Nom de la requête de recherche	Nombre d'articles retenus
DMod1	11 (IA + Multimedia)
DMod3	1 (Imagerie, reconnaissance des formes)
DMul	10 (5 IA + 5 Multimedia)
DMBig	5 entre 2016-2019
MLBig	5 IA
	Total =32 articles

TABLE 1.3 – Total des articles retenus

figure 1.1) :

1. l'acquisition des données depuis des sources multiples ;
2. la sélection des données, effectuée à l'aide de requêtes interrogeant des bases ou des entrepôts de données, afin de collecter les informations relatives au problème pour lequel on souhaite construire de nouvelles connaissances ;
3. le nettoyage des données, souvent appelée phase de pré-traitement. Elle cherche à éliminer les données inexploitables (par exemple, des images incomplètes suite à l'interruption de la source d'acquisition) ou non pertinentes à l'étude, réduire le bruit, corriger les valeurs erronées ou manquantes. Certaines tâches, notamment la détection des données non exploitables, se font manuellement par un expert du domaine d'application ayant une connaissance suffisante des techniques d'acquisition. D'autres tâches peuvent s'effectuer avec un ou plusieurs algorithmes. Ensuite, les données sélectionnées et nettoyées sont formatées et préparées pour la phase suivante ;
4. l'analyse des données est une étape centrale du processus. L'algorithme est choisi selon le type des données, la problématique applicative et l'environnement de déploiement de l'application finale. Les données sélectionnées et pré-traitées sont explorées avec un ou plusieurs algorithmes. Ces algorithmes peuvent, par exemple, générer un ensemble de motifs, des règles, un regroupement par classe, une conclusion ou une action. Même si l'étape d'analyse de données n'est qu'une partie du processus général, elle est celle qui stimule de large travaux de recherche dans la littérature ;

5. l'évaluation des nouvelles connaissances consiste tout d'abord à valider les connaissances par rapport à ce qui est déjà connu et à évaluer la pertinence des nouvelles connaissances via des mesures idoines ;
6. l'interprétation des connaissances via des algorithmes de visualisation facilitant la restitution des résultats par les utilisateurs.

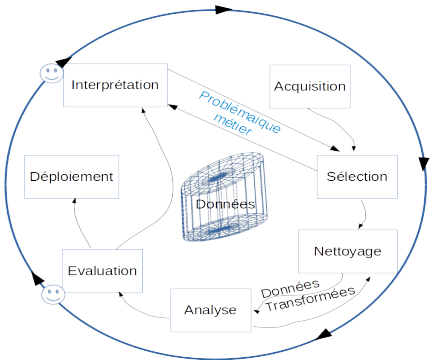


FIGURE 1.1 – Processus de base d'analyse des données inspiré de crisp-dm.org

Dans la section suivante, nous allons nous focaliser principalement sur les méthodes d'analyse. Une taxonomie de ces méthodes sera également présentée dans le but de bien illustrer la dépendance entre l'algorithme choisi et le type des données dont dispose l'application.

1.4 Approches d'analyse

Pendant très longtemps, nous avons utilisé le terme analyse de données (DA) pour dire fouille de données (DM) et inversement. De plus en plus, on parle désormais d'analytique de la donnée (*Data Analytics* (DAC)). C'est un nouveau terme qui s'ajoute au glossaire de l'analyse des données. Force est de constater que *Data Analysis* et *Data Analytics* se confondent tant dans leur sémantique que dans leur pratique [5]. Alors qu'en réalité ces trois termes ont des objectifs bien distincts. DM a pour objectif d'établir des associations et des relations entre les données, souvent cachées ou non évidentes, en fouillant de grands volumes de données réparties sur plusieurs bases de données relationnelles [HK06] [HTF01]. Les données, ici, doivent être structurées [HK13]. À l'inverse, DAC permet d'examiner des données brutes souvent en grand volume afin de mettre en lumière des informations

5. Les mots *Data Analytics* et *Data Analysis* renvoient sur la même page Wikipedia anglaise.

compréhensibles fondées uniquement sur ce qui est déjà connu [Del15]. Ici, les données peuvent être structurées ou semi-structurées. Actuellement, on parle même de *big DAC* [MNW+19].

DA est le processus complet que nous avons décrit dans la section 1.3, capable de considérer tout type de données (structuré, semi- ou non structurées). Certains travaux classent les trois approches de manière hiérarchique : $DM \subseteq DAC \subseteq DA$ [HK13] [WFHP16a]. Les différents niveaux hiérarchiques d'analyse ainsi que les techniques sont illustrés sont largement discutés dans [Del15] (*cf.* Figure 1.2).

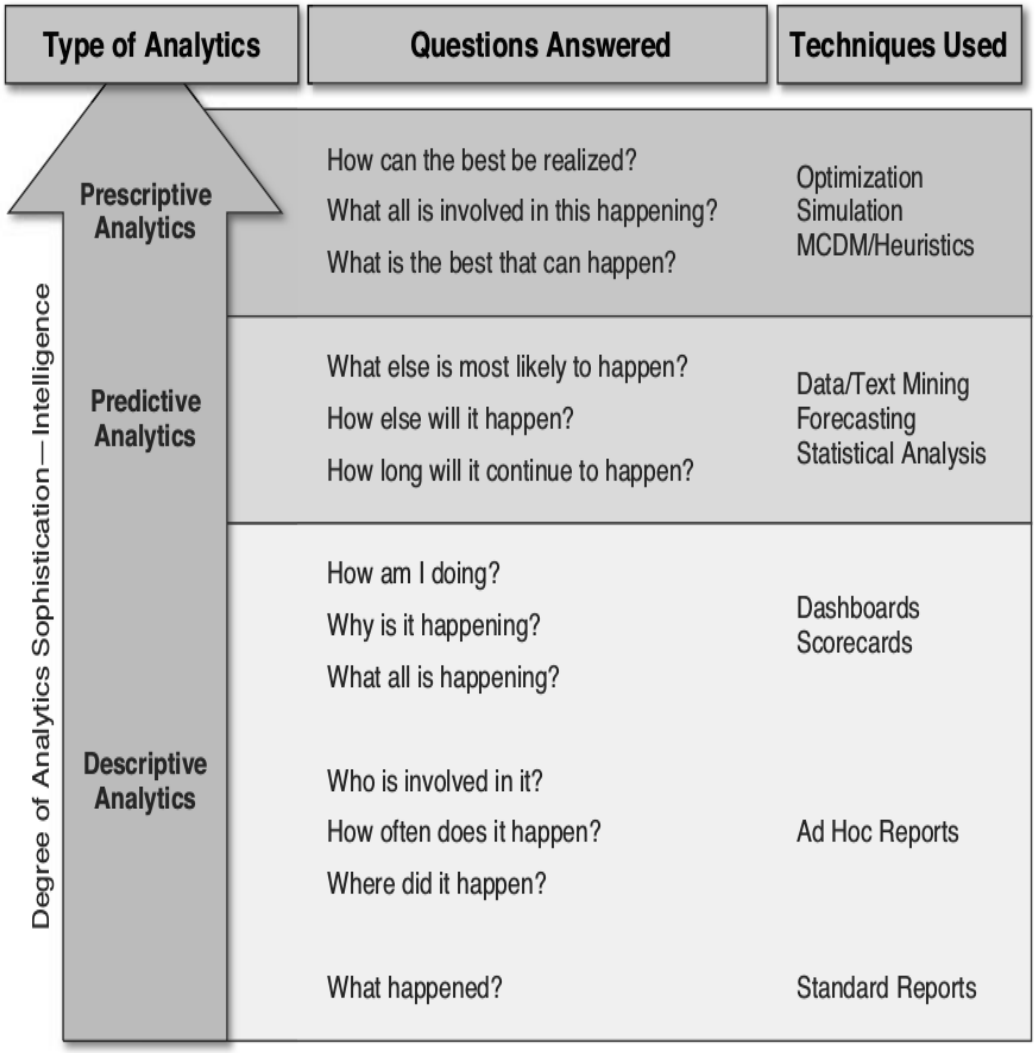


FIGURE 1.2 – Hiérarchie des analyses (source [Del15])

L'évolution de ces termes peut s'expliquer principalement (et non pas entièrement) par trois facteurs. Le premier facteur s'explique par l'augmentation des quantités de données

durant ces vingt dernières années. En effet, le passage à l'échelle a posé de vrais défis techniques quant à la représentation des données. Les entrepôts de données (ou *datawarehouse*), par exemple, ne sont plus adaptés pour gérer des collections gigantesques de données [Agn14]. Le second facteur est la transformation digitale, qui a favorisé l'hétérogénéité et la multiplication des sources de collection, tels que les réseaux sociaux, les logs, les capteurs, les objets connectés. Tous ces canaux de récolte des données favorisent la connexion des données et leur enrichissement. Le troisième facteur (sans doute pas le dernier) est lié aux changements des pratiques d'analyse et à la diversité des besoins d'analyse. Aujourd'hui la pratique de prise de décision dans les entreprises consiste d'abord à comprendre les chiffres (les données) puis à utiliser cette compréhension au service d'une prise de décision plus intelligente et cohérente avec les données. Malgré la multitude des définitions de ces trois termes dans la littérature, elles sont toutes convergentes en ce qui concerne la taxonomie des approches d'analyse (DA) [Shm10] [MNW+19] [PHMD18]. Elle se décline en trois grandes familles de méthodes permettant de répondre aux trois questions suivantes : 1) que s'est-il passé ? 2) que pourrait-il se passer ? 3) que devrions nous faire ?

1. **Analyse descriptive :** elle répond à la première question et vise à obtenir des renseignements décrivant une situation, un événement ou l'état d'un objet afin d'établir ce qui s'est produit. Plusieurs techniques en statistique simple permettent de décrire le contenu d'un ensemble de données ou d'une base de données. En particulier, les statistiques descriptives, y compris les mesures de tendance (moyenne, médiane, mode), mesures de dispersion (écart-type), les méthodes de tri, les distributions de fréquences, les distributions de probabilités et les méthodes d'échantillonnage, sont de bons outils. D'autres techniques sont également utilisées pour résumer les données ou bien pour les décrire. Issues des approches d'apprentissage automatique, les méthodes de recherche de motifs dans les données est très utilisée pour décrire les différentes formes d'association cachées dans les données. Le partitionnement de données (ou *clustering*), qui vise à identifier dans un ensemble de données des groupes homogènes partageant des caractéristiques communes, est une autre technique descriptive classique de la famille des approches d'apprentissage automatique. Dans [JDM00], une synthèse de ces techniques est largement décrite permettant

de comparer certaines des méthodes bien connues utilisées à différentes étapes d'un processus d'analyse descriptive. Le résultat de ce processus peut être utilisé pour trouver d'éventuelles opportunités décisionnelles. L'analyse descriptive aide à dégager le contexte nécessaire à l'analyste pour prendre une décision et réaliser des actions.

2. **Analyse prédictive :** répondant à la question deux, elle propose des modèles pour prédire de nouveaux événements et aider à choisir de futures actions en conséquence. Elle est de nature probabiliste et se décline en différentes méthodes telles que la classification, la régression, l'analyse de séries temporelles, de la recherche opérationnelle, etc. Elle permet de prévoir ce qu'il pourrait se passer à partir des données descriptives accumulées au fil du temps. Par exemple, la prédiction permet d'estimer les valeurs futures de certaines variables. Si la variable est catégorielle, on parle de classification, sinon de régression [CUD19] [ATB+19] [Mil14]. Si la variable prédite dépend du temps, on parle de prévision dans les séries temporelles. Ces prévisions sont généralement caractérisées par des mesures donnant des indications sur la confiance que l'on peut avoir dans la prédiction. Son but est donc d'identifier les variables prédictives et de construire des modèles prédictifs dans une analyse descriptive.

3. **Analyse prescriptive :** elle apporte une réponse à la troisième question qui cherche à prévoir la meilleure action quand un événement futur est susceptible de se produire. Elle se trouve en haut de la hiérarchie analytique, et vise à déterminer la ou les meilleures solutions déjà établies sous forme de plusieurs plans d'action par l'analyse prédictive ou descriptive [CD14]. Généralement des modèles mathématiques sophistiqués d'optimisation, de simulation et de modélisation décisionnelle heuristique sont employés pour parvenir aux solutions optimales. Pour plus de détails, nous renvoyons le lecteur sur une revue récente de cette analyse proposée par [LBAM20]. Même si l'analyse prescriptive est au sommet de la hiérarchie, les méthodes utilisées ne sont pas nouvelles.

Il est à noter que dans la littérature, **l'analyse diagnostique** ou causale, n'a jamais été considérée comme une approche d'analyse de données à part entière. Or, elle s'appuie sur l'analyse des données passées pour déterminer la cause de certains événements. Par

conséquent, l'analyse diagnostique augmente l'analyse descriptive en posant la question 1bis) *pourquoi certains événements se sont produits ?*, en utilisant les tendances des données recueillies. Le processus d'analyse diagnostique est très utilisé dans la surveillance médicale, la détection des pannes, etc. L'analyse diagnostique permet de comprendre le passé pour mieux prédire le futur. C'est pourquoi, nous l'ajouterons à la hiérarchie proposée par la communauté entre l'analyse descriptive et l'analyse prédictive. Les techniques employées sont généralement de la régression, de la classification ou du *clustering*. Enfin, ces quatre types d'analyse peuvent être utilisés séparément ou conjointement pour une prise de décision.

1.5 Données complexes et massives

Il est évident que nous vivons dans une ère de déluge de données. D'énormes quantités de données sont générées en permanence à des vitesses inédites constituant ainsi des collections de données à grande échelle. Nous parlons depuis la dernière décennie de *Big Data*, mais peut-on réellement dire que ces données sont complexes par ce qu'elles sont volumineuse ? Quelles caractéristiques peut-on définir sur les données complexe ? Quelles caractéristiques sont communes aux données complexes et au *big data* ?

Pour proposer des éléments de réponse, nous avons revu un ensemble d'articles sur le sujet. En s'appuyant sur les travaux de la littérature, il a été très difficile de trouver une définition claire et faisant consensus des données complexes malgré l'existence de l'atelier national « fouille des données complexes » depuis 2003. En revanche, les données massives sont largement étudiées. Il existe une littérature abondante traitant le *big data* à tous points de vues, caractéristiques, stockage, traitement, décision, analyse, etc. Dans cette section, nous avons souhaité donner une définition cadré à la notion de complexité des données. Ce cadre est comparé à la définition du *big data* afin de délimiter et situer le contexte de nos travaux. Faute de consensus sur la définition, il nous a semblé intéressant de partir de la définition du terme *complexe* proposée par le dictionnaire « Larousse » : *« qui contient plusieurs parties ou plusieurs éléments combinés d'une manière qui n'est pas immédiatement claire pour l'esprit ; difficile à comprendre »*. Dans cette définition, nous

retenons principalement trois notions à savoir la combinaison, le nombre et l'interaction. À l'instar de ces trois notions, les données complexes seraient tout d'abord une collection d'un grand volume de données mêlées, combinées dont les interactions ou les liens entre ces données sont difficiles à comprendre. Les données peuvent être numériques, symboliques, booléennes, multi-dimensionnelles, multi-échelle, multi-sources. Elles peuvent également diverger par leur structure comme, par exemple, les données *ensemblistes*, les données arborescentes, séquentielles, sous forme de graphes. Elles peuvent être redondantes, contradictoires, incomplètes et entachées d'erreurs. Elles peuvent être dynamiques évoluant dans le temps. Elles peuvent arriver en flots ou en séries. Bien que la notion de données complexes n'a de sens que dans un contexte précis, nous pouvons définir un certains nombre de caractéristiques. Dans le cadre de nos travaux, nous proposons de définir l'espace de caractérisation de la complexité d'une collection de données (remarquons que le terme collection de données lève la confusion avec base de données qui est forcément structurée) selon huit dimensions :

Définition 1. ***Données complexes***

Une collection de données est qualifiée de complexe quand les données qui la composent sont :

(C1) de plusieurs types : nominale (0/1), ordinales, discrètes, continue, ratio ;

(C2) de sémantiques diverses : suivant le contexte, les données peuvent représenter des points de vue différents ;

(C3) multi-structures : très structurées telles que les données relationnelles, semi-structurées tels que les fichiers xml pour représenter des méta-données et non structurées telles que la vidéo ou les images ;

(C4) multi-dimensionnelles : ce sont des données qui ont été agrégées suivant plusieurs dimensions telles que les données d'un entrepôt ;

(C5) multi-échelles : quand les données sont hétérogènes en terme de résolution temporelle ou spatiale ;

(C6) multi-sources : liées aux nouvelles technologies de l'information et aux progrès qu'ont connus les systèmes de recueil et de collecte de données. L'émergence de nouvelles

sources de données telles que nouveaux capteurs de mesure dotés de grande précision, capteurs embarqués à bord de mobile, localisation satellitaire, Web of Things *;*

(C7) multi-modales : une modalité est une collection de données agrégées par un outil d'acquisition [LAJ15]. Les modalités principales sont le langage naturel (texte écrit ou parlé), la vision (image, vidéo) et l'audio (son, musique, voix). Par exemple, la vidéo est multi-modales puisqu'elle résulte d'une acquisition d'image et de son ;

(C8) multi-dynamiques : en flots ou en série et changeant de valeurs sur des intervalles de temps divers. Par exemple, la température de l'atmosphère change toutes les secondes alors que les précipitations changent tous les jours voire tous les mois.

L'ensemble des éléments caractéristiques décrits par la définition 1 de la donnée complexe sont les conséquences de la prolifération de données massives. Si les données massives sont des données complexes, l'inverse n'est pas vrai. En effet, les données massives sont souvent décrites par leurs dimensions (qui réfèrent à leurs Vs) dont le volume est la dimension principale. Les premières définitions des dimensions se sont limitées aux $3Vs$ [NB14] à savoir volume, vélocité et variété. Depuis d'autres dimensions ont été ajoutées [FB13] [WZD14]. En particulier la valeur, a été très vite ajoutée comme une autre dimension (souvent apparaît comme une cinquième dimension) mais définie comme une dimension concernant la sortie désirée du traitement des données massives et non des données elles-mêmes [KUG14]. Nous allons nous limiter aux 4V's suivants définis ci-dessous.

Définition 2. *Caractéristiques des données massives (*Big Data*)*

(V1) le Volume est la quantité ou l'échelle des données qui peut être définie verticalement pour représenter le nombre de variables ou horizontalement pour définir la taille de la collection (de l'échantillon) ;

(V2) la Variété décrit à la fois quatre niveaux de variation, structurelle, de type, de source et sémantique ;

(V3) la Vélocité est liée d'une part à la vitesse de génération des données et d'autre part à leur traitement en temps-réel ;

(V4) la Véracité s'intéresse à tracer les origines des sources au regard de la qualité des données ;

Force est de constater l'intersection entre les caractéristiques d'une collection de données complexes et les données massives à l'exception du volume. [GH14] considère qu'une collection de données complexes (multi-sources ou multi-modales ou diversité sémantique) de petite taille est équivalente à une collection de données simples volumineuse. Les caractéristiques décrivant la vélocité peuvent être reliées en partie à la variation dynamique dans les données complexes. En conclusion, les caractéristiques des données massives ont simplement hérité des caractéristiques des données complexes dont les effets néfastes sur le processus d'analyse s'accentuent avec le volume des données. Dans ce contexte, l'analyse doit se doter de fortes exigences tant en terme de qualité que de performance. Ces exigences seront discutées dans la section 1.6 que nous consacrons principalement à la description des challenges à venir pour l'analyse des données complexes et massives. Dans la suite de ce manuscrit, nous utiliserons l'expression **données massives** pour à la fois prendre en considération les caractéristiques liées aux 4Vs et à la complexité.

1.6 Défis à venir pour l'analyse des données massives

Nous synthétisons et complétons les limites et challenges associés au méthodes d'analyse identifiés dans plusieurs revues de la littérature. Nous présentons dans les encadrés de couleurs nos travaux réalisés et discutés dans ce manuscrit d'habilitation à diriger des recherches et nous les situons au centre des défis et voies de recherche actuelles. Le passage à l'échelle est l'un des défis majeurs de la recherche actuelle tant du point de vue algorithmique, que méthodologique ou technologique. Il est fortement lié à la conjoncture numérique. Il a conduit à supposer que la disponibilité des données en masse contribuera à l'augmentation de la performance des algorithmes d'analyse des données [GHH+14]. Cette présomption déclenche alors une multitude de défis visant tout d'abord à adapter les algorithmes traditionnels qui, jusqu'ici, étaient conçus pour de petites collections de données. Les limites de ces algorithmes ont été discutées dans une multitude de travaux [AYM+15] [NVK+15] [Suk14] [QWD+16]. Parmi eux, certains se sont focalisés sur l'usage spécifique de ces algorithmes [WFHP16b] alors que d'autres ont discutés les limites et l'émergence de nouveaux outils et de nouvelles plates-formes nécessaires à l'analyse dans le contexte des données massives [SR15] [dAB15] [WZD14].

1. Le Volume (V1) est à l'origine de trois catégories de challenges visant respectivement la construction et la sélection des descripteurs (la phase d'ingénierie des données), l'adaptation des algorithmes d'apprentissage (en phase analyse des données) et, enfin, l'évaluation (phase de décision). En effet, le volume des données peut augmenter selon trois directions. La direction verticale représente le nombre de descripteurs ou les attributs caractérisant l'échantillon des données. Quand le nombre de descripteurs augmente considérablement relativement à la taille de l'échantillon, la performance d'apprentissage des algorithmes d'analyse diminue. Ce phénomène de Hughes [Hug68], connu bien avant les données massives, est un problème classique de sélection et de réduction de variables. Faute de modularité, la plupart des algorithmes d'analyse nécessitent le chargement de la totalité des données en mémoire ou bien de les avoir localement sur disque [Par12] [KGD13]. De ce fait, quand les données augmentent selon la direction horizontale, autrement dit la taille de l'échantillon, les traitements deviennent de moins en moins performants en temps de calcul [Agn14]. Enfin, quand les données augmentent selon les deux directions, la performance s'aggrave. Cela s'explique par la complexité polynomiale du temps de calcul et de l'espace exploratoire (qui est de l'ordre de $O(n^3)$ en temps et $O(n^2)$ en espace, n étant la taille de l'échantillon, pour l'algorithme *Support Vecteur Machine*, la régression logistique comme l'analyse en composantes principales est de $O(nm^2 + m^3)$, m est la dimension verticale [TKC05]). Un autre challenge causé par l'augmentation verticale et horizontale des données est lié à l'ingénierie des descripteurs. Il vise tout d'abord la construction des descripteurs en s'appuyant typiquement sur les connaissances évolutives du domaine et, ensuite, à sélectionner ceux qui sont les plus pertinents ;

Dans le chapitre 2, nous allons présenter nos différentes méthodes de construction des descripteurs qui permettent de s'adapter à la complexité des données selon (C2), (C3) et (C4), et ce, dans un contexte de données volumineuses. Nous décrivons également nos méthodes de parallélisation du processus de construction pour apporter davantage de modularité et pour réduire le temps de calcul. Nous abordons dans le chapitre 4 le problème de déséquilibre des

classes de données [GWC+13] et les solutions apportées dans le but de préserver l'équilibre de la distribution des données et de leur représentativité. Cette problématique, liée à l'ingénierie des données massives, suscite l'intérêt de beaucoup de travaux dont certains montrent, faute de correction, la sensibilité de la plupart des algorithmes d'apprentissage tels que l'analyse discriminante, les réseaux de neurones et le SVM [JS02]. Ils montrent également l'impact négatif d'un déséquilibre extrême sur les résultats de la généralisation, de la classification ou de la prévision [BCD+14]. Dans le chapitre 3, nous décrivons nos solutions de parallélisation des algorithmes d'analyse pour la prise en compte de données temps réels ainsi que leur évaluation.

2. la Variété (V2) des données s'intéresse principalement à la dimension syntaxique et sémantique des données. Son premier défi est lié à la localisation physique des données et leur distribution sur plusieurs fichiers voire sur plusieurs emplacements physiques de stockage. De ce fait, les données doivent être dotées d'un système d'indexation efficace qui soit performant dans un environnement distribué [GHTC13]. Le second défi est lié à l'hétérogénéité des données et la nécessité de réconcilier, d'unifier leur variation syntaxique et de décoder et agréger leur variation sémantique. Cette optique d'enrichissement des données par des sources hétérogènes d'information que nous abordons dans les chapitres 3 et 4, n'est pas la vocation principale des algorithmes classiques d'analyse [JGL+14] [Zhe15] ;

Étant donnée la variation des sources dont les données sont issues, ces dernières peuvent être de différents formats, résolutions et échelles (C2, C3, C5, C6, C7). Dans le chapitre 4, nous développons deux grandes approches d'apprentissage par fusion permettant de considérer la diversité structurelle et sémantique des données de type textes, images et données issues de capteurs. Nous présentons dans un modèle générique de fusion de données complexes à base de graphe qui intègre également la dimension temporelle et spatiale des données. L'avantage de cette structure sous forme de graphe du modèle est que cela permet également une visualisation globale comme locale des données. Sa pertinence

en analyse a été mesurée dans un contexte particulier de recherche sémantique visuelle.

3. la Vélocité est l'un des effets directs du recueil des données en temps (quasi-)réel. Elle peut se manifester par la nécessité d'analyser des données en temps réel. Les challenges dans ce cas seraient de suivre les changements de distribution conditionnelle des données, connus dans la littérature par le concept de *Drift* [DM14] [Tsy04] [GZB+14]. En effet, la distribution des données est souvent supposée aléatoire indépendante et identiquement distribuée pour simplifier les modèles. Cette hypothèse est rejetée dans le contexte des données temps réels où l'analyse doit être en mesure de répondre instantanément comme, par exemple, le contexte de détection des anomalies ou des événements rares. De plus, si les valeurs des données évoluent dans le temps, il est cependant nécessaire de concevoir des familles de modèles d'analyse adaptées ;

   Dans le chapitre 3, nous présentons nos travaux qui s'attellent à la prédiction et la prévision dans le contexte de données massives et temporelles. Une attention particulière est accordée à ces données en ce qui concerne la vérification permanente de leur cohérence avant la phase d'analyse. C'est une étape supplémentaire que nous proposons en amont de l'analyse afin de maintenir une performance de prévision constante dans le temps. Nous discuterons dans le chapitre 5 les pistes de recherche permettant de mettre en place des algorithmes d'apprentissage actifs en se basant sur des indicateurs de révision des modèles d'apprentissage dans le temps.

4. la Véracité qui s'intéresse à diminuer l'incertitude que peuvent avoir les données, en plus du bruit, en ayant la trace des sources d'origine [WCP+14] [BKT00]. Plus les sources augmentent, moins il est facile de relever l'origine des bruits et de quantifier ainsi facilement la certitude.

   Dans le chapitre 3, nous présentons le début d'une étude de valorisation des données basée sur les données ouvertes (*Linked Open Data* ou LOD). C'est une démarche qui nous paraît très intéressante puisqu'elle contribue à enrichir les données existantes et donc de leur donner plus de sémantique. Elle

contribue également à standardiser les formats des sources des données en leur permettant une meilleure interopérabilité et une pérennité sémantique.

Tous ces défis sont au centre de nos travaux réalisés et de nos travaux à venir.

Chapitre 2

Extraction de descripteurs visuels

Sommaire

2.1 Introduction **30**

2.2 Approche fondée sur les connaissances du domaine **31**

2.2.1 Extraction d'objets sémantiques 32

Images médicales 2D et 3D 33

Images photographiques 41

2.3 Vers une approche de segmentation sémantique d'images . . **46**

2.3.1 Approche *« split and merge »* 48

2.3.2 Apport du *watershed* 50

2.3.3 Distribution avec MapReduce 53

2.3.4 Vers une segmentation sémantique 57

2.3.5 De l'image segmentée à l'image labellisée par apprentissage profond 60

Apprentissage par transfert pour la segmentation sémantique . . 60

2.4 Conclusions **65**

2.1 Introduction

Les niveaux perceptifs (ce qui est perçu, *i.e.* description syntaxique du contenu visuel de l'image en fonction de descripteurs et de primitives visuelles) et les niveaux conceptuels ou

sémantiques (ce qui est interprété, la signification des éléments présents dans l'image) sont considérés comme notions phares de nos travaux de recherche. Cette vue modulaire de la sémantique est un axe directeur important contribuant en particulier à séparer les concepts génériques (niveau conceptuel ou sémantique) de leur représentation dans le domaine de l'image (niveau perceptif ou bas niveau), qui est très dépendante du domaine spécifique d'application. Par exemple, si nous considérons la relation spatiale générique *proche de*, sa sémantique dans l'image ne sera pas la même dans un contexte d'interprétation d'images satellitaires que dans un contexte d'interprétation d'images médicales. Outre cette séparation entre le niveau perceptif et le niveau conceptuel, ces représentations multi- niveaux considèrent aussi plusieurs niveaux conceptuels ou sémantiques. Une première distinction doit être faite entre le niveau de l'objet et le niveau de la scène, c'est-à-dire l'agencement structurel des différents objets et les différentes relations entre eux. Une seconde distinction, se fait au niveau conceptuel qui peut être d'ordre générique, spécifique ou abstrait. Ces différentes notions ont été largement étudiées dans le cadre de la thèse de Olfa Allani (2014-2017) dans l'objectif de proposer un modèle de représentation multi-niveaux de la sémantique d'une image (illustrée par la figure 2.1). Cette représentation fait une distinction entre le niveau perceptif (bas niveau) et le niveau conceptuel (sémantique) pour lequel nous distinguons aussi trois niveaux d'abstraction à savoir (i) le niveau de l'objet qui correspond à la description des objets présents dans l'image, (ii) le niveau de la sémantique partielle (*Partial Semantics*) qui inclut les relations entre les objets présents dans l'image et (iii) le niveau de la sémantique pleine (*Real (Full) Semantics*) qui inclut un raisonnement, un processus d'inférence pour construire la description de l'image dans sa globalité.

2.2 Approche fondée sur les connaissances du domaine

La représentation des connaissances est, en effet, un point crucial dans la mise en place d'un système d'aide à la décision pour un domaine donné. Il faut savoir identifier et caractériser les connaissances de l'expert, que ce soient des connaissances sur la théorie du domaine, ou au contraire, des méthodes de résolution très personnelles ou collectives, développées au cours d'une longue pratique et d'un retour d'expérience. Les connaissances sur le domaine

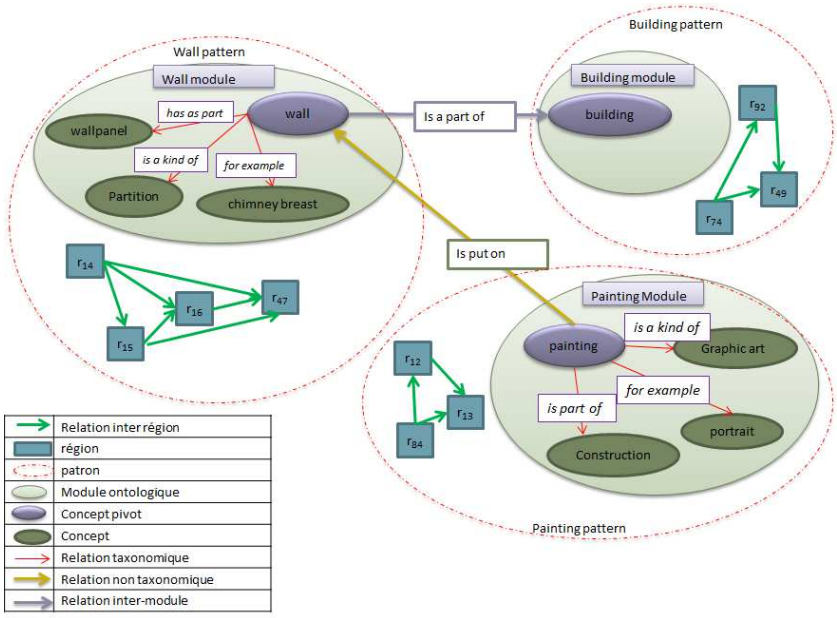


FIGURE 2.1 – Représentation multi-échelles de la sémantique des images

peuvent se décrire par deux types de descripteurs. Des descripteurs de bas niveau qui relèvent de l'exploitation des données disponibles du domaine. Ils sont construits par des modèles qui miment la connaissance théorique. Les modèles utilisés dépendent du type des données de l'étude. Par conséquent, les descripteurs de bas niveau sont fortement liés à la source d'acquisition de ces données. Par exemple, les données images dans un même domaine peuvent être acquises par des sources différentes et induisent des descripteurs de bas niveau associés différents. Ces descripteurs basiques sont souvent difficiles à interpréter et ne reflètent absolument pas la sémantique latente des données. Par sémantique, nous désignons les connaissances contextuelles basées sur les pratiques ou les expériences. C'est le but des descripteurs de haut niveau qui visent à capturer l'expertise même partielle, et à ajouter de la valeur sémantique aux descripteurs de bas niveau. Nous allons présenter dans les sections suivantes les méthodes de construction de chacun de ces descripteurs dans différents domaines et ce depuis des données massives vérifiant les critères de complexité (C1) à (C7).

2.2.1 Extraction d'objets sémantiques

Dans ce qui suit, nous cherchons à caractériser sémantiquement les objets perceptifs. C'est une phase qui a pour but d'enrichir l'objet perceptif de bas niveau par des connaissances

de haut niveau (explicites ou implicites) afin d'identifier le ou les concepts sémantiques associés.

Images médicales 2D et 3D

Dans le cadre d'une première collaboration avec le service de rhumathologie du CHR d'Orléans (IPROS) et l'école nationale des ingénieurs de Tunis via le projet CMCU [1], nous avons cherché à élaborer une méthode non-invasive permettant de détecter la maladie de l'ostéoporose.

lightgray Master/Thèse/Projet	Collaborations
Olfa Allani (2014-2017)	Thèse en co-tutelle Hajer Baazaoui (RIADI, Tunis) et Herman Akdag (LIASD, Univ. Paris 8)
projet CMCU (N° 03S1107)	Sylvie Sevestre, Anne Ricordeau (Map5, Univ. Paris 5), IPROS (CHU Orléans) et ENIT (Univ. Tunis)
projet ANR-05-BLAN-0017 « mipomodim » (Milieux Poreux : Modèles, Images.(2006-2009)	Anne Estrade, Anne Ricordeau (Map5, Univ. Paris Descartes), IPROS (CHU Orléans)

Table 2.1 – Résumé des projets et des encadrements sur la thématique de la construction de descripteurs

L'ostéoporose se manifeste par une diminution de la masse osseuse, une perte progressive de calcium et de collagène. Elle provoque une réduction considérable de la masse osseuse mais aussi une altération de la micro-architecture de l'os [BLR96]. Elle touche la femme dès l'âge de la ménopause, c'est-à-dire entre 50 et 55 ans, et l'homme plus tardivement dès l'âge de 75 ans. Cette maladie est devenue de plus en plus coûteuse à l'état puisque l'espérance de vie a largement augmenté (de 5 ans entre 2000 et 2015) selon l'Organisation Mondiale de la Santé (OMS). En effet, l'OMS a défini l'ostéoporose comme une dégradation significative de la densité minérale osseuse (DMO) et une modification de la micro-architecture osseuse chez les patients et en particulier chez les sujets jeunes (50 ans). Cependant, la quantification de la DMO n'est plus suffisante à elle seule pour l'identification des sujets malades à partir d'un âge avancé [Rec89], [HSJ98]. Le milieu médical a donc opté pour la caractérisation de

1. N° CMCU :03S1107

l'état de la micro-architecture de l'os afin d'extraire des informations pertinentes et surtout complémentaires aux descripteurs cliniques (DMO, âge, poids/taille) afin de diagnostiquer la présence de l'ostéoporose [MM97]. Pour ce faire, on a recours à l'IRM ou le rayon X à haute résolution mais à faible radiation. En effet, la modalité rayon-X la mieux adaptée en terme d'exposition et de coût pour la prévention a été utilisée dans le cadre de nos travaux afin de générer des radiographies numériques à partir desquelles nous étudions la texture de l'os du talon (calcanéum). Une analyse de la modalité du rayon-X montre que les caractéristiques 2D de la micro-architecture osseuse sont réparties et localisées sur plusieurs niveaux de résolution de l'image. L'os du calcanéum (un os de type trabéculaire) possède une architecture très complexes (*cf.* figure 2.2). Il est formé de travées qui se répartissent dans deux directions différentes (*cf.* figure 2.3). Ainsi, l'irrégularité progressive de la micro-architecture osseuse, la raréfaction des travées et leurs connectivités (liées aux croisements entre les travées) en terme de nombre et de distribution spatiale, sont des informations pertinentes pour la caractérisation des cas ostéoporotiques [KL98] [SDB+91]. En résumé, c'est l'évolution temporelle de la structure d'un état spatial isotrope à anisotrope que nous cherchons à prédire.

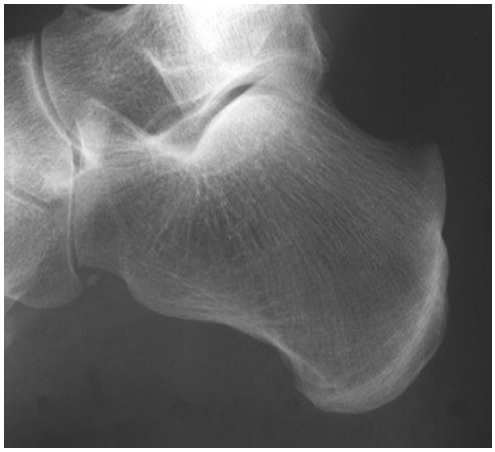


FIGURE 2.2 – Os trabéculaire du calcanéum

Toutes ces descriptions autour de la maladie portent des connaissances pouvant être représentées par des descripteurs de bas niveau tels que le paramètre de régularité d'un mouvement brownien fractionnaire modélisant les lignes de l'image [PLH+98] [Jen95] [PHB+01], les paramètres de cooccurrence et les longueurs de plages [Har79] [HSD73]. Ce sont des descripteurs fort intéressants du point de vue statistique puisqu'elles permettent de

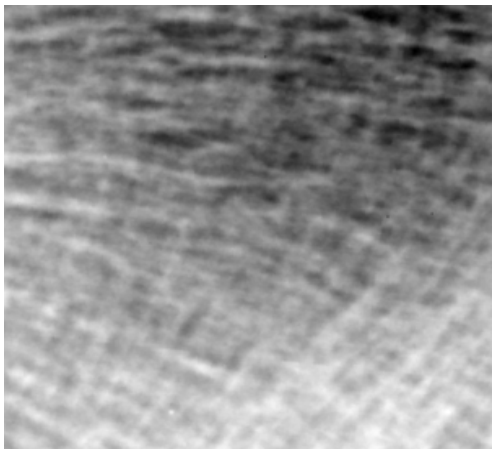


FIGURE 2.3 – Radiographie du calcanéum

moyenner l'état général de la micro-architecture et donc de fournir une analyse descriptive des images. Leur défaut majeur est qu'elles sont incapables de décrire la topologie de la micro-architecture, ce qui est un vrai frein pour le diagnostic de l'ostéoporose. Dans ce contexte, nous avons opté pour la valorisation de la nature de la structure en forme de réseau de travées. Nous avons pris le partie dans notre étude de résumer la structure par un réseau connecté des squelettes des travées. Pour calculer le squelette, les images ont été numérisées à l'aide d'un scanner Agf à DuoScan, avec une résolution de $100\mu m$. Une région d'intérêt carrée (ROI) de 256×256 pixels, située dans la grande tubérosité du calcanéum, a été préalablement définie par des repères anatomiques [BRL+94]. Pour la pertinence du contenu des images, deux ensembles de travées se trouvant à l'intérieur du ROI ont été repérés : un groupe compressif divergeant vers le bas de l'articulation et un groupe de tension de la partie inférieure de la grande tubérosité balayant vers l'arrière. À partir des images numérisées, nous avons procédé à une première modélisation du signal issu des travées osseuses par l'extraction d'un squelette morphologique du réseau osseux. Le squelette morphologique est le résultat d'un algorithme itératif préservant la connectivité par amincissement. Comme les autres algorithmes d'amincissement utilisés pour les images grises, il se réfère à la définition centrale de pixel simple définie par Rosenfeld dans [Ros79] pour les images binaires qui assure la préservation de la connectivité. Dans [SBCS01], l'adaptation pour les images grises a été utilisée pour accélérer le calcul. Même si la définition de la transformation homotopique telle que définie pour les images en niveaux de gris dans [Soi03] n'est pas strictement respectée, une adaptation de la phase

d'initialisation d'un tel algorithme d'amincissement séquentiel a conduit dans nos travaux à des squelettes présentant de bonnes propriétés de connectivité. Son usage en image est tout simplement lié à la vision de l'image comme un ensemble de points décrits par leur position x, y et leur valeur d'intensité $I(x, y)$. La transposition de l'homotopie en image numérique donne donc lieu à la définition suivante :

Définition 3. *Deux ensembles sont homotopes s'il existe une transformation bicontinue pour passer de l'un à l'autre, telle que :*

— *chaque objet (grain) contient le même nombre de trous que son transformé,*

— *chaque trous (pore) contient le même nombre d'objets que son transformé.*

L'homotopie en image numérique décrit l'organisation des grains et des pores entre eux, c'est-à-dire la préservation de la topologie. Elle dépend de la donnée de la règle de connexité à définir sur les objets. Il existe trois sortes de règles 4-*connexité*, 6-*connexité*, 8-*connexité*.

Définition 4. *Une règle est définie par :*

1. *8-connexité pour l'objet et 4-connexité pour le fond (le vide) ;*
2. *4-connexité pour le fond, 8-connexité pour la forme ;*
3. *6-connexité pour la forme et 6-connexité pour le fond.*

Deux propriétés inhérentes en découlent à savoir :

1. l'invariance par translation et rotation où toute translation d'un point ou rotation de 90° pour 4− et 8−, 60° pour 6−, autour du centre de l'objet, ne modifie pas sa connexité,
2. l'autodualité, c'est-à-dire le nombre d'objets connexes sera différent si la règle de connexité est inversée. L'autodualité est vraie seulement 6−.

Les trois règles de connexités sont illustrées par la figure 2.4.

Le squelette est vu comme une opération homotopique permettant la préservation de la topologie de nos travées. C'est une suite d'opération séquentielle d'amincissement homotopique (réduction de l'épaisseur des objets en préservant leur connectivité) en utilisant des éléments structurants morphologiques spécifiques.

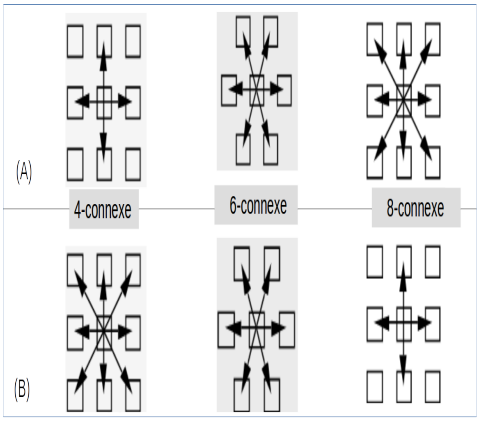


FIGURE 2.4 – Connexités : (A) pour l'objet et (B) pour le fond

Squelettisation par amincissement

L'image I peut être vue comme l'union de deux ensembles $I = I_0 \cup I_1$. Les pixels de l'objet dans l'ensemble I_1, prenant des valeurs de niveau de gris allant de 1 à 255, et les pixels de fond dans l'ensemble I_0 prenant la valeur 0. La séparation entre les pixels de fond et les pixels de l'objet candidat est une tâche de segmentation de l'image, vue comme pré-traitement indispensable pour extraire les objets visuels pertinents à la suite du traitement. Dans ce qui suit, nous supposons que cette étape est déjà élaborée que nous détaillerons dans la section 2.3.

Les pixels objets, candidats à la suppression (destructible), et qui seront déplacés en I_0, sont des pixels objets frontières ou de bord définis comme suit :

$$\delta I_1 = \{x \in I_1, \mathcal{V}_4(x) \cap I_0 \neq \emptyset\}$$

avec $\mathcal{V}_4(x)$ représente l'ensemble des pixels voisins en 4-connexité du pixel x. Il existe plusieurs opérateurs d'amincissement morphologiques dont l'érosion étant l'opération basique en morphologie mathématique. Intuitivement, l'amincissement consiste à effectuer des érosions successives avec un élément structurant symétrique en l'appliquant sur les pixels de bord. L'érosion est une opération morphologique qui peut s'écrire d'une manière algébrique très simple. On désigne $\mathcal{B}$ notre élément structurant défini dans {0,255}.

L'érodé ϵ_B d'un objet $\mathcal{X} \subseteq I$ par $\mathcal{B}$, est :

$$\epsilon_{\mathcal{B}}(\mathcal{X}) = \{x \in I \setminus \mathcal{B}_x \subseteq \mathcal{X}\} \tag{2.1}$$

avec $\mathcal{B}_x$ est le translaté de $\mathcal{B}$ au pixel x. L'érosion a pour effet de réduire l'épaisseur de l'objet et d'accentuer la taille des trous (les trous s'élargissent). Par dualité, on peut obtenir l'objet d'origine à partir de son érodé par l'opération de dilatation. La dilatation morphologique d'un objet $\mathcal{X}$ par l'élément structurant $\mathcal{B}$ est :

$$\Delta_{\mathcal{B}}(\mathcal{X}) = \{x \in I \setminus \check{\mathcal{B}}_x \cap \mathcal{X} \neq \emptyset\} \tag{2.2}$$

où $\check{\mathcal{B}}$ est le symétrique de $\mathcal{B}$ par rapport à l'origine du système et $\check{\mathcal{B}}_x$ son translaté au pixel x. L'amincissement est un cas particulier de l'érosion où l'élément structurant est un pixel translaté sur chaque pixel de bord de l'objet. Un pixel x est destructible s'il est un pixel de bord et si le nombre de composantes connexes $\mathcal{C}_x$ dans $\mathcal{V}_8(x) \cap I_1$ est égal à 1. Autrement dit, x est destructible ssi $x \in \delta I_1$ et $|\mathcal{C}_x| = 1$

Ce nombre $|\mathcal{C}_x|$ est facilement calculable grâce au nombre Yokoi [YTF73] :

$$|\mathcal{C}_x| = \sum_{i=0,2,4,6} \overline{I}(x_i)[1 - \overline{I}(x_{i+1}).\overline{I}(x_{i+2})]$$

Ce processus séquentiel s'arrête à une étape K où aucun pixel peut être supprimé, et aboutie à un squelette θ_K tel que pour tout $x \in I$:

$$\theta_K(x) = \begin{cases} 1 & si \;\; x \in I_1^{(K)} \\ 0 & si \;\; x \in I_0^{(K)} \end{cases} \tag{2.3}$$

Dans [Ser03], il a été montré que le squelette $\theta_K(x)$ est l'amincissement limite quand $k \to \infty$. En pratique, la limite moyenne de K approche l'épaisseur moyenne de l'objet. Bien que le réseau des travées peut être dense, la convergence de l'algorithme reste très raisonnable. Un résultat type du squelette est illustré par la figure 2.5. Nous avons étendu le squelette

$\theta_K(x)$ pour la caractérisation de la topologie des grains et des pores sur des images 3D composées de structures hétérogènes. Contrairement aux images 2D, les (volumes) images 3D contiennent à la fois des structures filaires (des poutres tubulaires plus ou moins fines), des structures surfaciques (des plaques surfaciques), et des structures sphériques (des zones de connexions plus ou moins denses).

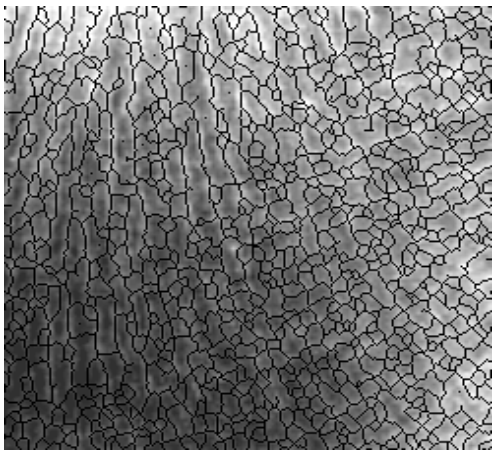


FIGURE 2.5 – Squelette sur la radiographie du calcanéum

Dans le cadre du projet ANR-05-BLAN-0017 « mipomodim« (Milieux Poreux : Modèles, Images.(2006-2009)), pluridisciplinaire fédérant des chercheurs en probabilités, des chercheurs en physique des matériaux et des chercheurs en informatique, nous nous sommes intéressés à mettre au point des méthodes de caractérisation de la structure osseuse en 3D. À partir de 200 échantillons 3D du calcanéum (même site étudié en 2D) prélevés sur des cadavres ne portant pas la maladie, une base d'images 3D reconstruites est constituée. Les échantillons sont des prélèvements cylindriques (carottes) de calcanéum.

La tomographie par rayon X est l'une des techniques permettant d'obtenir de telles reconstructions. Les images obtenues sont de résolutions fines et permettent de caractériser la micro-architecture trabéculaire osseuse à l'échelle des travées osseuses. Le micro scanner dont nous disposons nous a permis d'imager des échantillons cylindriques de $16mm$ de diamètre à une résolution maximale de $18.39\mu m$. La résolution maximale correspond à 1024 pixels répartis sur une distance de $18.83mm$. Cependant, le processus de simulation radiograhique utilisé ainsi que le temps de calcul nécessaire ont limité les reconstructions 3D à des volumes de 256x256x300 pixels (voir figure 2.6).

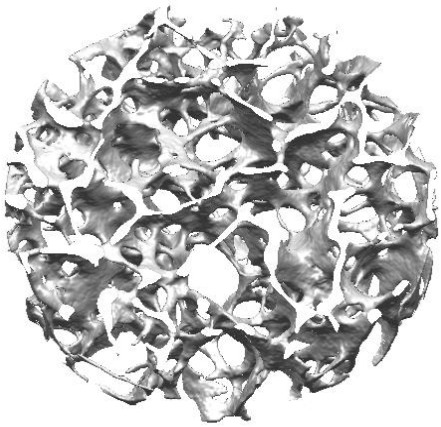


FIGURE 2.6 – Volume reconstruit du calcanéum

À l'issue de cette première étape, nous obtenons les données transformées en objets sémantiques utiles par la suite à l'extraction des descripteurs sémantiques. Grâce à notre algorithme squellétisation multi-échelle permettant d'obtenir des objets sémantiques 2D et 3D connectés formant ainsi un réseau d'objets hétérogènes (des segments, des surfaces, des points). Nous revenons sur le détail de ces objets dans la suite de ce chapitre. L'analyse d'image est une discipline comparable à la fouille des données, la seule différence est le format des données. Comme nous l'avons montré partiellement dans ce paragraphe, il est indispensable de transformer les images, qui sont rappelons-le des données non structurées, en objets sémantiques. Le contenu des images médicales est en général spécifique au domaine où les objets sémantiques à identifier restent homogènes. Mais, plus l'image est multi-domaines, plus il est difficile de les analyser. Typiquement, les images photographiques contiennent des scènes du monde réel et se composent d'objets hétérogènes. Dans la même image, des animaux, des personnes, du sable, de l'eau, des maisons, etc., peuvent être trouvés (voir illustration 2.7). Chacun de ces objets appartient à un domaine différent et décrit par des caractéristiques différentes. Il en résulte que l'analyse d'image photographique est confrontée également au passage à l'échelle, compte-tenu du nombre d'images, de la diversité de leurs formats et de l'hétérogénéité de leur contenu. Le premier défi que nous abordons dans la section suivante porte principalement sur l'extraction des objets sémantiques.

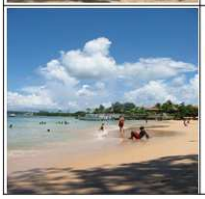
Image	Concepts descriptifs	Conclusion
	(sand, diver, walk, water , beach, boat, sea)	Chaque élément (concept) représente un aspect de l'image (texture, forme, couleur etc.). Il est possible d'appliquer durant la recherche des descripteurs compte tenu du contenu sémantique (les concepts) au lieu d'appliquer un descripteur arbitrairement
	(sand, beach, tree, boat, wood, slope, bay)	
	(people, sand, beach, house, palm, tree)	

FIGURE 2.7 – Diversité des objets dans des images photographiques

Images photographiques

Dans ce contexte d'image réelle, il est supposé que nous disposions seulement d'une représentation visuelle des objets sans connaissance *a priori*. En terme de perception, les images photographiques sont des images aussi complexes que les images médicales puisque leur interprétation est dépendante de la perception visuelle humaine. En effet, la perception du contenu d'une image peut différer d'une personne à l'autre. Dans un contexte où la production des images et leur usage se démocratisent, leur croisement avec d'autres ressources telles que les tweet ou les réseaux sociaux favorise l'enrichissement des informations implicites de la scène. Dans le cadre de la thèse d'Olfa Allani, nous nous sommes particulièrement intéressés à apporter des réponses à deux questions majeures. La première question a porté sur l'extraction dynamique des descripteurs pour des images photographiques afin d'associer des descripteurs sémantiques appropriés et pertinents à ces images toujours dans le but de faciliter la recherche d'information sémantique visuelle. La seconde question concernait plus particulièrement l'enrichissement de l'indexation de ces images et le passage à l'échelle. Les descripteurs sémantiques jouent un rôle double dans un système d'aide à la décision. Le premier rôle est de décrire fidèlement les données sous forme de connaissances permettant de construire un modèle de prédiction ou de classification fiable. Le second rôle est une amélioration de l'indexation des données par leur contenu sémantique afin de pouvoir y accéder rapidement tout en trouvant des informations pertinentes

répondant à la requête émise. C'est de cette façon que les descripteurs ont été pensés dans le cadre de nos travaux, contrairement à la plupart des démarches exposées dans la littérature, où le modèle prédictif et le modèle d'indexation sont souvent décorrélés, en particulier quand les données sont visuelles. Dans ce contexte, les descripteurs visuels sont importants pour la recherche à base de contenu. En effet, ils portent l'information numérique tirée des images et permettant de les comparer, les classifier, les compresser, etc. Une multitude de descripteurs visuels ont fait leur apparition depuis l'émergence du traitement et de l'analyse d'image. Certains ont été particulièrement conçus à des fins de recherche [KS04] [LW04] [NMMB98] mais qui, finalement, ont été écartés faute de performance [Vin11]. Les limitations majeures de ces travaux se résument, d'une part, à une élaboration heuristique des descripteurs qui se déclinent en descripteurs locaux et globaux ; et, d'autre part, à considérer une liste de descripteurs prédéfinis, de surcroît décorrélés du contenu et des caractéristiques prépondérantes des images. Outre ces caractéristiques artisanales, des approches par apprentissage ont été largement développées en s'appuyant sur des banques d'images parfaitement segmentées et labellisées. Bien que ces approches soient très efficaces pour la recherche textuelle où la requête est constituée d'un ensemble de mots-clés ou sous forme de phrase, elles ne permettent pas d'outrepasser la recherche par le contenu où la requête est une image sans labels. Dans cette optique, nous avons proposé une approche originale fondée sur la recherche de motifs de descripteurs à partir des travaux de la communauté via des règles d'association. Le processus de construction des règles d'association se décompose en quatre étapes illustré par la figure 2.8.

Partant de deux hypothèses de travail, une cinquantaine d'articles de recherche a été retenue constituant ainsi notre corpus. La première hypothèse a permis de limiter les articles de la littérature à ceux qui ont traité les descripteurs locaux et globaux pour divers domaines et ont proposé leur implémentation sous accès libre. Avec la seconde hypothèse, nous avons restreint les travaux de la littérature à ceux qui utilisent des bases d'images de référence (*benchmark datasets*) avec une attention particulière aux critères d'évaluation utilisés. Le corpus d'articles a été analysé semi-automatiquement afin de construire une table de transactions. Une transaction T est de la forme : $< transaction_{id}, concept, descripteur >$, avec $concept \in Concepts$, $descripteur \in Descripteurs$. Une fois la génération de toutes les

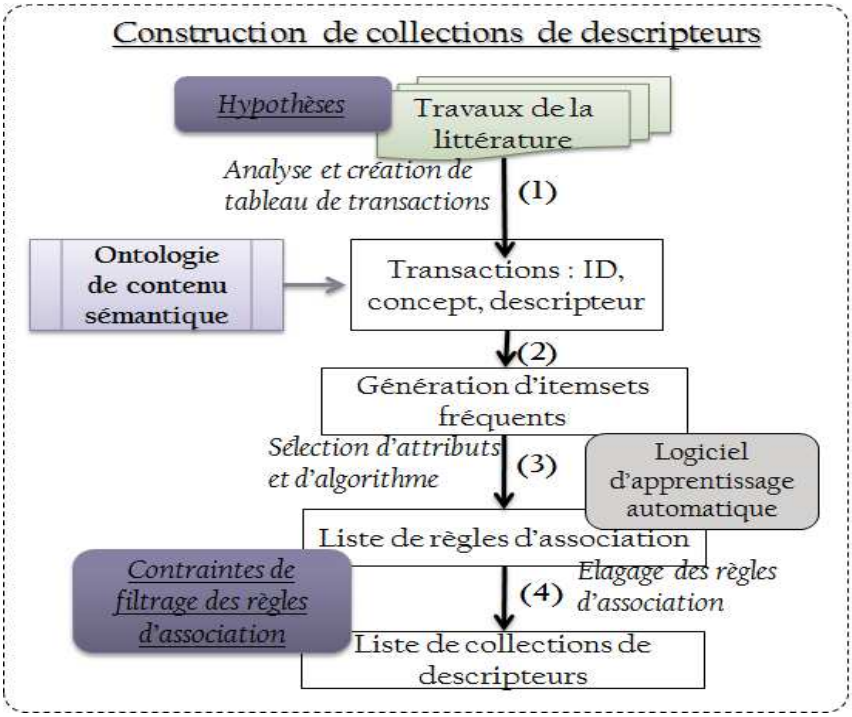


FIGURE 2.8 – Étapes de construction des règles d'association

transactions réalisées, ces dernières sont transformées pour qu'elles soient exploitées par un outil d'apprentissage automatique. L'apprentissage des règles d'association repose sur l'idée d'extraire tous les ensembles d'items (on appelle $Itemset$ un ensemble d'$Items$) fréquents puis d'en déduire les règles d'association. Nous cherchons à déterminer les règles de type $concept \rightarrow descripteur$ pour tout $concept \in Items \backslash Descripteurs$, tout $descripteur \in Items \backslash Concepts$ et $Concepts \cap Descripteurs = \emptyset$. La collection de descripteurs notée $\mathcal{CD}$, déduite de ces règles, est composée de $\mathcal{CD} = \{concept_i \rightarrow descripteur_j \vee descripteur_k\}$. Les règles d'association peuvent être évaluées au-delà des mesures classiques de support de confiance par des critères exprimant leur intérêt à savoir la mesure d'intérêt qui caractérise une forme de causalité ; c'est-à-dire la connaissance de l'antécédent apporte de l'information supplémentaire sur la connaissance du conséquent. Il s'agit de la mesure LIFT [KK06]. C'est un rapport de probabilité calculé pour filtrer les règles obtenues par le processus d'apprentissage. Par ailleurs, comme nous pouvons le constater sur la figure 2.8, l'étape (1) de collecte des descripteurs et des concepts nécessite d'avoir recours à une ressource extérieure (ou plusieurs) afin de couvrir tous les concepts et les domaines associés. Étant donnée la difficulté de cerner les concepts d'un domaine par une seule liste et

de conserver les relations de synonymie ou d'hyperonymie entre eux, il est nécessaire de recourir à un modèle permettant la représentation de la connaissance de manière à garder les relations et garantir leur extension. Pour ce faire, nous avons opté pour l'utilisation d'une ontologie de domaine inspirée des travaux de [EHG+10]. Elle servira à la modélisation de l'information sémantique et au filtrage des règles. Un extrait de cette ontologie est présenté dans la figure 2.9.

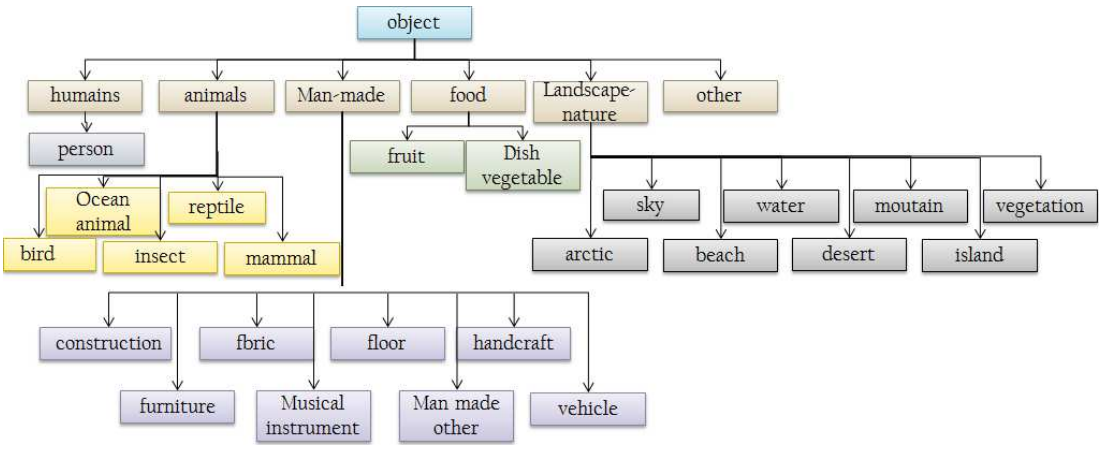


FIGURE 2.9 – Extrait de l'ontologie de domaine

Voici un exemple d'usage de l'ontologie en cas d'ambiguïté des règles. Supposons que nous disposions d'une règle concernant *« castle »* et d'une règle concernant *« building »*. Parallèlement nous savons que le concept *« Tower »* est une partie *« part of »* du concept *« Castle »* qui est de type *« is a »* du concept *« building »*. Grâce à l'ontologie, nous dégageons désormais le concept le plus proche de *« tower »* via la distance sémantique $d(\text{tower, building}) = 2$ et $d(\text{tower, castle}) = 1$. Par conséquent la règle sur *« castle »* sera retenue au final puisqu'elle est générique. Enfin, cette phase de construction de collection de descripteurs par des règles d'association (*cf.* l'extrait de ces règles en figure 2.10) reliant un concept à des descripteurs a pu dégager 117 règles couvrant 117 concepts présents dans les 1871 transactions. Les transactions ont été construites à partir de 50 articles retenus selon différents critères que nous avons discutés dans [AMBA17] [AMB+15].

Les règles obtenues ont été par la suite filtrées automatiquement conformément aux contraintes suivantes :

1. l'agrégation des règles d'association à l'aide de relations évaluées en fonction du de-

```
wildcats ==> LocalBinaryPattern
winter ==> LocalBinaryPattern
streetview  ==>   OpponentSIFT
3Dobjects  ==> SIFT
tomato ===> SURF
flowers  ==> DominantColor
logo  ==> SIFT
plant  ==>  GaborFeature
home   ==>   ReducedSIFT
indoorscene  ==>   ReducedSIFT
university  ==>  ReducedSIFT
leaf  ==> ShapeContext
buses  ==> DominantColor
dinosaurs  ==> DominantColor
mountains 8 ==> DominantColor
```

FIGURE 2.10 – Extrait des règles d'association

gré de préférence d'une règle par rapport à une autre. C'est une mesure d'agrégation par préférence comparée. En effet, de nombreuses mesures ont été définies afin de pouvoir classer les règles dites intéressantes. Très hétérogènes, elles produisent des classements forts variés. C'est pourquoi, plutôt que de privilégier une mesure, il nous paraît intéressant de tenir compte des différentes informations apportées par l'ensemble des mesures. Elles permettent d'une part de retranscrire la nature numérique des mesures, et d'autre part, de réduire les problèmes de non comparabilité entre elles ;

2. les valeurs de confiance sont considérées comme étant des coefficients de priorité et permettent de garder une trace de l'importance de la transaction afin de pourvoir filtrer en cas de conflit entre les règles d'associations obtenues. Nous utilisons un système fondé sur la logique floue qui consiste en l'utilisation d'une t-conorme ;
3. les relations hiérarchiques définies au niveau de l'ontologie de domaine sont exploitées en vue de générer les règles pour tous les concepts ;
4. toute règle comportant plusieurs concepts est dupliquée.

La collection créée $\mathcal{CD}_{finale}$ est certes dépendante des travaux de la littérature choisis, mais

néanmoins elle présente plusieurs avantages. Tout d'abord, elle permet un premier transfert entre articles scientifiques et choix des descripteurs en fonction du domaine abordé. Ensuite, ce transfert pourra évoluer de manière continue en assurant l'enrichissement des règles d'association et, par conséquent, d'affiner le processus de choix des descripteurs.

2.3 Vers une approche de segmentation sémantique d'images

La segmentation d'image peut être définie comme une technique de traitement d'image spécifique utilisée pour séparer une ou plusieurs régions significatives. La segmentation d'image peut également être considérée comme un processus de définition des bords entre des entités sémantiques distinctes dans une image. D'un point de vue plus technique, la segmentation d'image est un processus d'attribution d'un label à chaque pixel de l'image de sorte que les pixels ayant le même label soient reliés par rapport à une propriété visuelle ou sémantique. La segmentation d'image englobe une grande classe de problèmes finement reliés en vision par ordinateur. La version la plus classique est la segmentation sémantique [GOO+17]. Dans la segmentation sémantique, chaque pixel est classé dans l'un des ensembles prédéfinis de classes de sorte que les pixels appartenant à la même classe auront la même entité sémantique dans l'image. Il est également intéressant de noter que la sémantique dépend non seulement des données mais aussi du problème visé. Par exemple, pour un système de détection des piétons, l'ensemble du corps d'une personne devrait appartenir au même segment. Cependant pour un système de reconnaissance d'action, il peut être nécessaire de segmenter les différentes parties du corps en différentes classes. D'autres formes de segmentation d'image peuvent se concentrer sur l'objet le plus important d'une scène. D'autres sur la morphologie des objets pertinents à détecter [PCT+15] [KGPP14]. Une catégorie particulière de ces problèmes, appelée détection des points saillants [Bor15], fournit des algorithmes pour la segmentation des objets à partir de régions d'intérêts automatiquement détectées. D'autres variantes de ce domaine peuvent être des problèmes de séparation entre l'objet et le fond. Dans de nombreux systèmes comme la recherche d'image ou la réponse visuelle à des questions (*Visual Query Answering*), il est souvent

nécessaire de compter le nombre d'objets. La segmentation par instance spécifique permet de résoudre ce problème. En effet, la segmentation par instance est souvent associée à des systèmes de détection d'objet pour séparer plusieurs instances d'un même objet [DHS16] dans une scène. Dans le domaine de la segmentation avec un niveau sémantique faible, la sur-segmentation est également une approche courante où les images sont divisées en régions extrêmement petites pour assurer l'adhérence des limites, au prix de la création d'un grand nombre de faux bords. Les algorithmes de sur-segmentation sont alors combinés avec des techniques de fusion de régions pour élaguer les faux segments dans les images. Notre approche de segmentation, inspirée de ces différentes méthodes a pour but de proposer une implémentation distribuée et parallèle de l'algorithme de segmentation en exploitant une architecture dédiée sous Spark. Dans notre contexte, nous supposons disposer de domaines distincts d'images (médicales, naturelles, satellitaires, Lidars, etc.). Pour ce faire, notre démarche se base sur quelques hypothèses de travail comme suit. (*h1*) La moyenne des tailles des objets d'intérêts présents dans les images est déduite à partir de la résolution des images et leur outils d'acquisition. (*h2*) Le domaine de l'image est défini par l'utilisateur en tant que méta-donnée. En particulier, les labels des objets sont prédéfinis pour les images médicales et les images satellitaires. Par exemple, dans le cas de nos images médicales, nous disposons à ce stade de deux labels $objet = I_1$ et $fond = I_0$. Alors que les images satellitaires contiennent plusieurs objets dont le nombre de labels dépend de la résolution des images. En effet, les images à trois bandes contiennent peu de labels tels que *la végétation, les routes, les bâtiments, l'eau, désert*. Plus le nombre de bandes est élevé, plus il y a de labels à distinguer dans l'image. Quant aux images spectrales et hyperspectrales, le nombre de labels peut dépasser plus d'une trentaine. Face à cette diversité de contenu liée aux images que nous traitons, notre approche propose une sur-segmentation de l'image et combine un algorithme de fusion spatial et sémantique guidé par des connaissances sur le domaine. L'algorithme de sur-segmentation est basé sur une approche hybride de type *« split and merge »* avec un critère d'homogénéité combinée avec une approche de segmentation par ligne de partage des eaux (*watershed*). L'algorithme de *« split and merge »* peut être vu comme une méthode de segmentation grossière des objets visuels candidats. Il permet d'identifier d'abord les objets visuels à forte densité spatiale (gros objets). Cette segmentation est combinée avec l'algorithme *watershed* qui

permet de distinguer les objets granulaires (petits objets). L'algorithme résultant utilise par conséquent trois critères : (a) l'homogénéité, (b) la taille des objets et (c) le domaine. L'algorithme que nous proposons est de nature récursive nécessitant ainsi une structure dédiée de stockage des segments intermédiaires. Cette structure doit rendre compte des relations de voisinage mais aussi des dépendances hiérarchiques entre segments. Pour cela, nous avons choisi une structure de type *quadtree*. En effet, il s'agit d'une structure courante pour des problèmes de segmentation utilisant une approche *« split and merge »*.

2.3.1 Approche *« split and merge »*

Un algorithme de type *« split and merge »* consiste à diviser une image en régions plus petites (*split*), que l'on considère comme étant homogènes. Elles sont par la suite regroupées selon leur proximité. Dans le cas du *split*, nous considérons deux critères d'homogénéité. Un critère global de l'image défini par sa variance Var et un critère local défini par l'entropie de Shannon H associée au segment traité. On utilisera les relations suivantes :

$$Var = \sum_{dimensionsImage} (valeurPixel[i][j] - moyImage)^2$$

Nous considérerons, dans un second temps, l'entropie de Shannon associée au segment considéré. On considérera la relation suivante :

$$H = -\sum_{dimensionsImage} (p_k * log(p_k), \text{ avec } p_k \text{ probabilité d'avoir un pixel de valeur } k.$$

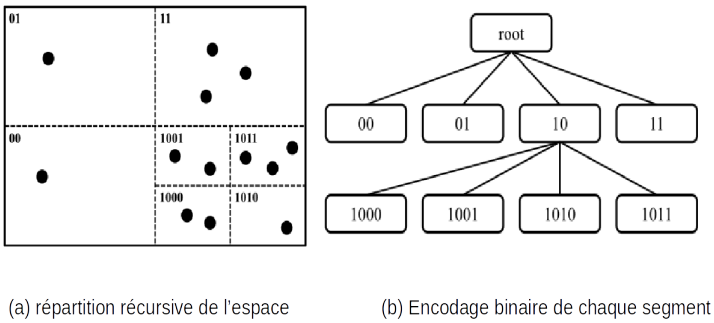


FIGURE 2.11 – Exemple d'un arbre quadtree

subsubsection*Split Il sera procéder alors par récurrence afin de diviser chaque segments en quatre qui seront enregistrés dans notre arbre (*cf.* figure 2.11). Les quatre segments pourront être de dimensions identiques ou bien présenter un *offset*. En effet, l'ajout d'un offset

permet de tenir compte de la présence d'un recouvrement d'un segment avec ses voisins ce qui permet d'améliorer la segmentation. L'offset devant diminuer à chaque récursion, il a été envisagé deux approches. La première consiste à prendre un même pourcentage de chaque dimension du segment à diviser. La seconde consiste, quant à elle, à diminuer de manière exponentielle un offset initial à chaque récursion. Enfin, l'algorithme de division s'arrête lorsque la dimension du segment est inférieure à un seuil prédéfini $seuil_{split(t_D)}$ en fonction du domaine. En effet, s'il s'agit d'images médicales ou d'images satellitaires, le seuil est déduit à partir de la résolution des images et correspond au label associé aux objets dont la moyenne des tailles est la plus petite. En revanche, dans le cas des images réelles, ce seuil est fixé arbitrairement. On considérera ici que le segment n'est plus intéressant dès que l'une de ses dimensions est strictement inférieure à $seuil_{split}$. L'algorithme devra également s'arrêter si les segments fils du segment considéré sont tous homogènes, c'est-à-dire de variance ou d'entropie inférieure à un seuil que l'on appellera $seuil_{split}(h)$. Le pseudo-code décrivant l'algorithme de la phase de division est présenté à Algorithme 1.

Algorithme 1: Algorithme *« split »*

Data : arbre, noeudCourant, $seuil_{split}(t_D)$, $seuil_{split}(h)$, offsetPrecedent, CritereSplit
Result : arbre complété contenant les couches de la segmentation

```
1 begin
2     while critereSeuilSplit(noeudCourant)=( h(noeudCourant)> seuil_split(h) (n'est
      pas homogène) ou de dimension supérieure à seuil_split(t_D)) do
3         offsetCourant ⟵ diminuer offsetPrecedent;
4         noeuds ⟵ diviser noeudCourant en 4 en tenant compte de
          offsetCourant;
5         ajouter les noeuds à l'arbre avec noeudCourant pour parent;
6         for noeudFils ∈ noeuds do
7             if valeur du Val(CritereSplit(NoeudFils)) == val(critereSeuilSplit) then
8                 Split(arbre, noeudfils, SeuilSplit, offset, CritereSplit);
```

Merge

Concernant la phase de fusion (*merge*), elle consiste à regrouper les segments de la dernière couche de l'arbre avec leurs voisins. Ces segments seront regroupés selon leur homogénéité,

c'est-à-dire de la valeur de leur variance/entropie. Ainsi, on pourra construire une enveloppe séparant chaque élément de l'image. Le pourcentage d'écart relatif entre chaque critère est utilisé pour assumer la fusion locale. Deux approches candidates sont alors applicables. La première consiste à utiliser un graphe d'adjacence (dit RAG pour *Region Adjacency Graph*). Le graphe sera construit récursivement en ajoutant, à chaque fois, les fils de chaque sommet du graphe, les arêtes reliant les sommets voisins l'un de l'autre et ayant pour valeur l'écart d'homogénéité entre les deux segments. Enfin, l'algorithme parcourt le graphe, et pour chaque sommet, il regarde s'il existe des arêtes inférieures à un certains seuil et regroupe les deux sommets (voir Algorithme 2). L'algorithme procède, ensuite, récursivement jusqu'à n'avoir que des voisins d'écart supérieur à notre seuil. Ce seuil est appelé $seuil_{Merge}$. Ainsi, la différence d'homogénéité entre deux segments $s1$ et $s2$ est calculée entre deux segments par :

$$h_{diff} = (n_1 + n_2)seuil_{Merge} - (n_1 h_1 + n_2 h_2) = n_1(seuil_{Merge} - h_1) - n_2(seuil_{Merge} - h_2)$$

avec n_1, n_2 sont respectivement le nombre de pixels dans chacun des segments $s1$ et $s2$. Cette homogénéité peut être très facilement étendue à différents canaux d'une image ayant N canaux, comme c'est souvent le cas pour les images satellitaires, par :

$$h_{diff} = \sum_{i}^{N} w_i[n_{1,i}(seuil_{Merge,i} - h_{1,i}) - n_{2,i}(seuil_{Merge,i} - h_{2,i})]$$

où w_i est le poids d'importance du $i^{ème}$ canal.

L'autre approche consiste à utiliser une matrice de même dimension que l'image à segmenter. Cette matrice est complétée, ensuite, en utilisant la dernière couche de l'arbre issue de l'algorithme de *split* en assignant à chaque position (i, j) le noeud correspondant. Ainsi, il suffira de parcourir récursivement la matrice et de regrouper dans une même liste les voisins dont l'écart est inférieur à $seuil_{Merge}$. Le pseudo-code est décrit dans l'algorithme 3 de fusion dans le cas de l'utilisation d'une matrice « de voisinage ».

2.3.2 Apport du *watershed*

Notre méthode de segmentation nommée SM de type *« split and merge »* présentée ici est orientée objet. Chaque décision de fusion est basée sur les attributs récents des objets

Algorithme 2: Algorithme de création du *Rag*

```
Data : arbre résultant du Split
Result : le RAG correspondant à l'arbre
1  begin
2      graphe ⟵ nouveauGraphe(arbre.racine);
3      nbreSommetsSuppr ⟵ 0;
4      while nombre de sommets du graphe + nombre de sommets supprimés <
       nombre de noeuds de l'arbre do
5          for Sommet ∈ graphe do
6              if la valeur du Sommet (= un noeud) a des fils then
7                  ajouter 4 sommets contenant chacun un fils au graphe;
8                  for Arete ∈ Sommet.aretes do
9                      déconnecter Sommet et le Voisin correspondant à Arete ;
10                     connecter les fils au Voisin s'ils sont en contact;
11                 ajouter 1 à nbreSommetsSuppr;
```

Algorithme 3: Traitement par Merge

```
Data : arbre résultant du split
Result : Liste contenant des listes de noeuds regroupés selon leur écart
         d'homogénéité
1  begin
2      matriceVoisinage ⟵ creerMatriceVoisinage(arbre);
3      listeNoeudsLibres ⟵ derniereCouche(arbre);
4      resultat ⟵ creerListeVide();
5      while il reste des noeuds dans listeNoeudsLibres do
6          regroupement ⟵ creerListeVide();
7          noeudCourant ⟵ listeNoeudsLibres[0];
8          enlever noeudCourant de listeNoeudsLibres;
9          ajouter noeudCourant à regroupement;
10         parcourir tous les voisins de noeudCourant grâce à matriceVoisinage et les
           ajouter à regroupement;
11         en faire de même pour chaque voisin ajouté à regroupement;
12         ajouter regroupement à resultat;
```

fusionnés lors des étapes précédentes. Cela s'avère être un avantage méthodologique décisif de la méthode. En effet, chaque décision est basée sur les attributs de structures homogènes d'une échelle locale. Ce qui fait de cette méthode une segmentation multi-échelle et adaptative à la résolution de l'image. Malheureusement, les algorithmes de type *split and merge* sont très peu efficaces pour différencier des objets de granularité fine et de forme

linéaire tels que les cours d'eau. Comme discuté précédemment, ils sont efficaces pour extraire des segments à forte densité spatiale puisque l'homogénéité est pondérée par le nombre de pixels de l'objet. Pour lever cet inconvénient, nous avons amélioré l'algorithme en intégrant une approche de filtrage de régions à base de *watershed*. En effet, cette technique, couplée à une approche *« split and merge »*, pourrait apporter de meilleurs résultats et notamment pour les structures filaires, linéaires telles que les routes ou les rivières. Elle consiste à utiliser des marqueurs assimilant certaines régions à des pics (de fort gradient) et d'autres à des vallées (zones de faible gradient). Les vallées seront ensuite *inondées* avec des eaux de couleurs différentes. Lorsque deux bassins de couleurs d'eau différentes se rencontrent, une frontière est alors construite ; et cela, jusqu'à ce que tous les pics soient inondés. Les frontières construites constituent la séparation des objets de l'image. La mise en place technique de ce nouvel algorithme (nommée SWM) est comme suit. Tout d'abord, les seuils nécessaires pour réaliser une première segmentation grossière sont construits. Ensuite, le bruit présent sur l'image est supprimé. L'arrière-plan et les parties de l'image dont on est certain qu'elles sont à l'intérieur d'un élément à segmenter sont distingués. Ainsi, par différence, les zones dans lesquelles se situent les frontières, aussi appelées zones indéterminées, sont obtenues. Nous pourrons alors marquer l'arrière-plan, les zones intérieures aux objets à segmenter et les régions indéterminées. L'algorithme de *watershed* utilisera ces marqueurs pour délimiter les différents objets les uns des autres et de l'arrière-plan selon le principe précisé précédemment. Notre approche de *« split and merge »* complétée par l'algorithme de *watershed* a été appliqué en sortie de phase de *split*, sur chaque segment de la dernière couche de l'arbre. En effet, ce segment est supposé homogène, mais si avec un critère peu sélectif, nous obtiendrons dans notre segment un objet (ou une partie de cet objet) ainsi qu'une partie de l'arrière plan. L'ajout du watershed permet à la fois de séparer cet objet de l'arrière plan, mais aussi de gagner en rapidité d'exécution car il a besoin d'un maillage moins fin pour obtenir une segmentation précise suivant les contours des éléments de l'image (en particulier les cours d'eau). Après la phase de *merge*, il suffit de regrouper les régions dans une matrice de taille la dimension du groupe. Par dilatation des contours, nous assurons la continuité d'un segment à un autre du groupe. Enfin, un algorithme de remplissage de trous pour construire un cache permet de sauvegarder l'élément segmenté. La figure 2.12 illustre les différentes étapes de

notre algorithme de segmentation.

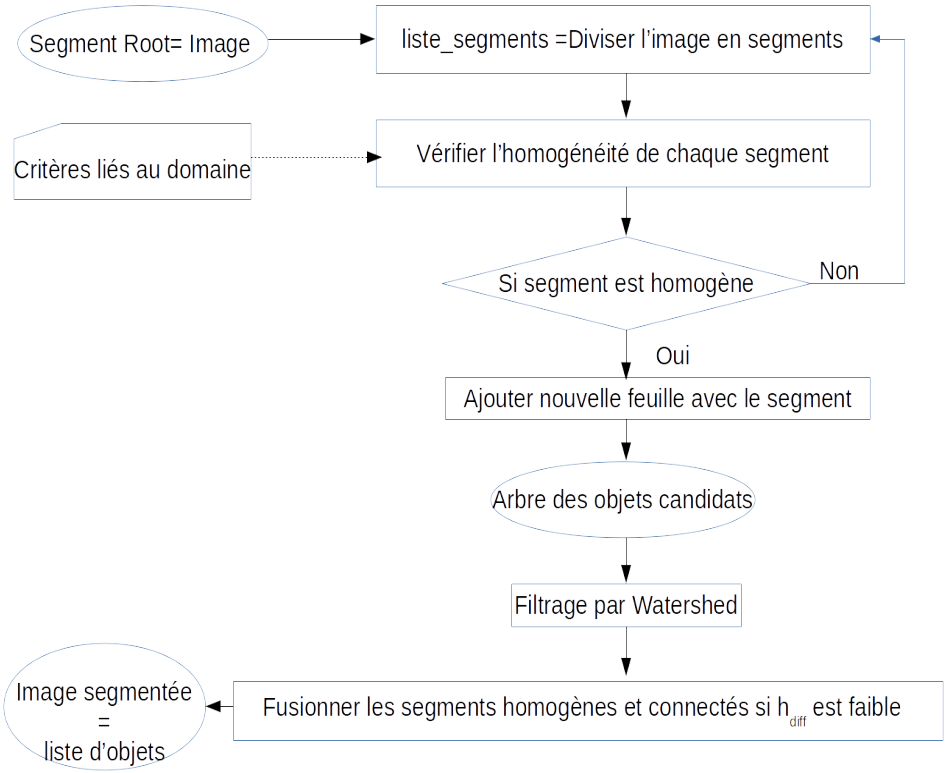


FIGURE 2.12 – Processus de l'algorithme de segmentation SWM

2.3.3 Distribution avec MapReduce

MapReduce est un modèle de traitement distribué doté d'un environnement d'exécution parallèle dans un graphe de machines indépendantes et distribuées. *Hadoop* [VMD+13] est implémenté dans la communauté du libre et supporte l'infrastructure MapReduce. *Hadoop Distributed File System* (HDFS), est un système de fichier, de stockage distribué émanant de Hadoop Apache qui a été conçu pour prendre en charge de gros fichiers. En utilisant HDFS, un fichier de grande taille est initialement partitionné en plusieurs fragments, appelés *chunk*, et sont stockés dans plusieurs machines de manière redondante pour plus de fiabilité. La taille d'un fragment est généralement de 64 Mo. En plus des opérations coûteuses en temps et en ressources, s'ajoute la contrainte de traitement en temps réel que Hadoop ne permet pas de la lever. *Spark* appartient à une nouvelle génération

d'informatique distribuée [ZCF+10], il est devenu un projet Apache en 2013. Spark est compatible avec les modules Hadoop, tels que *YARN* et *HDFS*, mais il dispose également d'un mode autonome. La principale motivation du projet Spark était de renforcer l'exécution de charges de travail itératives par le biais de la mémoire de calculs (in-memory). En raison des nombreux accès au disque nécessaires pour traiter une demande, Hadoop est assez inefficace pour de telles charges de travail [Pat19]. Ces calculs en mémoire reposent sur une structure de données appelée *« Resilient Distributed Dataset »* (RDD) qui sont des ensembles d'éléments tolérants aux pannes pouvant être exploités en parallèle et qui peuvent être utilisés pour mettre en cache un ensemble de données pendant les opérations. Les RDD peuvent faire référence à un ensemble de données dans un système de stockage externe, tel qu'un HDFS, et peuvent être créés à partir de n'importe quelle source de stockage supportée par Hadoop. Les RDD sont calculées en cas de besoin en ce sens que la séquence de transformations qui y est effectuée ne sera traitée que lorsque les données associées devront être collectées. De plus, si une partition de données d'un RDD est perdue à cause d'erreurs physiques, le cache de Spark peut être automatiquement recalculé en exécutant à nouveau sa séquence de transformations respective [2]. Entretemps, des alternatives au modèle MapReduce ont été proposé, en particulier le calcul distribué Apache Spark [3] qui a attiré beaucoup d'attention ces dernières années surtout en raison de sa capacité à surpasser Hadoop dans de nombreuses applications. Ceci est dû à sa capacité de partager la mémoire entre les noeuds d'un cluster au lieu de restreindre la communication à l'accès aux fichiers de données comme c'est le cas avec Hadoop et dans des cadres distribués alternatifs tels que Apache Tez [SSS+15]. Afin de paralléliser l'algorithme (nommé $SWMP$), nous avons fait le choix de commencer par la phase de division et donc nous nous intéressons pour l'instant à l'algorithme 1 Pour cela, nous disposons d'un cluster Spark avec un manager YARN composé d'une machine maître et 3 machines esclaves. La machine maître est équipée d'un processeur Intel Xeon E3-1220 de 3,1 GHz, mémoire de 16 Go et disque dur de 500 Go. Chaque esclave est équipé d'un processeur Intel Core i5 à 3,2 GHz, d'une mémoire de 4 Go et d'un disque dur de 500 Go. Toutes les machines

2. https://www.datamation.com/data-center/hadoop-vs.-spark-the-new-age-of-big-data.html

3. Apache Spark Development Team, 2019. Apache spark. Apache Software Foundation. spark.apache.org (15 February 2019)

sont connectées par un commutateur Ethernet 1 Gbps. Chaque machine fonctionne sous Linux (Ubuntu 10.04 Lucid). Nous avons utilisé Hadoop 2.0.0 pour l'implémentation de l'infrastructure MapReduce obtenu à partir du site Apache [4]. Fondamentalement, MapReduce se compose d'un *job tracker* et de plusieurs *task trackers*. Chaque *task tracker* est exécuté sur une machine esclave. Un esclave traite les données à l'aide d'une fonction *map* et d'une fonction *reduce*, chacune d'entre elles étant activée par *job tracker*. Ce dernier fonctionnant seulement sur la machine maître, prend en charge la détection des erreurs, l'équilibrage des charges, l'occupation des processeurs, etc. Le fonctionnement de *map* et *reduce* se résume par les deux instructions suivantes :

(1) $map(key1, value1) \rightarrow list((key2, value2))$

(2) $reduce(key2, list(value2)) \rightarrow (key3, list(values3))$

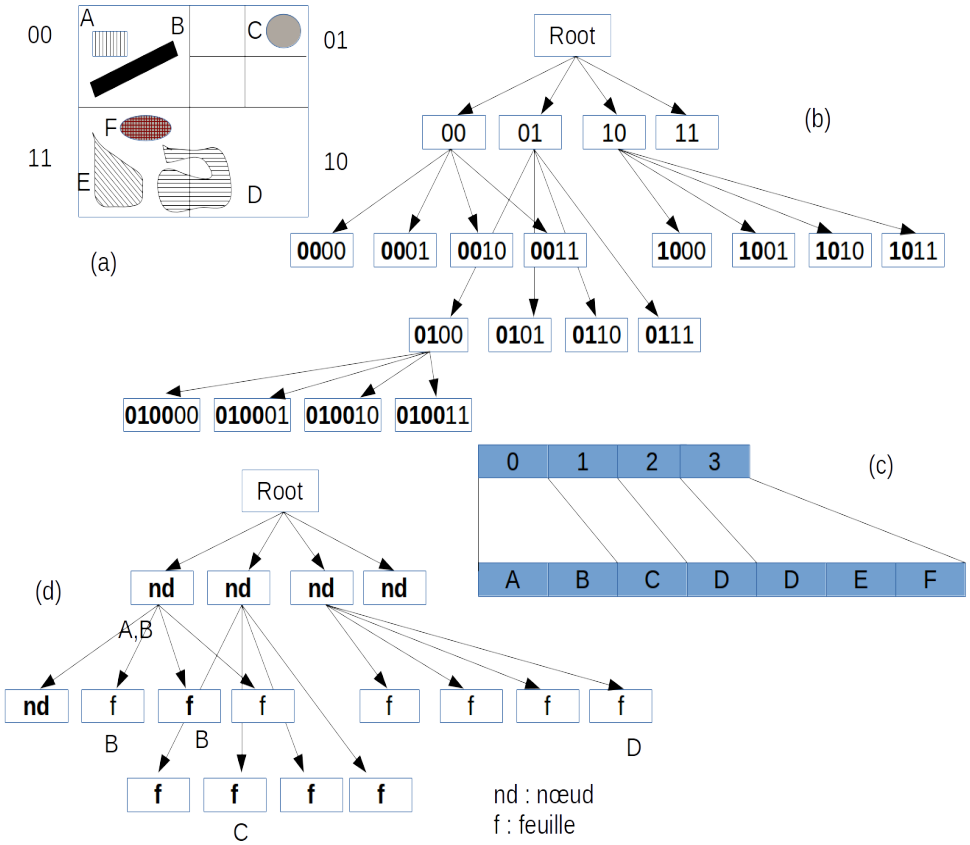


FIGURE 2.13 – (a) Image à segmenter ;(b) représentation encodée des quadtrees ; (c) représentation logique ; (d) fonction map() dans MapReduce

La répartition des segments d'une image dépend de la taille d'un bloc de données dans

4. http://hadoop.apache.org

HDFS. Si la taille du bloc est $size_b$ et qu'un segment fait $size_s$ alors le segment est divisé en nb blocs tel que $nb(s) = \frac{size_s}{size_b}$ où chaque bloc est pris en charge par la fonction $map()$ de MapReduce (figure 2.13). Chaque fonction $map()$ appelle l'algorithme de *split* (*cf.* Algorithme 1) et sépare les objets dans l'un des segments. Chaque segment constitue un arbre local ayant une liste de clés et dont les valeurs sont le contenu des feuilles enregistrées dans des fichiers. Afin de reconstituer l'arbre global, la fonction $reduce()$ trie les clés et connectent les noeuds intermédiaires appartenant au même noeud père. Des premières comparaisons sont effectuées sur des images satellitaires de dimension $3,393px$Ö$3,349px$ et de taille $4092,39KB$ (4 190 610 bytes) (*cf.* figures 2.14 et 2.15). Ces images proviennent d'une collection fournie par Kaggle dans le cadre du challenge *« Dstl Satellite Imagery Feature Detection »*. Le temps moyen d'exécution de l'algorithme de division sans parallélisation sur une image est de 6 minutes, sachant que le temps d'exécution global moyen est de 12 minutes pour obtenir plus de 40000 objets. Avec la parallélisation sur trois machines, nous obtenons un temps moyen d'exécution de 21 secondes. Ce gain est très prometteur pour explorer plus en avant la parallélisation de la totalité de l'algorithme de segmentation. Toutefois, cette étape de segmentation est très utile en tant que pré-traitement pour une tâche de reconnaissance des objets et d'étiquetage des objets visuels spatiaux. En effet, cette étape de pré-traitement permet de cerner les objets cibles pour leur caractérisation sémantique.

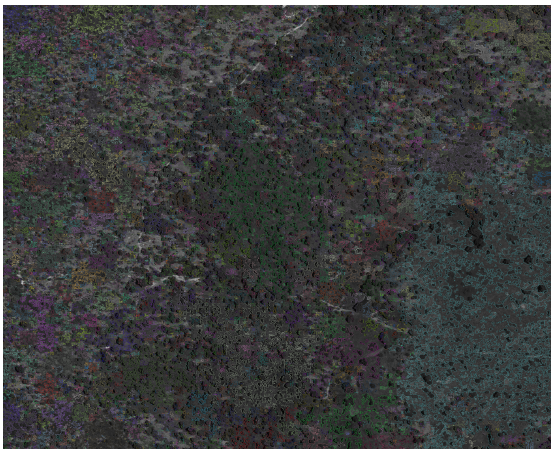


FIGURE 2.14 – Meilleure segmentation obtenue pour le type « forêt »

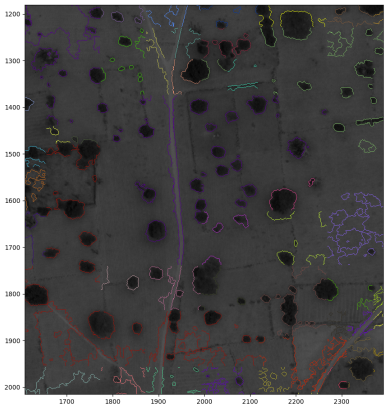


FIGURE 2.15 – Meilleure segmentation obtenue pour les routes et les forêts

2.3.4 Vers une segmentation sémantique

La segmentation des données est une étape indispensable pour toute application décisionnelle. Elle permet de séparer les données pertinentes du bruit. Le type de bruit est très dépendant de la source d'acquisition des données. La segmentation des objets utiles devra s'appuyer sur des critères élaborés en fonction du domaine. Par exemple, s'il s'agit de données de capteurs, les critères d'homogénéité seront formulés, d'une part, par rapport à la cohérence des valeurs avec les valeurs tolérées par capteur, et, d'autre part, par rapport à leur homogénéité temporelle. Les neurosciences ont fait beaucoup de progrès dans la compréhension du système visuel et de la façon dont les images sont transmises au cerveau. On pense que le filtre de différence gaussien (DoG) simule la façon dont la rétine humaine traite les images observées et en extrait les détail. Parmi ces types de filtre, l'approche WLD (*Weber Local Descriptor*) basée sur le fait que la perception humaine d'un modèle dépend non seulement de la modification d'un stimulus (comme le son, l'éclairage) mais aussi de l'intensité initiale du stimulus. Le WLD se compose de deux éléments à savoir l'excitation différentielle et l'orientation. Une autre méthode pour améliorer la force du descripteur de texture consiste à effectuer un pré-traitement efficace. Ci-dessous un aperçu rapide du point de vue des neurosciences [CGTon] :

— les photo-récepteurs composés de bâtonnets et de cônes ont des propriétés très différentes : les bâtonnets ont la capacité de voir la nuit, dans des conditions de très faible éclairage ; les cônes, quant à eux, ont la capacité de traiter des signaux lumineux.

La couche photo-réceptrice joue donc le rôle d'un filtre d'adaptation à la lumière.

— la couche plexiforme externe (PE) : le photo-récepteur joue le rôle d'un filtre passe-bas. Les cellules horizontales effectuent un second filtre passe-bas. Dans le PE, les cellules bipolaires calculent la différence entre le photorécepteur et la réponse des cellules horizontales (bruit lié à l'éclairage à basse fréquence). Généralement, pour modéliser les processus du PE, on utilise deux filtres passe-bas Gaussien avec des écarts-types différents correspondant aux effets des photo-récepteurs et des cellules horizontales. Ainsi, les cellules bipolaires agissent comme un filtre DoG.

— la couche plexiforme interne (PI) : le PI fonctionne de manière similaire au PE mais agit sur l'information temporelle plutôt que sur l'information spatiale comme c'est le cas du PE. Il calcule la deuxième dérivée spatiale d'une image. Dans les zones où l'image a une intensité constante, la réponse du filtre sera nulle (régions homogènes). Partout où un changement d'intensité se produit, le filtre donnera une réponse positive sur le côté sombre et une réponse négative sur le côté clair. À une limite raisonnablement nette entre deux régions d'intensités uniformes mais différentes, la réponse du filtre sera : (i) zéro à une grande distance des bords ; (ii) positive d'un côté du bord ; (iii) négative de l'autre côté du bord ; (iv) zéro sur le bord lui-même.

Ainsi, PE et PI peuvent être modélisés par deux images filtrées que nous notons I_{bf}^{+} et I_{bf}^{-} tels que :

$$I_{bf}^{+}(p) = \begin{cases} I_{bf}(p) & \text{si } I_{bf}(p) \geq \epsilon \\ 0 & \text{sinon} \end{cases}$$

$$I_{bf}^{-}(p) = \begin{cases} |I_{bf}(p)| & \text{si } I_{bf}(p) \leq -\epsilon \\ 0 & \text{sinon} \end{cases}$$

$$I_{bf} = DoG * I$$

$$DoG = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}}$$

Dans le cas des images satellitaires, les critères d'homogénéité seront explicités par rapport à la nature des objets attendus dans l'image segmentée. Ces critères tels que l'indice de concentration de chlorophylles, l'indice de végétation ou encore l'indice de réflection de l'eau apportent une précision supplémentaire à l'homogénéité des valeurs des pixels. Quant aux images naturelles (photographiques), ces critères sont plus complexes à définir et dépendent essentiellement de l'observateur. En effet, chaque observateur peut définir par des mots-clés ou bien par une description synthétique les caractéristiques des objets recherchés. À partir de là, se pose donc une question fondamentale sur la validation des objets segmentés et principalement si la qualité de la segmentation est acceptable par rapport aux buts applicatifs. En imagerie médicale comme en imagerie satellitaire, le critère d'acceptabilité peut être quantifié par des mesures classiques tels que le rappel, la précision et la $F-$mesure tandis que pour les images naturelle, il nous semble important d'impliquer l'avis de l'observateur via un critère sur le retour de pertinence. Ces différents critères seront présentés et discutés dans le chapitre 4. Par ailleurs, les images satellitaires ont les inconvénients d'être multi-sources, multi-résolution, de taille très élevée et quelquefois nécessitent du traitement en temps réel. Notre algorithme $SWMP$ permet de segmenter d'une manière satisfaisante les images en :

— détectant tous les objets pertinents dans l'image ;

— de tenir compte des inconvénients cités précédemment puisque le critère d'homogénéité que nous utilisons est applicable à n'importe quelle résolution, et via MapReduce, l'image est divisé en blocs quelle que soit la taille de l'image traitée.

Cependant, notre algorithme $SWMP$ possède deux problèmes majeurs. Le premier problème concerne le traitement temps réel où il est recommandé d'étendre le code de $SWMP$ à l'environnement Spark streaming. Le second problème est lié à la labellisation des objets. Dès lors que la segmentation est vue comme l'identification de chaque pixel de l'image, il est donc nécessaire de créer une carte de labellisation de tous les objets différents détectés dans l'image. Or le résultat que nous obtenons avec $SWMP$ consiste à identifier toutes les instances d'un même objet sans pour autant les catégoriser. Notre algorithme de segmentation est de type *« instance-segmentation » versus « semantic-segmentation »*. C'est le but de cette dernière étape de notre algorithme qui consiste à identifier les labels des

objets.

2.3.5 De l'image segmentée à l'image labellisée par apprentissage profond

L'approche naïve de labellisation est de réduire la segmentation à une tâche de classification. Les classes correspondent aux labels distincts des pixels de l'image. Pour les images satellitaires et les images médicales, les classes sont définies au préalables, contrairement aux images photographiques où les objets sont dépendants de la scène. Pour cette raison, nous devons nous intéresser à deux types de classification à savoir supervisés et non supervisés. Différents travaux existent dans la littérature dont certains visent des applications spécifiques alors d'autres visent de nouveaux techniques et algorithmes. Parmi eux, nous trouvons les méthodes de classification par apprentissage profond. Ces méthodes sont de plus en plus intéressantes puisque l'amélioration de leur performance est sous-jacente au volume très élevé d'images disponibles davantage pour entraîner ces modèles. Par opposition aux modèles traditionnels, l'apprentissage profond nécessite une large quantité de données pour apprendre le modèle. Les données d'apprentissage doivent au préalable être correctement collectées et leur segmentation validée par des experts. Dans ce contexte, la prolifération des jeux de données et des défis a favorisé l'apparition de diverses technologies de pointe pour mettre en oeuvre la segmentation dans divers domaines. Pour les lecteurs intéressés, nous recommandons les travaux d'Alberto Garcia-Garcia [GOO+17] pour une vue d'ensemble de certaines des meilleures techniques de segmentation d'images dans le but de labellisation d'objets grâce à un apprentissage profond. Nous recommandons également les travaux de [SNIU19] pour avoir analysé l'ensemble des méthodes en expliquant pourquoi, quand et comment ces techniques fonctionnent face à divers défis.

Apprentissage par transfert pour la segmentation sémantique

Les réseaux de neurones à convolution sont très répandus en vision par ordinateur et leurs architectures ont évolué pour s'adapter à des problèmes de segmentation. L'architecture basique et commune à ces réseaux est le *fully convolutionnal layers*, nommé dans la littérature par FCN [LSD15]. La sortie de la dernière couche de convolution est utili-

sée pour identifier la classe d'un pixel. Plusieurs familles à base de FCN existent telles que *DeepMask* et *SharpMask* développées par Facebook [PCD15], *R-CNN* (pour *Region based Convolutionnal Neural Network*) [GDDM14], *DeepLab* [CPSA17] [CPK+18], *PNSP-Net*, *RefineNet* [LMSR17] et *VGGNet* [SZ14]. Pour utiliser efficacement ces architectures, il faut tout d'abord sélectionner un modèle pré-entraîné, ensuite choisir la mesure *Cross Entropy* comme la fonction de perte permettant de calculer la probabilité d'appartenance d'un pixel à une classe, enfin d'utiliser le principe du *fine-tune* sur le réseau pour l'adapter au problème spécifique. Une approche de segmentation à base de réseaux de neurones profonds existe. Les auto-encodeurs à convolution ont été utilisés traditionnellement pour l'extraction des descripteurs. Un auto-encodeur est composé classiquement d'un encodeur dont le but est de réduire la dimension de représentation des données en entrée sous la forme d'une représentation intermédiaire et d'un décodeur dont il a la charge de reconstruire les données d'origine à partir de la représentation intermédiaire. L'un des principaux avantages de l'utilisation de l'approche fondée sur les auto-encodeurs par rapport aux modèles FCN est la liberté de choisir la taille de l'entrée. Grâce à une utilisation intelligente des opérations de sous-échantillonnage et de sur-échantillonnage, il est possible de produire une distribution de probabilité à l'échelle du pixel qui soit de la même résolution que l'image d'entrée.

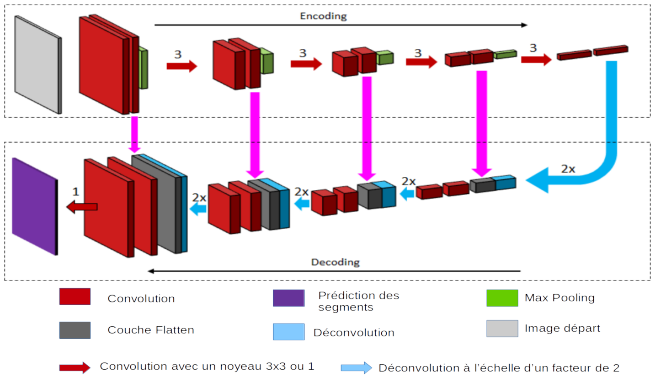


FIGURE 2.16 – Macro-architecture du réseau UNET inspirée de [SNIU19] [5]

Afin de choisir entre ces deux approches de réseaux de neurones profonds, nous avons posé

5. Le détail de l'architecture est donné par Olaf Ronneberger dans https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

les critères suivants :

1. la performance du modèle choisi se base sur les résultats déjà obtenus sur des collections utilisées dans le cadre des challenges ;
2. la possibilité d'étendre des modèles pré-entraînés ;
3. les modèles sont capables de traiter des images de tailles et de formats variables ;
4. prendre en compte l'aspect multi-domaines.

Ce travail a fait l'objet d'un stage de Master 2 recherche mené par Hanen Balti et a donné lieu à deux publications en 2019 [BMC+19] [Mel19a]. Le premier objectif visé par cette étude porte sur la comparaison de la performance de deux méthodes VGGNet (*cf.* figure 2.17) et UNET (*cf.* figure 2.16) à segmenter des images satellitaires. Ces deux modèles ont été largement utilisés à des fins de segmentation sémantique et ont été comparés sur la base de plusieurs collections d'images satellitaires dans le cadre de plusieurs campagnes d'évaluation. Le second objectif consiste à mesurer l'apport d'une architecture distribuée pour notre approche en terme de réduction du temps de calcul sans détérioration de la qualité de la segmentation. Notre étude a visé en premier lieu les images satellitaires afin de répondre en même temps aux critères de complexité $C2$ à $C6$ (voir le chapitre 1 Définition 1).

Notre approche $Deep - SWMP$

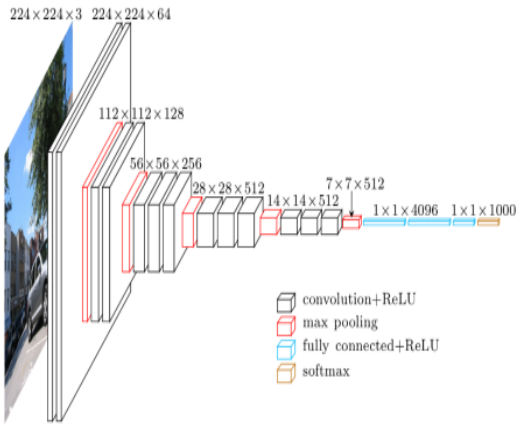


FIGURE 2.17 – Macro-architecture du réseau VGGNet [SZ15]

Notre objectif est de choisir parmi les deux architectures celle nous permettant de détec-

ter le plus d'objets significatifs à partir d'images. UNET (architecture en forme de U) est un auto-encodeur à convolution permettant de construire une carte de descripteurs située avant la première couche de décodage. La sortie du décodeur est une image de même taille que l'image d'entrée contenant les segments labellisés. Afin d'adapter l'architecture à notre problème de segmentation, nous avons ré-entraîné le modèle sur les images. Parallèlement, le modèle VGGNet, composé de 22 couches dont 16 couches d'apprentissage des poids (13 couches de convolution et 3 couches denses fortement connectées), construit également une carte de descripteurs représentée par la dernière couche précédant la couche *SOFTMAX*. Comme pour le modèle UNET, nous avons restreint les sorties aux labels attendus par les images étudiées. Les cartes respectives des descripteurs obtenues sont analysées et réutilisées comme descripteurs sémantiques. La comparaison entre les deux architectures a été réalisée sur trois collections d'images. La collection SIRI-WHU [6] contient 12 classes. Chaque classe contient 200 images de $200x200$ pixels à 2 m de résolution spatiale. La collection AID [7] est constituée de 10000 images de résolutions variables et hétérogènes, réparties sur 30 classes ; enfin, une collection propriétaire *SATSPECT* que nous avons composée d'images multispectrales, issues de différents capteurs tels que SPOT4, SPOT5, LANDSAT8, etc. Notre collection SATSPECT compte 3974 images réparties en 10 classes, de résolutions différentes multispectrales de $1kmx1km$ en 3 et 16 bandes. Les images à 3 bandes sont les images traditionnelles en RGB. Quant aux images à 16 bandes(multibandes), elles sont issues de l'imagerie multispectrale (400-1040 nm) et de l'infrarouge à ondes courtes (SWIR) (1195-2365 nm).

	VGG	UNET
Précision Moyenne	0.71	0.78
Rappel Moyen	0.66	0.77
$F1 - score$ Moyen	0.68	0.77

TABLE 2.2 – Moyenne des résultats obtenus par VGG et UNET sur les trois collections d'images.

Les résultats obtenus montrent l'intérêt de l'architecture UNET par rapport à un problème

6. www.lmars.whu.edu.cn/
7. project.inria.fr/aerialimagelabeling/

tel que la segmentation. UNET permet non seulement d'avoir plus d'objets détectés mais il présente deux avantages précieux pour notre étude. Le premier avantage est la compatibilité de l'architecture par rapport à la diversité des formats de nos images. UNET permet de segmenter des images dont les résolutions sont hétérogènes. Le second avantage est liée à la dimension des images en entrée où UNET fournit une carte de descripteurs ou une carte des objets segmentés de même taille que l'image d'origine, contrairement à VGG où les images traitées doivent être limitées en taille ce qui peut expliquer en partie ses résultats insuffisants face à des images de haute résolution telles que les images multi-bandes. Par ailleurs, nous avons constaté une difficulté de UNET à segmenter les images multi-bandes et les images spectrales de tailles dépassant les 1024 x 1024 pixels (en réalité si l'image contient n bandes, l'image fait $nx1024x1024$). Ce travail a été effectué sur une machine Amazon AWS ayant 16 GB NVIDIA K80 GPUs, 64 vCPUs et 6O GB de RAM. Toutefois, quand les images sont de taille importante, le réseau nécessite plus de couches supplémentaires en phase d'encodage et par conséquent en phase de décodage. Cette contrainte non négligeable impacte directement le nombre de paramètres du réseau qui explose. De ce fait, la qualité des résultats se dégrade et le temps de calcul augmente considérablement. La grande question que nous nous sommes posée est de quelle(-s) manière(-s) pourrions nous réduire le nombre de couches en traitant des images de grandes tailles. Notre réflexion nous a conduit à proposer deux pistes de solution qui visent à restreindre la phase de pré-traitement du réseau par une phase extérieure et en amont de la phase de segmentation. En effet, la première solution consiste à paralléliser la phase de segmentation où une architecture UNET sera attribuée par bande. Chaque bande sera donc traitée par une architecture UNET appropriée. Chaque image d'une bande sera découpée en un flot d'imagettes de tailles réduites. Cette première solution a l'avantage de réduire, de fait, le temps de calcul et favorise l'extraction des caractéristiques locales par bande des objets. Néanmoins, les résultats sont moins performants puisqu'ils sont exposés à la réduction de la résolution (aussi bien spatiale, que fréquentielle) des images engendrée par le traitement séparé de chaque bande. La seconde solution que nous proposons s'inspire des méthodes de *Pansharpening*. C'est un processus de fusion d'images panchromatiques de haute résolution et d'images multi-spectrales de basse résolution pour créer une seule image couleur de haute résolution. Par exemple, Google Maps ainsi que toutes les entreprises de création de

cartes utilisent cette technique pour améliorer la qualité des images. De façon comparable, nous présentons en entrée du réseau UNET$^{(k)}$, l'image pré-segmentée par $SWMP$ en plus de l'image de la bande.

2.4 Conclusions

Le manque des données correctement étiquetées dans beaucoup de domaines complique la tâche de classification. Ce manque est souvent lié à une intervention humaine très fastidieuse et complexe comme c'est souvent le cas pour les images et tout particulièrement les images satellitaires. C'est pourquoi la phase de segmentation est une étape décisive pour la suite de la chaîne de traitement. La qualité de la segmentation varie de la nature de l'image, de la complexité des objets présents dans l'image, et de sa résolution. Alors d'imaginer un algorithme de segmentation serait une aberration, en revanche orienter la ou les techniques de segmentation à appliquer pour les images du « domaine » en fonction des objets attendus serait plus approprié. C'est typiquement les difficultés que nous avons rencontrés dans la segmentation des images satellitaires qui ont la particularité d'agréger des objets différents à différentes échelles. Nous constatons par exemple que L'algorithme de segmentation des zones forestières n'est pas du tout adapté à la segmentation des fleuves avec des ramifications. Dans cette perspective, nous travaillons sur une piste d'amélioration de la segmentation de telles images en proposant d'intégrer des méta-connaissances (comme la connexité, l'échelle, propriétés morphologiques, propriétés topologiques, calorimétrique, etc.) sur les objets attendus et de proposer un algorithme de segmentations par échelle. Par exemple à l'échelle des rivières, on peut proposer des algorithmes de segmentation à base de graphe acyclique inspirés par les travaux [PCT+15].

Au delà du choix des algorithmes, la contrainte temps réel et le volume des données posent de nouveaux défis algorithmiques. En effet, dans un monde en pleine accélération, le volume des données collectées explose. Leur traitement en temps réel devient une exigence et une variable essentielle de la prise de décision. Une autre nouveauté liée à la masse des données disponibles, consiste à la démultiplication des potentiels de leur croisement dans une perspective de création de valeur. C'est pourquoi l'identification des objets perceptifs

est une étape que l'on peut qualifier de pré-traitement utile à la caractérisation sémantique de l'objet visuel. Cette double identité permettrait par la suite de construire des index sémantiques ouverts (dans le sens d'une possibilité d'une mise à jour régulière et cohérente) et augmentés par tout type de connaissance ajoutée depuis des données non structurées telle que les réseaux sociaux. Ces derniers donnent lieu à un champ immense de données non structurées et encore insuffisamment exploités en terme d'enrichissement de connaissance.

Chapitre 3

Enrichissement sémantiques des descripteurs visuels

Sommaire

3.1 Sémantique, texte et images : quelles correspondances ? . . . 69

3.2 Images médicales et descripteurs sémantiques 71

3.2.1 Images 2D 72

3.2.2 Images 3D 73

3.3 Descripteurs sémantiques d'images photographiques 77

3.3.1 Relations spatiales 78

3.3.2 Formalisation des relations spatiales 79

3.3.3 Relations sémantiques 85

Formalisation 86

Construction des domaines $\mathcal{D}$ et leurs vocabulaires 87

3.4 Enrichissement des domaines et module ontologique 91

3.4.1 Conceptualisation du lexique 92

3.4.2 Détection des relations 94

Les relations taxonomiques 94

Les relations sémantiques 95

3.5 Représentation multi-dimensionnelle des connaissances . . . 97

3.5.1 Raisonnement 99

3.6 Conclusions 100

Nous avons décrit dans le chapitre précédent le processus d'extraction d'un objet sémantique et ce pour deux catégories d'images. Pour l'image médicale, l'objet visuel consiste en un graphe connexe d'éléments 2D composé de points et de segments ou bien d'éléments 3D composé de boules et de surfaces. Quant aux images photographiques, nous avons pu associer aux concepts visuels des descripteurs visuels ciblés et appropriés. Des informations d'ordre sémantique gravitent autour de ces objets visuels quel que soit le type de l'image et le domaine auquel elle appartient. L'information sémantique permet de mieux caractériser l'objet et facilite sa comparaison par rapport à d'autres objets similaires. Extraire la sémantique de données multimodales à des fins d'interprétation est un problème complexe qui ne dépend pas seulement des données elles-mêmes. Il dépend également de connaissances *a priori* sur le domaine d'application (connaissances de haut niveau) d'une part, et du contexte applicatif d'autre part. Grâce aux progrès récents en ingénierie des connaissances, en particulier le web sémantique, on assiste à une valorisation des approches s'appuyant sur la modélisation de connaissances *a priori* sur le domaine étudié. En particulier, les techniques du web sémantique à travers l'usage de ressources ontologiques permettent de franchir un tel problème en offrant un cadre normalisé de formalisation des connaissances contextuelles d'un domaine donné [KH08]. En imagerie médicale, l'information sémantique est issue principalement des connaissances théoriques souvent exprimées par les médecins experts et publiées dans des articles de journaux spécialisés. Alors que l'information sémantique liée aux images photographiques, peut prendre différents formes. Quand elle est interne à l'image, elle peut renseigner la forme de l'objet perceptif, sa couleur ainsi que ses interactions avec d'autres objets de la même image. Dans des ressources externes, l'information peut se présenter sous la forme d'hyperlien, de tag, ou encore sous la forme d'annotations visuelles, textuelles, voire même sonore. Ce sont ces informations qui ont motivés nos travaux et nous présentons la méthodologie que nous avons adoptée dans la suite de cette section.

TABLE 3.1 – Résumé des projets et des encadrements sur la thématique de l'enrichissement sémantique des descripteurs visuels

Master/Thèse/Projet	Collaborations
Olfa Allani (2014-2017)	Thèse en co-tutelle Hajer Baazaoui (RIADI, Tunis) et Herman Akdag (LIASD, Paris8)
projet ANR-09-BLAN-0029 MATAIM (Modèles Anisotropes de Texture : Applications à l'Imagerie Médicale. (2010-2013))	Frédéric Richard, Anne Ricordeau Map5, Univ.ParisDescartes, IPROS (CHU Orléans)
Houssem Zitoun (2015-6 mois)	Stage de Master2 en co-encadrement avec Lamia Belouaer post-doctorant à l'École Navale de Brest
Amira Khemissi (2015-6 mois)	Stage de Master2 en co-encadrement avec Riadh Farah, ISAMM, Université de la Manouba

3.1 Sémantique, texte et images : quelles correspondances ?

Le problème de la mise en correspondance entre le niveau perceptif des données visuelles discuté dans le chapitre 2 (pixels ou voxels, groupes de pixels caractérisés par des descripteurs de bas niveau associés) et leur niveau sémantique (description linguistique en utilisant le vocabulaire du domaine d'application) est un problème bien identifié dans le domaine de l'interprétation et de l'indexation d'images. Il est connu par la fossé sémantique et est défini dans [LZLM07] comme la divergence entre les informations perceptuelles extraite des images et leur interprétation par un utilisateur *lambda* dans une situation déterminée.

Le fossé sémantique est un problème ancien qui a été soulevé dans les travaux fondateurs du domaine de la vision par ordinateur [Mar10] [TG80] [KPD+11]. Il est également identifié comme une problématique d'ancrage de symboles en intelligence artificielle et en robotique [CS99], vu comme la création et la mise à jour de la correspondance entre les symboles et les données perçues (par un capteur visuel ou non) représentant le même objet physique (ou concept abstrait). De manière similaire, dans [BH11], une étude des correspondances entre l'ancrage de symboles et l'annotation sémantiques consiste à associer dynamiquement une information linguistique de haut niveau à un ensemble de primi-

tives perçues (et extraites) dans l'image. Cette information linguistique fait référence aux concepts du domaine d'application ainsi qu'à leur définition.

Toutefois, la notion de sémantique d'une image, bien qu'elle soit indispensable, reste assez vague et varie suivant le contexte applicatif. Par exemple, l'annotation d'images par le contenu se limite souvent à un ensemble de mots-clés indépendants désignant les objets présents dans l'image ; mais elle peut aussi prendre la forme d'une longue phrase décrivant de manière linguistique le contenu de l'image (*cf.* figure 3.1) incluant des relations entre les différents concepts.

```
1  <DOC>
2  <DOCNO>annotations/00/116.eng</DOCNO>
3  <TITLE>Termas de Papallacta</TITLE>
4  <DESCRIPTION>a rock hole in the ground
   surrounded by sand; one man in a white tee-
   shirt, grey pants and a white cap is holding a
   shovel, the other one is wearing black pants, a
   white jumper and a traditional hat, is walking
   towards him;</DESCRIPTION>
5  <NOTES></NOTES>
6  <LOCATION>Papallacta, Ecuador</LOCATION>
7  <DATE>April 2002</DATE>
8  <IMAGE>images/00/116.jpg</IMAGE>
9  <THUMBNAIL>thumbnails/00/116.jpg</THUMBNAIL>
10 </DOC>
```

FIGURE 3.1 – Images extraites de la collection Image CLEF 2008

Les dimensions sémantiques auxquelles nous nous intéressons sont inspirées des travaux de Biederman [Bie72] [Bie87] :

— le domaine d'application : la sémantique d'une image peut être définie comme une caractéristique émergente de l'interaction entre des données observées et des connaissances sur ces données. Pour extraire cette sémantique, il est indispensable de recourir à un cadre formel et méthodologique permettant de représenter et de raisonner sur cette connaissance ;

— l'information spatiale : un objet visuel existe toujours dans le contexte de son environnement et jamais de manière isolée. Il existe donc des relations spatiales entre l'objet et son voisinage, c'est-à-dire avec les autres objets en terme de coexistence spatiale. Il a été démontré dans de nombreuses études cognitives [Tor03] [Bar04] [OT07] que c'est une tâche incontournable en reconnaissance de scènes et demeure sources de motivation de nombreux travaux [CdFB04] [GB10] [DJM18]. C'est en particulier, la connaissance structurelle de la scène et des différents objets pouvant coexister

dans cette même scène que nous cherchons à identifier et modéliser par des relations spatiales entre ces différents objets ;

— l'information contextuelle : dans [GB10], l'information contextuelle est définie comme toute information provenant indirectement de l'apparence d'un objet. Cette information peut être liée au voisinage visuel de l'objet dans l'image, à des informations associées à l'image tels que des tags ou des annotations textuelles ou encore à des mots-clés. Cette information permet souvent de réduire voire de lever l'ambiguïté entre deux objets d'apparence similaire.

Dans notre modèle de représentation de la sémantique, nous considérons trois niveaux d'abstraction à savoir le niveau sémantique primaire de l'objet visuel qui correspond à la description des objets présents dans l'image, le niveau de la sémantique secondaire représenté par les relations entre les objets de l'image et le niveau de la sémantique globale qui inclut tout raisonnement à partir des ressources extérieures et les annotations permettant d'inférer la description de l'image dans sa globalité. Nous allons détailler ces aspects pour trois domaines d'application que sont les images médicales, photographiques et satellitaires.

3.2 Images médicales et descripteurs sémantiques

À partir des connaissances sur l'évolution de l'ostéoporose que nous avons pu collecter à partir des articles spécialisés, puis validées par les médecins, nous avons proposé de définir des descripteurs sémantiques pour des objets visuels. Ces descripteurs permettent tout d'abord de labelliser ou d'associer une étiquette de type de structure à chaque élément de l'objet visuel. Comme exemples d'étiquette pour l'objet 2D, nous citons les noeuds de jonction reliant deux ou plusieurs structures, les segments de tension et les segments de compression. La labellisation concernera donc chaque élément pertinent du squelette qui n'est rien d'autre qu'un réseau connecté d'éléments structurels. Par conséquent, les relations spatiales sont inférées à partir de la caractérisation du réseau. Parmi les relations spatiales les plus importantes que nous cherchons à caractériser citons, par exemple, *les segments de compression sont alignés dans une direction privilégiée*, les *les segments de tension sont peu denses*, les *trous sont très étendus*, etc. Pour réaliser ces objectifs, et

après avoir élagué le squelette θ_i de chaque image I_i, nous cherchons à le caractériser de la manière suivante :

1. extraire les segments ;
2. labeliser les segments en fonction de leur orientation, c'est-à-dire de tension et de compression ;
3. Estimer leur nombre et le nombre d'interconnexions ;
4. Estimer le nombre de trous ;
5. Estimer la taille des trous ;
6. localiser les zones avec peu de segments de tension et beaucoup de trous.

3.2.1 Images 2D

Tout d'abord, le squelette θ peut être vu comme l'union de trois ensembles selon $\theta = \{point_{Segments}\} \cup \{point_{Intersections}\} \cup \{point_{Extremites}\}$. Les segments sont de deux types : des segments entre deux intersections ou bien un segment simple avec une intersection et une extrémité. Tous les segments sont labellisés en segment de compression (direction $\mathcal{D}^1$) et en segment de tension (direction $\mathcal{D}^2$). La méthode de labellisation des segments en $\mathcal{D}^1$ et $\mathcal{D}^2$ se déroule comme suit. À chaque segment est associé un angle calculé entre l'axe horizontal et l'axe majeur de l'ellipse contenant les pixels. Une fois tous les pixels de chaque segment labellisés par l'orientation de leur segment, nous procédons à une étape de classification basée sur l'histogramme des orientations (*cf.* figure **??**).

Les intersections sont des points qui connectent un ou plusieurs segments. Les extrémités sont les points terminaux d'un segment. Sur ces trois ensembles, nous cherchons à caractériser l'anisotropie de la structure projetée 2D. Ce descripteur d'anisotropie se caractérise souvent par une répartition désordonnée ou déséquilibrée des points d'intersection mais aussi des segments dans la direction de compression des travées. Pour le quantifier, nous avons choisi d'utiliser l'entropie et de l'appliquer aux points d'intersections et aux segments de compression. En effet, pour décrire la répartition spatiale des nœuds, les entropies spatiales sont calculées à partir des blocs de taille $N \times N$. Pour chaque valeur de $N = 2, 4$ et 8, l'entropie spatiale de Shannon est calculée en prenant en considération la proportion

TABLE 3.2 – Paramètres descriptifs de la structure 2D à base de segments et de points d'intersection.

Paramètres morphologiques	paramètres d'anisotropie
$\mathcal{W}$	$\mathcal{W}^{D1}$, $\mathcal{W}^{D2}$
$\mathcal{L}$	$\mathcal{L}^{D1}$, $\mathcal{L}^{D2}$
$\mathcal{L}_{90\%}$	$\mathcal{L}^{D1}_{90\%}$, $\mathcal{L}^{D2}_{90\%}$
$\mathcal{L}_{95\%}$	$\mathcal{L}^{D1}_{95\%}$, $\mathcal{L}^{D2}_{95\%}$
$\mathcal{L}_{99\%}$	$\mathcal{L}^{D1}_{99\%}$, $\mathcal{L}^{D2}_{99\%}$
$n.S$	$n.S^{D1}$, $n.S^{D2}$,$\mathcal{T}^{D2} = \frac{n.S^{D2}}{n.S}$
$n.Intersections$	$Entropie(N), N = 2, 4, 8$

des points d'intersection présents en chaque bloc. On considère n le nombre total de points d'intersection, i le numéro du bloc et p_i la proportion de points d'intersection dans le bloc i :

$$Entropie(N) = -\sum_{i=1}^{N^2} p_i.ln(p_i).$$

L'entropie vérifie l'inégalité suivante : $0 \leq Entropie(N) \leq \ln(n)$. Quand $Entropie(N) = 0$, cela signifie que les points d'intersection sont concentrés dans un bloc. À l'inverse, une répartition uniforme des points d'intersection sur les n blocs est atteinte quand $Entropie(N) = \ln(n)$. Les paramètres issues des segments en terme de longueur, largeur et direction sont résumés par le tableau 3.2

3.2.2 Images 3D

La caractérisation des objets en 2D s'avère peu fiable puisque l'image 2D est une réduction de l'architecture plus complexe en 3D. Comme elle résulte d'une projection dans une direction fixée, l'orientation de la structure est de ce fait biaisée par l'accumulation de la matière (os). Pour cela, nous avons cherché à identifier l'orientation de la structure osseuse 3D en séparant les deux structures plaque et poutre. C'est une étape de discrimination effectuée sur les voxels situés à la surface (les voxels de type grain en contact avec au moins un voxel de type pore). Ce problème peut être vu comme une succession de deux tâches, d'abord de segmentation pour extraire les voxels de surface puis de classification des voxels en deux classes plaque et poutre. Différentes méthodes plus ou moins com-

plexes qui proposent de résoudre ce problème existent dans la littérature. On distingue en effet deux classes de méthodes. Les méthodes de segmentation orientées structure locale, sont pour la plupart, basées sur le calcul du squelette tel que « l'axe médian » (AM) et couplées à des algorithmes d'analyse topologique des voxels du squelette (dit DTA pour *Digital Topological Analysis*) [PACB10]. Différents algorithmes de DTA existent [SYH+10] et proposent des classifications très subtiles des voxels en points de jonctions, en points de surface, en points de courbes et en d'autres types secondaires. Il faut cependant mentionner que les méthodes de calcul du squelette AM sont très complexes et nécessitent le réglage de plusieurs paramètres. De plus, DTA est un algorithme très précis mais très sensible au moindre bruit généré par le calcul du squelette. Compte tenu de sa grande sensibilité aux bruits, cette solution ne peut pas être appliquée à nos images où le bruit est inhérent à la technique d'acquisition. Pour remédier à ce problème, nous avons choisi d'utiliser des méthodes statistiques qui permettent de moyenner le bruit et de réduire considérablement son impact. Parmi ces méthodes, les tenseurs de structure sont très utilisés en physique des matériaux pour caractériser leur orientation. Les tenseurs ont été également appliqués sur des images binaires 2D pour quantifier l'anisotropie globale de la structure osseuse en utilisant la méthode *« Mean Intercept Length »* (MIL). Le MIL est une méthode ancienne proposée par [Whi74]. Elle consiste à compter le nombre de points d'intersection entre les bords de la structure et des des lignes parallèles dans une direction donnée. Un paramètre mesurant le degré d'anisotropie (DA) est ensuite obtenu à partir d'une estimation du tenseur de structure (*MIL Fabric tensor*). Pour une image en niveaux de gris, un tenseur fondé sur l'image du gradient est principalement utilisé pour décrire l'anisotropie globale de sa structure. Les vecteurs de gradient sont en effet normaux à la surface. c'est pour cette raison que le MIL et les approches fondées sur le gradient quantifient l'anisotropie des bords des structures et non la structure locale. Pour une discussion plus détaillée sur ces méthodes, nous renvoyons le lecteur auxs deux travaux suivants [RM16] [Tab10]. Notre choix s'est porté sur l'estimation locale du tenseur autour d'un voisinage d'un voxel à partir de l'intensité de l'image du squelette 3D θ_K et ce pour tous les voxels du squelette. L'avantage d'utiliser le squelette est de nous permettre une réduction de la dimension de l'image de départ sans perte tout en préservant les caractéristiques morphologiques et topologiques de la structure initiale. Nous notons notre tenseur $\mathcal{ILS}$ pour *Inertia Local*

Tensor. $\mathcal{ILS}$ est estimé en utilisant une boule $\mathcal{B}r$ de rayon r. Le choix de la boule est justifié par le fait que tous les voxels ont la même probabilité d'être considérés indépendamment du type de la structure locale (surfacique ou curviligne). Le tenseur d'inertie $\mathcal{ILS}(p)$ au voxel p est la matrice de covariance associée aux positions des voxels présents dans un voisinage $N(p) \in \mathcal{B}_r$. Par conséquent, $\mathcal{ILS}(\mathcal{P})$ est un tenseur symétrique positif de second ordre qui peut être décomposé selon le théorème spectral avec λ_i les valeurs propres associées aux vecteurs propres u_i :

$$\mathcal{ILS}(p) = Var_{N(p)} = \sum_{i=1}^{3} \lambda_i u_i u_i^t \text{ avec } \lambda_1 \geq \ \lambda_2 \geq \lambda_3 \geq 0. \tag{3.1}$$

Comme le montre la formule (3.1), l'intersection entre la boule en un voxel donné et ses voisins du squelette peut être visualisée par une ellipsoïde ayant les trois vecteurs propres comme axes principaux. Il en résulte que la structure locale centrée en un voxel p est :

1. de type curviligne ou linéaire si et seulement si $\lambda_1 \gg \lambda_2 = \lambda_3 = 0$;
2. de type surfacique si et seulement si $\lambda_1 = \lambda_2 \gg \lambda_3 = 0$;
3. isotrope si et seulement si $\lambda_1 = \lambda_2 = \lambda_3 \gg 0$.

L'ensemble $\mathcal{C} = \{\mathcal{L}, \mathcal{S}, \mathcal{I}\}$ des trois classes, ainsi que les types de structure sont synthétisés par le tableau 3.3. La formule (3.1) peut se décomposer en une somme des trois structures

TABLE 3.3 – Trois classes de structures à l'issue des valeurs propres idéales.

Condition des valeurs propres	Classe de structure	Type de structure
$\lambda_1 \gg \lambda_2 = \lambda_3 = 0$	linéaire	poutre ou bord de structure
$\lambda_1 = \lambda_2 \gg \lambda_3 = 0$	surface	plaque
$\lambda_1 = \lambda_2 = \lambda_3 \gg 0$	isotrope	zones de jonctions

locales permettant une meilleure interprétation du tenseur d'inertie de la manière suivante :

$$\mathcal{ILS}(p) = (\lambda_1 - \lambda_2)\mathcal{ILS}_1 + (\lambda_2 - \lambda_3)\mathcal{ILS}_2 + \lambda_3 \mathcal{ILS}_3 \tag{3.2}$$

Étant donné que λ_1 est une valeur propre strictement positive, on peut calculer à partir de l'équation (3.2), le degré d'appartenance d'un voxel p à chacune des trois classes de structure en divisant $\mathcal{ILS}$ par λ_1. Nous obtenons ainsi trois degrés d'appartenance $\mathcal{MF}_L$,

$\mathcal{MF}_S$ et $\mathcal{MF}_I$ tels que :

$$\begin{aligned}
\mathcal{MF}_L &= \frac{\lambda_1 - \lambda_2}{\lambda_1} \\
\mathcal{MF}_S &= \frac{\lambda_2 - \lambda_3}{\lambda_1} \\
\mathcal{MF}_I &= \frac{\lambda_3}{\lambda_1} \\
\mathcal{MF}_L + \mathcal{MF}_S + \mathcal{MF}_I &= 1.
\end{aligned}$$

le maximum $mf(p) = \max_{\mathcal{C}_i \in \mathcal{C}} (\mathcal{MF}_{\mathcal{C}_i})$ est le degrés d'appartenance calculé en un voxel p, et exprime sa certitude d'appartenance à l'une des trois classes, c'est-à-dire la classe $\mathcal{C}(p) = \operatorname{argmax}_{\mathcal{C}_i \in \mathcal{C}}(\mathcal{MF}_{\mathcal{C}_i})$, dont le degré d'appartenance est maximal (*cf.* figure 3.2).

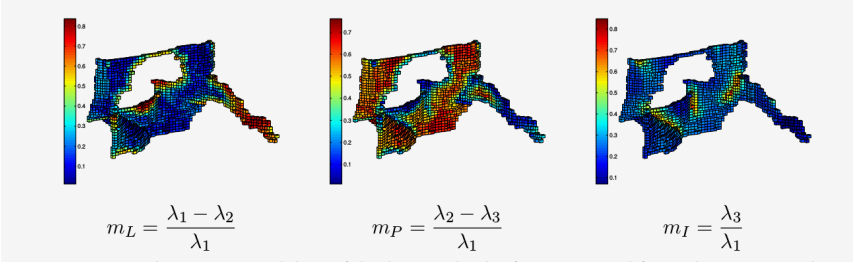


FIGURE 3.2 – Les trois cas de figures d'ILS

Grâce à notre méthode de tenseur d'inertie appliquée directement au squelette du volume 3D initial, chaque voxel du squelette est labellisé et nous disposons de trois labels contrairement aux travaux de recherche existants qui se contentent seulement de deux labels (plaque et poutre). La labellisation se fait en deux passages. Le premier consiste à attribuer un label à chaque voxel indépendamment de ses voisins. Au second passage, nous cherchons à corriger les labels contradictoires par un algorithme de propagation des labels de chaque voxel à ses voisins selon une boule isotrope de rayon 3. La dernière tâche de cette étape de labellisation consiste à labelliser les voxels du volume initial connaissant les labels des voxels de son squelette. Cette tâche inverse s'effectue simplement par propagation géodésique du volume localement centrée en un point x de la surface vers l'ensemble des points du squelette 3D associés (*cf.* figure 3.3 correspondant au volume initial de la figure 2.6).

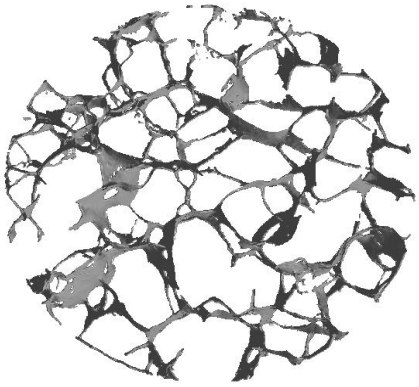


FIGURE 3.3 – Squelette du volume initial

3.3 Descripteurs sémantiques d'images photographiques

Différents types d'information peuvent être associés aux images [Del99] et être source d'enrichissement sémantique. Tout d'abord, citons la quantification et la qualification des relations spatiales entre les objets visuels au sein d'une même image. Ensuite, la quantification de relations temporelles entre images et extraites depuis les métadonnées. Ces dernières peuvent contenir des informations propriétaires telles que le nom du propriétaire, la localisation de l'image (information spatiale), la résolution, le type de l'appareil, et des informations temporelles telles que la date de création (savoir la saison) et l'heure (nuit ou jour). Enfin, la modélisation des relations sémantiques entre les objets issues des descriptions textuelles des images quand elles existent, ou bien directement déduites à partir des concepts représentant les objets visuels sur l'ensemble de la collection d'images. Les descriptions sont souvent explicitées en texte libre sans aucune structure prédéfinie pour l'annotation ou bien par des mots-clés choisis à partir d'un vocabulaire fermé ou ouvert. Ce sont ces informations qui ont guidé notre démarche exploratoire afin d'enrichir la sémantique de l'image en respectant sa cohérence avec les objets visuels déjà extraits à l'étape précédente. Contrairement aux images médicales, le mécanisme permettant d'extraire et d'inférer cette connaissance directement à partir des données est désormais possible grâce à l'explosion des masses de données multimédia disponibles. Le cadre méthodologique et expérimental pour l'extraction des relations spatiales ont été élaborés dans le cadre d'un stage de Master 2 co-encadré par Lamia Belouaer post-doctorant à l'École Navale de Brest

et moi-même. Ces relations ont été étendues à des relations spatiales floues et ont fait l'objet du stage de Master 2 d'Amira Khemissi co-encadré par Riadh Farah de l'Université de la Manouba de Tunis. Les relations sémantiques contextuelles ont été découvertes dans le cadre de la thèse d'Olfa Allani où des bases d'images annotées avec du texte libre et des bases d'images avec des pages web associées ont été examinées. L'ensemble de ces travaux s'appuie sur des images segmentées où l'on extrait l'ensemble des objets visuels pertinents. Cependant, l'étape de segmentation ne sera pas détaillée ici mais fera l'objet d'une section à part entière dans la suite de ce chapitre.

3.3.1 Relations spatiales

Admettons que vous ayez pris une photo d'un groupe de personnes derrière une voiture et que la voiture est à droite d'une église. Lorsque vous effectuez une interrogation d'un système de recherche inversée d'images avec la photographie en question, vous vous attendez normalement à des images similaires. C'est là qu'interviennent les relations spatiales qui constituent une base de descriptions linguistiques contribuant à l'enrichissement sémantique du contenu visuel des objets. Afin de définir la relation spatiale, il faut tout d'abord définir l'entité spatiale. Une entité spatiale est tout objet de l'espace qu'on veut caractériser. Elle peut être statique (un bâtiment, un bureau, une ville, une forêt, etc.) comme elle peut être dynamique (un humain, une voiture, un robot, etc.). Dans notre cas, ce sont les objets visuels déjà identifiés à l'étape précédente. Les relations spatiales représentent l'ensemble des caractéristiques déterminant les positions relatives de deux ou plusieurs objets dans un espace donné. Elles se déclinent en deux types de relations topologiques et métriques (illustrées par la figure 3.4) [HAB08].

Parmi les relations métriques, distinguons les relations de direction et les relations de distance. Outre la simplicité de cette typologie des relations spatiales, les relations de direction restent limitées à des configurations d'ordre (binaire ou tertiaire) basiques telles que *avant*, *après*, *en dessous* ou *en dessus*. Or, dans certaines configurations, notamment en navigation autonome de robots dans un espace dynamique, en détection des objets à partir d'images satellitaires, ou encore dans la détection des alignements des plaques et des poutres dans les images médicales de l'ostéoporose, il est utile de déterminer les

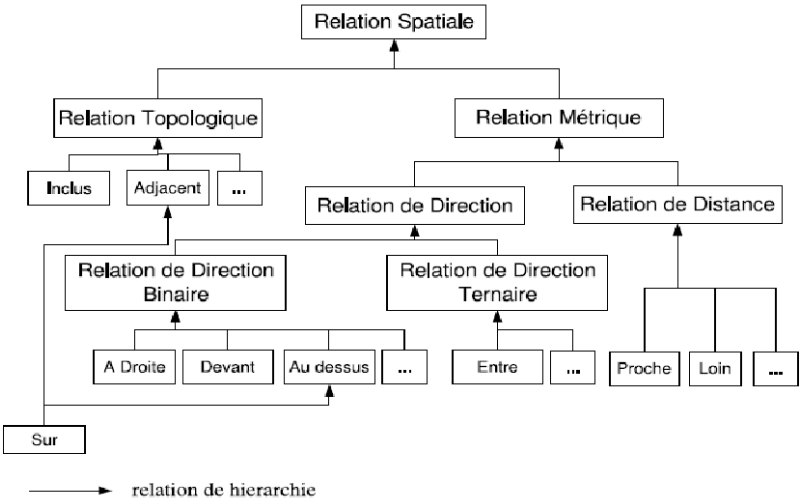


FIGURE 3.4 – Classification des relations spatiales [HAB08]

orientations des objets et de définir des alignements orientés entre les objets. Partant de ce postulat, nous avons proposé d'élargir le modèle proposé par [HAB08] et par [Bel11] où les relations de direction se distinguent en relations d'ordre et en relations d'alignement orienté (*cf.* figure 3.5).

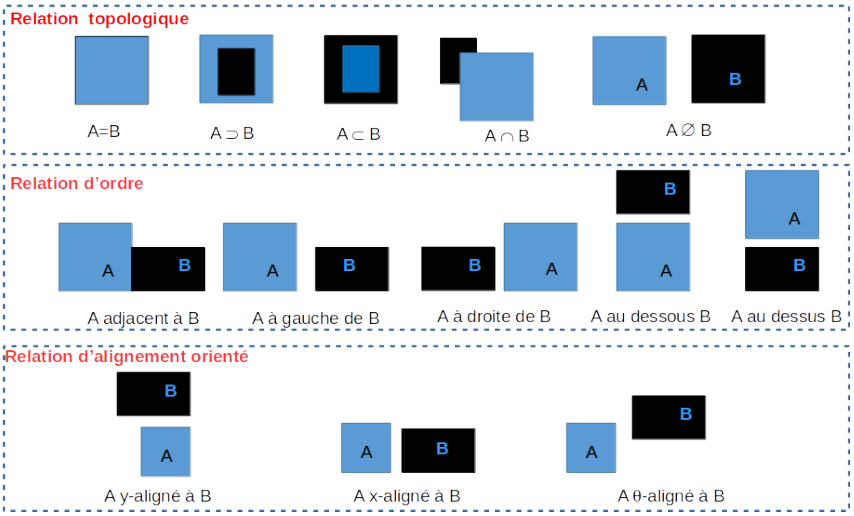


FIGURE 3.5 – Relations spatiales orientées

3.3.2 Formalisation des relations spatiales

Les relations topologiques sont les relations les plus simples à détecter automatiquement. En utilisant principalement des opérations ensemblistes telles que l'intersection et l'inclu-

sion entre deux ensembles, nous avons labellisé chaque paire d'objets avec une relation parmi les cinq présentées dans la figure 3.5. Nous avons considéré deux types de formalisation à savoir par des ensembles classiques et par des ensembles flous.

Distance floue

Les relations de distance permettent de décrire la distance séparant deux objets. Elle peut être quantifiée par un nombre sur $\mathcal{R}^+$ ou encore qualifiée par des relations linguistiques telles que *Près* ou *Loin*. D'autres relations peuvent être générées en changeant le degré de granularité ou bien en ajoutant des quantifieurs linguistiques (*très Loin* ou *peu Loin*). Nous considérons que nos objets visuels sont des objets flous à cause de l'étape de segmentation qui reste un pré-traitement approximatif et, de ce fait, bruité. Une distance floue est justifiée pour tenir compte de l'imprécision topologique des objets visuels. Plusieurs distances entre ensembles flous ont été proposées dans la littérature. Celles qui nous intéressent doivent : (1) permettre de prendre en considération l'hétérogénéité topologique et spatiale des structures ; (2) être invariantes par changement d'échelle ; (3) être symétriques. Les dilatations morphologiques sont des fonctions qui permettent de générer par dualité, des distances vérifiant les trois propriétés fixées. L'avantage du formalisme morphologique est que les distances sont alors exprimées sous forme algébrique et donc plus faciles à étendre au cas flou en préservant leurs propriétés ainsi que les expressions analytiques usuelles. La dilatation floue est tout simplement l'extension de la formule (2.2) à des sous-ensembles flous :

$$\forall x \in I, \Delta_{\mathcal{B}}^{\mathcal{F}}(O)(x) = \underset{y \in I}{sup}(t[B(y-x), O(y)]) \tag{3.3}$$

où $\Delta^{\mathcal{F}}$ est la dilatation floue par l'élément structurant flou $\mathcal{B}$ et t est l'opérateur $t-$norme. La distance floue $\mathcal{D}_{\mathcal{F}}$ entre deux sous-ensembles flous O_1 et O_2 peut être vue comme le nombre minimal de dilatations floues appliquées à O_1 pour intercepter la première fois O_2 (resp. dilatations successives de O_2 jusqu'à intercepter O_1 la première fois). Autrement dit, nous cherchons le point le plus proche soit dans O_1, soit dans O_2. Explicitée de cette façon, la distance floue par dilatations successives n'est autre que la distance de Hausdorff.

$$\forall x \in I, \mathcal{D}_{\mathcal{F}}(O_1, O_2) = \underset{n \geq 0}{inf}(\Delta_{\mathcal{B}}^{\mathcal{F}}(O_1)^n(x)) \cap O_2 \neq \emptyset \tag{3.4}$$

$$= \inf_{n \geq 0} (\Delta_{\mathcal{B}}^{\mathcal{F}}(O_2)^n(x)) \cap O_1 \neq \emptyset$$

Le coût de calcul de la distance de Haussdorff est de $\mathcal{O}(taille(O_1) \times taille(O_2))$ tandis qu'une dilatation est de $min(\mathcal{O}(taille(O_1)), \mathcal{O}(taille(O_2)))$. Par conséquent, la dilatation floue est une solution optimisée d'implémentation de la distance de Hausdorff. Par ailleurs, les distances floues que nous avons utilisées ont été représentées linguistiquement et varient entre *très proche* à *très loin* Quelques résultats de tests sont illustrés par la figure 3.6 où la requête textuelle consiste à chercher des images contenant des enfants très proches des ballons.

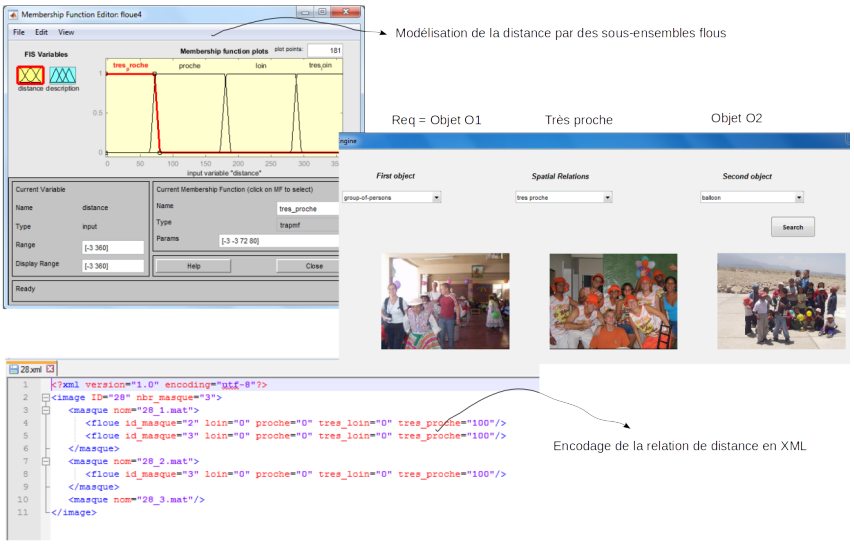


FIGURE 3.6 – Encodage de la distance en sous-ensembles flous et résultats de recherche avec la relation *très proche*

Relation d'ordre

Pour définir une relation d'ordre binaire ou tertiaire, il est indispensable de fixer un objet de référence. En effet, certaines relations d'ordre ne sont pas symétriques. Par exemple, si on souhaite décrire quel objet se trouve à droite d'un autre objet, l'objet de référence impose la sémantique de la relation puisque, si nous inversons la référence, l'ordre n'est plus vrai. Les relations d'ordre non symétriques sont essentiellement *à droite, à gauche, en haut, en bas, devant et derrière*. Tandis que les deux relations *à côté* et *adjacent* sont

symétriques. Pour détecter automatiquement les différentes relations d'ordre, nous avons procédé en trois étapes. La première étape consiste à délimiter les objets par des boîtes englobantes facilitant la détection des collisions entre les objets (voir l'exemple de la figure 3.7). La seconde étape consiste à poser un sens de référence lié à l'observateur et qui sera associé à l'objet de référence. La troisième étape consiste à évaluer la relation en comparant les positions des boîtes englobantes et de les transcrire en XML. Pour déterminer la boîte englobante, nous réutilisons le tenseur d'inertie $\mathcal{ILS}(p)$ présenté par la formule (3.1), appliqué à des objets en deux dimensions. La matrice de covariance associée aux positions des pixels p présents dans un voisinage $N(p) \in \mathcal{B}_r$ est calculée. Par conséquent, $\mathcal{ILS}(\mathcal{O})$ est un tenseur symétrique positif de premier ordre qui peut être décomposé selon le théorème spectral avec λ_i, les valeurs propres associées aux vecteurs propres u_i :

$$\mathcal{ILS}(p) = Var_{N(p)} = \sum_{i=1}^{2} \lambda_i u_i u_i^t \text{ avec } \lambda_1 \geq \ \lambda_2 \geq 0.$$

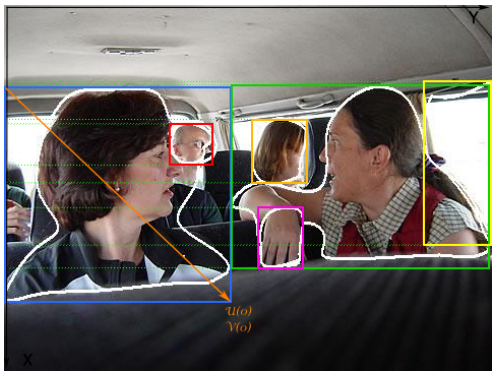


FIGURE 3.7 – Exemples de boîtes englobantes

La boîte englobante est le rectangle qui circonscrit à l'ellipse définie par ces deux vecteurs propres $\vec{u_1}$ et $\vec{u_2}$. Dans la seconde étape, le sens de référence associé à l'objet de référence peut être désigné par l'observateur lui-même. À défaut, il est confondu avec le sens de l'observateur de l'image (droite et gauche de l'observateur projetés sur l'objet de référence). Un exemple de relation d'ordre est illustré par la figure 3.8.

Alignement orienté

Par alignement orienté, nous cherchons à repérer des objets qui sont disposés dans une même direction. Pour cela, nous avons associé à chaque objet l'angle d'orientation de sa

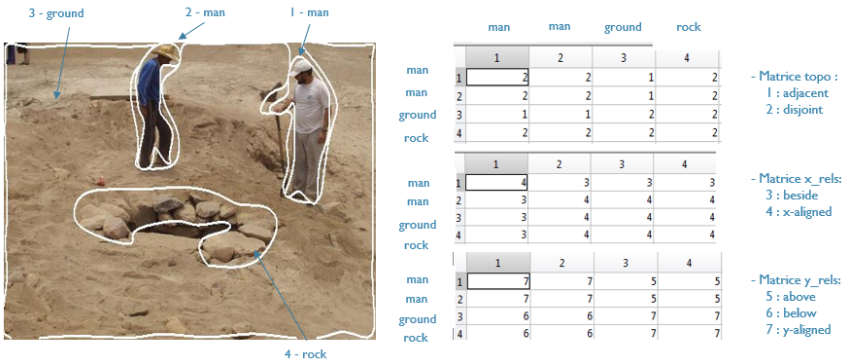


FIGURE 3.8 – Exemple de relations d'ordre

propre boîte englobante. Pour simplifier le problème, nous avons considérer les valeurs absolues des angles et nous les avons réduits en prenant leur *modulo* par rapport à 180°. Deux objets sont détectés alignés dans une direction donnée si et seulement si la différence entre les angles est inférieure à un seuil fixé (*cf.* figure 3.9). Ce seuil a été fixé arbitrairement et correspond à 18° afin de considérer les alignements horizontaux et verticaux. L'histogramme des angles nous a permis d'ajouter une connaissance supplémentaire qui est la dispersion des objets dans l'image. En effet, l'histogramme permet de refléter la répartition des objets dans l'espace de l'image ainsi que la densité de présence des objets dans une ou plusieurs directions privilégiées. C'est exactement la même idée d'anisotropie abordée au niveau des images médicales de l'ostéoporose dans le cas 2D et 3D. Via le degré de dispersion que nous avons associé à chaque objet en plus de la distance entre les objets, nous induisons facilement les objets composant un groupe, comme, par exemple, un groupe de personnes (soit une foule en fonction de la densité et de la distance).

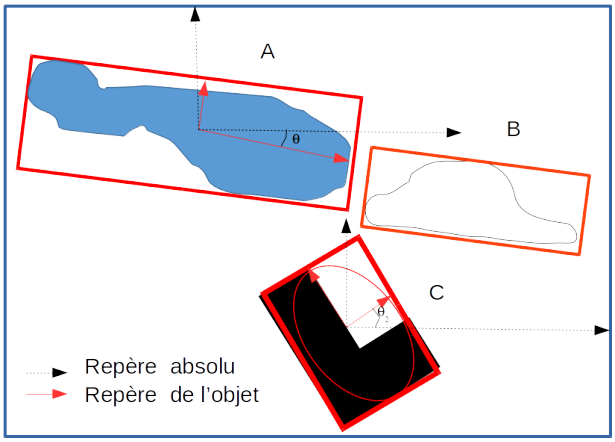


FIGURE 3.9 – Exemple de relations d'ordre : l'objet A est aligné avec l'objet B.

L'ensemble des relations que nous avons traitées sont essentiellement binaires. Les relations ternaires nécessitent davantage de traitement et de formalisation même si dans notre implémentation nous avons considéré la relation simple *« entre »*. Mais des perspectives d'extension de ces travaux sont déjà en cours pour considérer des relations plus complexes. Dans la figure 3.11, nous présentons quelques résultats pour les deux derniers types de relations spatiales que sont l'ordre et l'alignement.

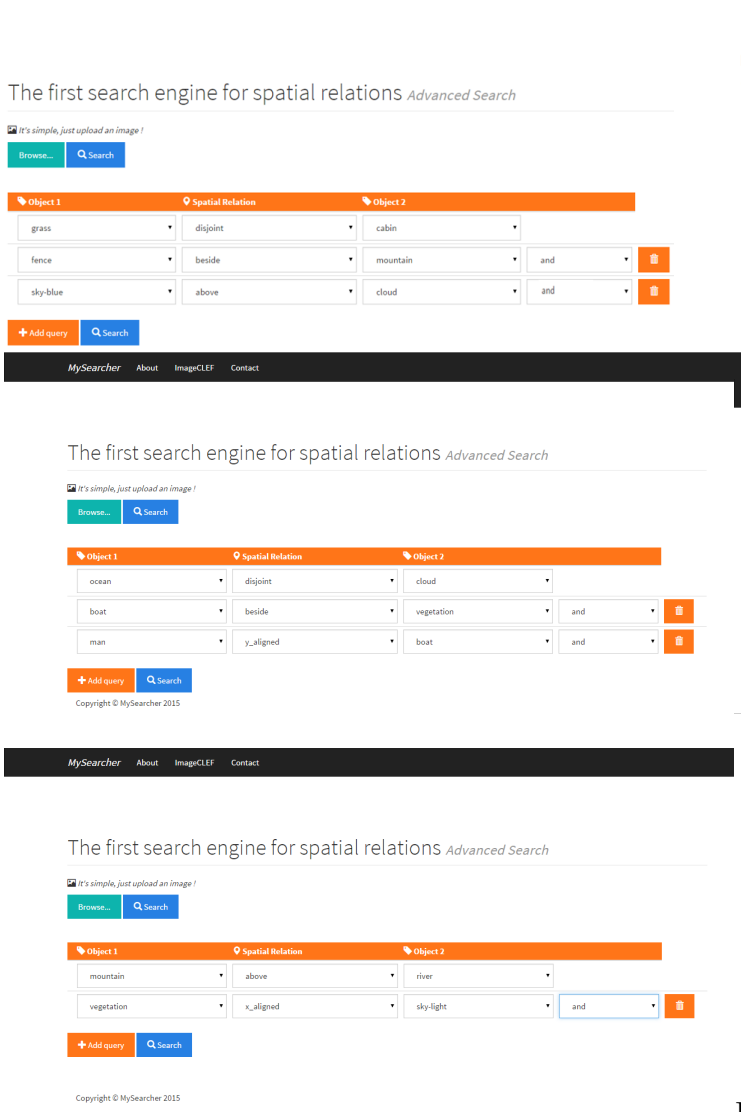


FIGURE 3.10 – Requêtes spatiales

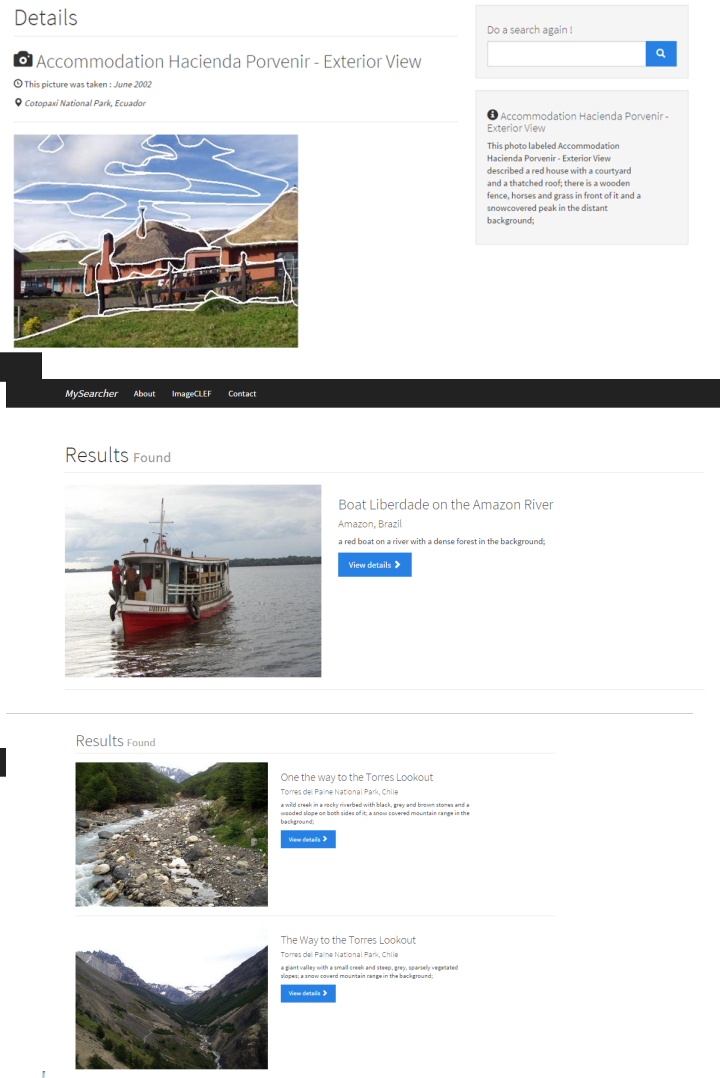


FIGURE 3.11 – Images résultats des requêtes

3.3.3 Relations sémantiques

Le but de cette étape est de construire des relations sémantiques homogènes entre les concepts. Nous allons nous focaliser principalement sur les relations taxonomiques et les relations contextuelles. Les relations taxonomiques permettent d'établir une hiérarchisation des concepts. Elles organisent les concepts dans une structure arborescente. Les relations contextuelles permettent d'attribuer du sens à la terminologie. La terminologie est le résultat de l'extraction des termes décrivant les objets visuels ainsi que la construction de corpus textuels depuis les annotations. Pour atteindre cet objectif, nous allons tout d'abord procéder à une première classification des objets visuels selon les concepts sémantiques et le domaine. Rappelons que notre objet visuel consiste en une région d'intérêt identifiée pertinente dans l'image. Supposons que cette région présente une tour et sera naturellement rattachée au concept *Tour*. Ce concept pourrait faire référence à *une tour de télécommunication* ou bien à *une tour médiévale* ou encore à un événement *le tour de France*. Ce même concept peut donc être associé à trois objets visuels différents laissant penser qu'il s'agit des trois mêmes. Le besoin de distinguer ces objets selon leur concept mais aussi leur sens paraît primordial et indispensable. C'est cette notion de contexte sémantique que nous cherchons à extraire à partir des concepts, des régions d'une même image et la projection des concepts sur le domaine des vocabulaires extrait et analysé depuis les annotations textuelles (*i ;*. le corpus textuel) ainsi que les méta-données. La première classification sémantique des concepts sera basée seulement sur les termes. En effet, chaque terme peut avoir différents sens, sous la forme d'une liste d'éléments appelés *synsets*, portant les sens plausibles associés aux concepts, relativement au contexte d'usage du terme. Afin de lever l'ambiguïté sémantique, une étape de filtrage sémantique de l'ensemble des *synsets*, consiste à éliminer les sens non pertinents en s'appuyant sur le contexte d'usage du terme. Par exemple, pour le terme *Eclipse*, nous disposons des deux annotations suivantes : (1) *Eclipse provides a universal toolset for development and is an Open Source IDE* et (2) *The solar eclipse occurs when the moon passes between the sun and Earth, and the Moon fully or partially blocks the Sun.* Ces deux phrases constituent le contexte d'usage du terme permettant de discriminer le sens le plus probable. Toutefois, il est force de constater que le même terme *Eclipse* appartient à deux domaines différents,

Informatique et *Astronomie*. Par conséquent, la classification des images par ensemble de domaines homogènes est une étape importante et utile pour la *désambiguïsation*. Cette étape ne devra en aucun cas se faire à l'échelle d'une région compte-tenu de sa complexité ($\mathcal{O}(nLog(n))$, n étant le nombre total de régions). Le résultat de la phase de désambiguïsation est l'attribution à chaque concept, un ensemble homogène de paires de contexte et synsets associés.

Formalisation

Avant de détailler dans ce qui suit les différentes étapes d'élaboration des relations contextuelles des concepts, nous proposons de formaliser notre problème de la manière suivante : Soit $\mathcal{O} = \{o_1, o_2, ..., o_N\}$, l'ensemble des objets visuels sur toute la collection d'images. Soit $\mathcal{C}_{min} = \{c_1, c_2, ..., c_M\}$, l'ensemble minimal des concepts de la collection d'images. Comme exemple de concepts, nous pouvons définir C_{min} = {*building, wall, wallpanel, chimneybreast*}.

Chaque concept est relié à un ou plusieurs objets visuels (voir l'exemple du concept *Tour*). On note $\mathcal{C}_{max}$ l'ensemble des concepts $\mathcal{C}_{min}$ enrichi par une ressource lexicale extérieure. Un exemple d'enrichissement de $\mathcal{C}_{min}$ par WordNet [1] est $\mathcal{C}_{max}$ = {*building, wall, wallpanel, chimneybreast, painting, graphicart, portrait, construction*}.

Étant donné qu'une base d'images peut couvrir un ou plusieurs domaines et s'étaler sur un ensemble élargi de thématiques, nous proposons de traiter le cas d'une base multi-domaines. En effet, une seule image couvre généralement plus d'une thématique. Par exemple, une image de paysage avec une montagne ou une forêt contient les éléments *ciel, verdure, eau*, etc. Désignons par $\mathcal{D} = \{d_1, d_2, ..., d_P, \}$ l'ensemble des domaines couverts par la collection.

On considère $\mathcal{X}$ l'ensemble des contextes définis de $\mathcal{O} \times \mathcal{C} \rightarrow \mathcal{D}$ tel que $\mathcal{X}(o_i, c_j) = d_k$ est un sous-ensemble de $\mathcal{D}$, défini comme étant les objets visuels du domaine d_k.

1. https://wordnet.princeton.edu/

Construction des domaines $\mathcal{D}$ et leurs vocabulaires

Nous proposons ici de construire dynamiquement les domaines de la collection. Nous partons toujours des concepts issus des termes qui correspondent aux objets visuels obtenus par segmentation suivie de reconnaissance d'objets dans l'image. Par nature, $\mathcal{C}$ est un ensemble hétérogènes de concepts qui peut être structuré en groupes de concepts homogènes. Ce problème est vu comme une classification non supervisée puisque nous ne disposons d'aucune information *a priori* sur le nombre de domaines. Il s'agit donc de partitionner un ensemble de concepts en un ensemble de groupes de concepts sémantiquement homogènes en se fondant sur la distance sémantique inter-concepts et la distance intra-concepts. Ainsi un domaine d_l est un ensemble de concepts tel que la similarité entre deux ensembles de termes $sem(c_i, c_j)$ est maximale et $sem(c_i, c_k)$ est minimale, pour tout $i \neq j, i \neq k, c_i, c_j \in d_l$ et $c_k \notin d_l$.

La mesure de distance sem doit rendre compte de l'écart sémantique entre chaque paire de termes t_{c_A}, t_{c_B} appartenant respectivement aux concepts c_A, c_B, pour tout $A \neq B$. La similarité sémantique peut être facilement comprise comme *combien un mot A est lié au mot B ?*. La détermination des similarités sémantiques est souvent évoquée dans les applications du traitement du langage naturel (TLN). Plusieurs types de mesures de similarités existent :

- les mesures sémantiques distributives (statistiques) : se fondent sur l'analyse de texte et sur l'hypothèse que les mots fréquemment présents ensemble sont liés sémantiquement. Elles sont utilisées dans l'évaluation de proximités entre textes. Citons la méthode LSA (*Latent semantic analysis*) [Lan06], PMI (*Pointwise mutual information*) [Bou09], SOC-PMI (*Second-order co-occurrence pointwise mutual information*), NGD (*Normalized Google distance*) [CV07], etc. ;
- les mesures sémantiques à base de connaissance : elles utilisent les ontologies, taxonomies, dictionnaires, thésaurus, etc., pour les comparer. Deux types sont à distinguer :
  - les mesures sémantiques fondées sur des graphes (topologiques) sont utilisées pour comparer des concepts ou des concepts définis dans un graphe mais aussi pour comparer des concepts définis dans une taxonomie ou un graphe sémantique. Elles peuvent être basées nœuds ou arcs. Parmi les méthodes à base de

noeuds, citons les méthodes de Resnik [Res95], de Jiang et Conrath [RJY15], les méthodes DiShIn (*Disjunctive Shared Information between Ontology Concepts*) [CS11] et *Align, Disambiguate, and Walk* [VJS+14]. Les lecteurs intéressés par plus de détails sur les mesures de similarité sémantique peuvent consulter les articles proposés par [EAM14], puis récemment par [Far19] ;

— les mesures sémantiques logiques sont généralement utilisées pour comparer des expressions plus complexes et ce, lorsque les informations définies sur les éléments à comparer (concepts ou instances) ne peuvent pas être réduites à un graphe.

La synthèse de ces mesures est exposée dans le tableau 3.4.

La méthode de classification non supervisée doit prendre en compte la hiérarchie sémantique entre les concepts. Nous avons appliqué l'algorithme de clustering hiérarchique avec lequel les groupes sont construits selon des critères de proximité sémantique. Nous obtenons à ce stade, l'ensemble des domaines $d_1, d_2, ..., d_P$ où P est le nombre minimal de domaines obtenus par l'algorithme de clustering hiérarchique. Contrairement à l'algorithme k-means, le nombre de groupes n'est pas fixé à l'avance et il résulte de la construction récursive de la hiérarchie des domaines. Chaque domaine $d_i, avec i = 1..P$, est composé d'un ensemble de concepts sémantiquement proches $\{c_1^i, c_2^i, ..., c_I^i\} \subseteq \mathcal{C}$, mais qui peuvent être hétérogènes en terme de contexte. Pour assurer une cohérence contextuelle de ces concepts pour un même domaine, nous procédons à une étape de désambiguïsation des synsets associés aux concepts.

Désambiguïsation des domaines

À l'étape précédente, nous avons obtenu un ensemble initial $\mathcal{D}$, composé d'un nombre minimal de domaines. Chaque domaine pourra encore se décliner en sous-domaines par élimination de l'incohérence contextuelle. Pour lever l'incohérence contextuelle qui peut y avoir entre concepts, nous allons nous appuyer sur les synsets. En effet, Les instances des synsets sont des regroupements de mots synonymes qui expriment le même concept. Ainsi, à tout concept c_i lui est associé un ensemble de mots noté $synsets(c_i) = \{t_1^i, t_2^i, ..., t_K^i\}$, où t est synonyme du concept c_i. Par ailleurs, deux objets visuels différents peuvent être

TABLE 3.4 – Mesures de similarité sémantique

Similarité	Classe	Description
Tversky (Tve) [Tve77]	Topologique basée arcs	La similarité entre deux objets est exprimée comme le nombre pondéré de propriétés en commun, auxquelles on retire le nombre pondéré de propriétés spécifiques à chaque objet. Il propose donc un modèle de similarité non symétrique, appelé modèle de contraste basé sur l'intersection et la différence entre deux concepts $(c1, c2)$ en terme de taxonomie. À noter que la position des termes dans la taxonomie et le contenu informatif du terme sont ignorés. La métrique de Tve est $sem_{Tve} = \frac{\lvert c1 \cap c2 \rvert}{\lvert c1 \cap c2 \rvert + \alpha \lvert c1 - c2 \rvert + \beta \lvert c2 - c1 \rvert}$
Leakcock et Chodorow (LC) [LC98]	Topologique basée arcs	la mesure LC a pris en compte la profondeur maximale de la taxonomie et a proposé la métrique suivante : $sem_{LC}(c1, c2) = -log \frac{len(c1,c2)}{2 * DeepMax}$
Hirst et ST-Onge (HSO) [HS98]	Topologique basée noeuds	La mesure de HSO calcule la relation entre les concepts en utilisant la distance entre les noeuds du concept, le nombre de changements de direction du chemin reliant deux concepts $(c1, c2)$ et l'admissibilité du chemin en se fondant sur WordNet. Un chemin autorisé est un chemin qui ne s'éloigne pas de la signification du concept de source et devrait donc être pris en compte dans le calcul de la parenté. Soit d le nombre de changements de direction dans la trajectoire qui relie deux concepts $c1$ et $c2$, soient C et k deux constantes, et $sem_{HSO} = C - SP - k * d$.
Wu et Palmer (WP) [WP94]	Topologique basée noeuds	C'est un score qui tient compte de la position des concepts $c1$ et $c2$ dans la taxonomie par rapport à la position du substitut le moins courant $(c1, c2)$. Il suppose que la similarité entre deux concepts est fonction de la longueur (Len) et de la profondeur$(Depth)$ du chemin dans les mesures basées sur le chemin. $Sem_{WP}(c1, c2) = \frac{(2 * Depth(LCS(c1,c2)))}{(Len(c1,c2) + 2 * Depth(LCS(c1,c2))}$ où $LCS(c1, c2) =$ Noeud le plus bas de la hiérarchie qui est un hyperonyme de $c1, c2$

représentés par un même concept. Cependant, la cohérence des objets visuels au sein d'une même image peut être vue comme la vraisemblance qu'un objet soit présent dans certaines scènes mais pas dans d'autres. Cette information à caractère statistique est souvent liée aux co-occurrences entre différents objets au sein d'une même image ou aux occurrences d'un objet dans une image donnée. C'est une information importante pour l'interprétation d'images car elle peut permettre de déduire des connaissances de plus haut niveau sur une image donnée et assurant désormais la cohérence de son interprétation. Pour ce faire, nous

avons choisi BabelNet [NS12] qui est un dictionnaire encyclopédique multilingue et un réseau sémantique intégrant automatiquement plus de douze ressources qui sont les plus utilisées pour construire les synsets et la désambiguïsation. Nous avons calculé par la suite une matrice de co-occurrences basée sur la distance entre les synsets des concepts telle que $dis(s_{c_i}, s_{c_j}) = 1 - sem(s_{c_i}, s'_{c_j})$, pour tout $s \in synsets(c_i)$ et $s' \in synsets(c_j)$, avec $i \neq j$, $i, j = 1..I$. La matrice de co-occurrences est parcourue pour fusionner les concepts dont les synsets sont très proches. Enfin, nous obtenons ainsi des sous-domaines déclinés du domaine initial. Les sous-domaines résultants sont caractérisés par des ensembles disjoints de paires $(c_i, synsets(c_i))$.

Construction d'un graphe d'objets visuels

La phase de construction de graphe d'objets visuels permet de créer des index hiérarchiques visuels et représentatifs des domaines. Elle fait la transformation d'un ensemble d'objets répartis selon les différents domaines en un ensemble de graphes dont les noeuds correspondent aux objets et les arcs correspondent aux similarités visuelles. Étant donné qu'un domaine est représenté par un vocabulaire de concepts, les objets représentant chaque concept sont organisés en sous-graphe fortement connexes. Les sous-graphes sont ensuite résumés par les objets les plus significatifs de chacun d'eux. Ce sont les objets ayant une similarité visuelle maximale qui sont désignés comme objets représentatifs. La similarité visuelle se calcule en fonction du type du descripteur visuel extrait des images et disponible dans les collections de descripteurs visuels introduites au niveau du chapitre 2. Le mécanisme d'extraction des descripteurs visuels des objets est implémenté en utilisant les bibliothèques jFeatureLib [2] et LIRe [3]. La similarité inter-objets représente la moyenne des mesures de similarité adaptées à chaque descripteur. L'extraction des descripteurs et le calcul de similarité seront tous les deux détaillés au niveau du chapitre 4.

2. http://jfeaturelib-api.locked.de/v1.6.0/
3. www.lire-project.net/

3.4 Enrichissement des domaines et module ontologique

Jusqu'à présent, nous n'avons pas encore exploité les informations contenues dans les fichiers d'annotation des images. Les annotations apportées aux images consistent en un ensemble de phrases descriptives de la scène d'intérêt dans l'image. Un corpus de phrases, $corpus_i$ est associé à chaque domaine d_i identifié par l'étape précédente. L'analyse de l'ensemble des fichiers d'annotation distingue des phrases grammaticalement correctes et des phrases sans verbes. Afin d'évaluer la qualité du corpus textuel obtenu, nous avons fixé deux critères : lexical et thématique [FR07]. Pour effectuer l'analyse thématique, des outils automatiques dédiés ont été comparés. Il s'agit de *Computer Aided Qualitative Data Analysis Systems* (CAQDAS [4]) où plusieurs outils CAQDAS existent ; parmi eux, citons *QDA Miner Lite* [5], *Coding Analysis Toolkit* [6], *Computer Aided Textual Markup and Analysis (CATMA)* [7]. Notre choix s'est porté sur CATMA compte-tenu de sa popularité et de sa performance. L'analyse lexicale du corpus consiste à décomposer une chaîne de caractères en entités lexicales puis de leur appliquer un analyseur lexical. L'analyseur lexical est une fonction qui contient toutes les informations sur les séquences de caractères qui peuvent être contenues dans les entités lexicales. Les deux analyses sont réalisées séquentiellement et permettent d'éliminer les bruits du corpus et de reconnaître les mots et les caractères spéciaux. Les paires $(d_i, corpus_i)$ seront nos données d'entrée pour la phase de construction du module ontologique $\mathcal{OM}_i$ associé. La figure 3.12 décrit les détails de la construction du module ontologique organisé en cinq étapes.

Les étapes (1) et (2) permettent de dégager le vocabulaire nécessaire pour la construction de l'ontologie à savoir les concepts et les relations ainsi que l'enrichissement. Il s'agit des phases de conceptualisation lexicale. Une fois que les données nécessaires pour la construction d'ontologie sont constituées, l'étape suivante (3) consiste en la formalisation permettant de passer d'un vocabulaire source à un vocabulaire cible respectant la structure des éléments d'une ontologie, à savoir, les classes, les sous-classes, les individus et

4. onlineqda.hud.ac.uk/IntroCAQDAS/
5. provalisresearch.com/fr/produits/logiciel-d-analyse-qualitative/logiciel-gratuit/
6. cat.texifter.com/
7. catma.de/

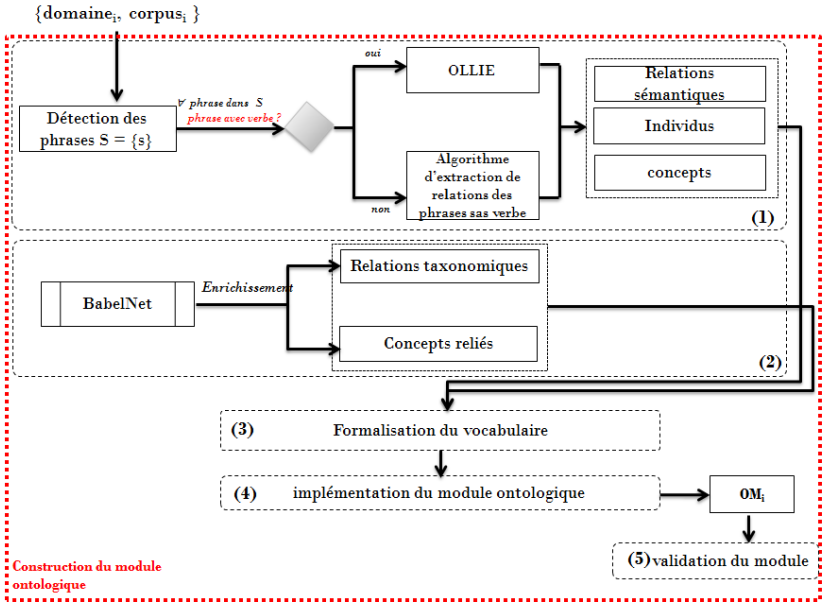


FIGURE 3.12 – Processus de construction du module ontologique en 5 étapes.

les relations. À l'étape (4), une ontologie est générée dans un langage de modélisation ontologique valide. L'ontologie résultante sera évaluée en utilisant un raisonneur (5). Nous reprenons chacune des étapes dans les sections qui suivent.

3.4.1 Conceptualisation du lexique

La conceptualisation du lexique permet l'union des concepts de domaine en un seul ensemble. Ceci inclut les concepts définissant le domaine et les concepts syntaxiquement et sémantiquement reliés. Dans le but de construire des modules ontologiques couvrant le maximum de concepts et de relations du domaine, nous avons été amenés à enrichir le vocabulaire du domaine dont la construction a été décrite dans la section 3.3.3. L'efficacité de l'étape d'enrichissement repose principalement sur le choix de la ressource utilisée. Plusieurs ressources sont disponibles et libres d'utilisation en tant qu'API web. Le choix de la ressource est fortement lié à nos besoins à savoir la complétude, l'extensibilité, la diversité des catégories lexicales présentes et des types de relations ainsi que l'adéquation avec la problématique de désambiguïsation. La comparaison selon les critères cités nous a permis de sélectionner la ressource BabelNet [AMBA17]. Ce dernier est le plus complet avec un maximum de ressources incluses. Il offre une architecture structurée en graphe où les éléments sont des concepts de différentes catégories lexicales, les relations sémantiques,

hiérarchiques et lexicales. Il offre également un outil de désambiguïsation raisonnable en temps de calcul et dont l'efficacité a été approuvée. L'approche de désambiguïsation proposée par BabelNet est basée également sur les graphes. Une heuristique repère des sous-graphes de densité maximale au sein du réseau sémantique multilingue BabelNet2.5.1 [8] et sélectionne des interprétations sémantiques cohérentes. Il associe à chaque noeud de BabelNet une signature sémantique (ensemble de noeuds reliés). Ensuite, il extrait tous les fragments du texte d'entrée pouvant être liés et liste leurs sens possibles. L'étape qui suit consiste à créer une interprétation sémantique du texte basée sur des graphes en liant les sens des fragments candidats extraits. Enfin, il extrait un sous-graphe dense à partir duquel le meilleur sens candidat pour chaque fragment est sélectionné [Nav09]. Cependant, BabelNet est lent quant à la manipulation des listes de concepts reliés. En effet, le grand nombre de ressources intégrées peut être à l'origine d'une explosion du nombre de traitements à prévoir et causer des redondances durant la phase d'enrichissement ; ce qui explique entre autres sa lenteur. La phase d'enrichissement à partir de BabelNet est directement faite en introduisant le concept et ses synsets. Une liste des concepts et des expressions reliées avec les informations qui les concernent et le type de relation est retourné. Comme BabelNet est composé de plusieurs ressources, les relations extraites sont les mêmes que celles offertes par les ressources qui le composent. Le tableau 3.5 présente les relations dérivées de chaque ressource intégrée à BabelNet.

Table 3.5 – Relations dérivées des ressources intégrées à BabelNet

Type de relations	Ressource
Taxonomiques	WordNet
Taxonomiques et sémantiques	Wikipedia, Open Multilingual WordNet, OmegaWiki, WikiData, Wikiquote, VerbNet, Microsoft Terminology, GeoNames, WoNeF, ItalWordNet

Les concepts reliés ainsi que les relations sont utilisés ensuite pour la construction du module ontologique. Nous enregistrons les concepts résultant de la phase d'enrichissement dans une liste et les relations en tant que liste de triplets. Un triplet relation présente deux concepts et la relation qui les connecte se lit ainsi : *« concept A » possède la relation*

8. pyBabelNet

« relation » avec « concept B ». Le filtrage des données résultant de l'enrichissement est une étape qui permet d'assurer l'homogénéité interne du vocabulaire de domaine. Elle repose sur la vérification de la non redondance des éléments et l'appartenance des concepts déduits au domaine en question.

3.4.2 Détection des relations

Au niveau d'une ontologie, les relations représentent les liens que les objets peuvent avoir entre eux. Ce sont les liens entre les classes/sous-classes de concepts. À travers les travaux de la littérature, nous avons établi la hiérarchie représentée par la figure 3.13 qui illustre les différents types de relations entre concepts. Ces relations étant classées en deux types majeurs qui sont les relations taxonomiques et les relations sémantiques. Les relations taxonomiques permettent d'établir une hiérarchisation des concepts. Elles organisent des concepts dans une structure arborescente. Les relations sémantiques permettent de donner du sens à une terminologie. Ce sont en général des relations d'équivalence, de hiérarchie ou d'association.

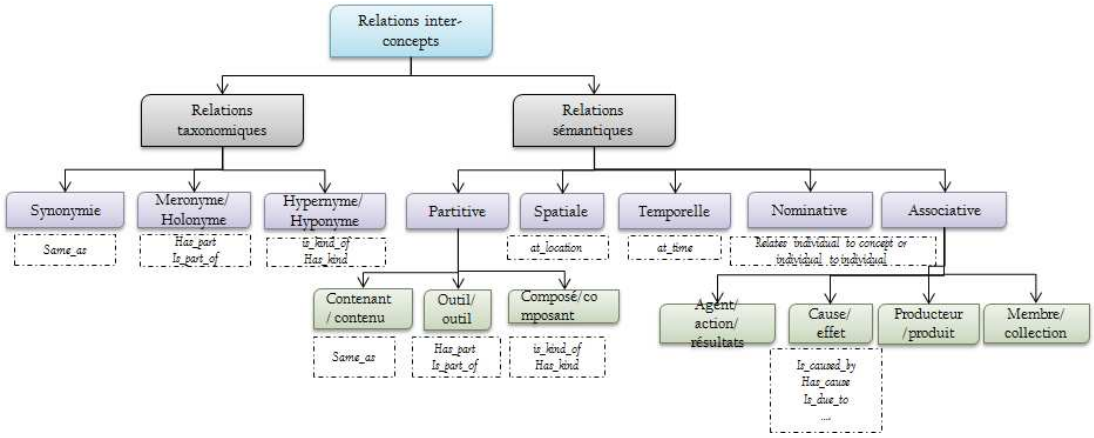


FIGURE 3.13 – Catégorisation des relations taxonomiques et sémantiques entre concepts

Les relations taxonomiques

D'un point de vue mise en place de systèmes, les relations taxonomiques sont déduites de la ressource lexicale durant l'étape d'extraction des concepts reliés. On distingue les :

— relations d'hyponymie ou de spécialisation généralement connues sous *« is kind of »* ou *« is a »*. Par exemple, une enzyme est une sorte de protéine, qui est à son tour

une sorte de macromolécule ;

— relations partitives ou de méronymie décrivent les concepts qui font partie d'autres concepts.

Les relations sémantiques

Les relations sémantiques, non taxonomiques, peuvent être classées en :

— relations nominatives décrivant le nom des concepts qui sont des entités nommées. Par exemple : *« Pays avoir-capitale Paris »* ;

— relations locatives qui décrivent l'emplacement d'un concept. Par exemple : *« lit est-situé-dans chambre-à-coucher »* ;

— relations associatives qui correspondent à des propriétés entre concepts ou à des attributs dans le cas où elles associent à un concept un type de données prédéfini ; des propriétés logiques sont associées à ces relations telles que la transitivité, la symétrie.

L'étape d'extraction des relations sémantiques du corpus textuel pour un domaine est une tâche multidisciplinaire puisqu'elle s'inscrit dans le cadre d'une des applications du traitement automatique de la langue naturelle (TALN) appliquée à la recherche d'information et la fouille de texte. Le TALN est l'ensemble des méthodes et des programmes qui permettent un traitement par l'ordinateur des données langagières en tenant compte des spécificités du langage humain [Cor08]. Le TALN se construit autour de deux catégories de techniques linguistiques et par apprentissage statistique. Les techniques linguistiques consistent à se doter d'une représentation, une modélisation des langues et des données langagières. Elles permettent de développer une recherche linguistique pure de modélisation des données langagières ; ensuite, de définir des modèles et des algorithmes en conséquence. L'objectif de cette étape est d'analyser et filtrer le corpus textuel dans le but d'en déduire les relations sémantiques entre les concepts. Il s'agit ensuite de stocker ces relations dans des triplets de même structure que les triplets extraits à la fin de l'étape d'enrichissement. Plusieurs outils d'extraction de ces relations existent. Parmi eux, citons OpenNLP, Treetagger, RelEx [FKZ07], Stanford nlp, Opencalais Submission Tools, Ollie, qui sont les plus utilisés à l'heure actuelle. À travers la comparaison de ces outils, nous avons choisi

d'utiliser Ollie nous permettant de le faire évoluer pour une extraction des relations de dépendance dans des phrases sans verbe. L'ignorance de ces phrases dans notre traitement engendre la perte d'informations importantes du corpus (63% des phrases sont sans verbe). C'est pourquoi nous avons proposé d'étendre Ollie afin d'extraire des relations depuis des phrases sans verbes en s'appuyant sur une analyse de la grammaire anglaise. La figure 3.14 décrit le processus d'extraction de relations sémantiques pour des phrases sans verbes. En effet, une phrase sans verbes est décomposée à travers un outil d'analyse de la

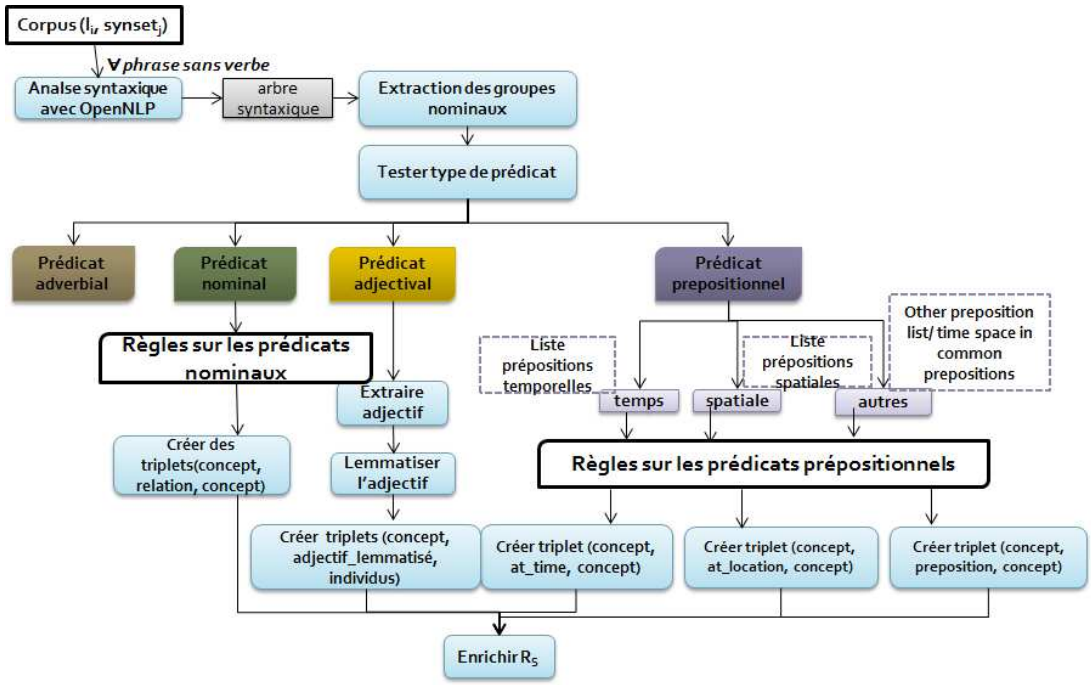


FIGURE 3.14 – Extraction des relations sémantiques pour des phrases sans verbes

langue naturelle en des phrases nominales. Ensuite, nous analysons les différents prédicats afin de détecter leur type en se basant sur l'utilisation ou non des prépositions. Quatre types de prédicats sont susceptibles de se présenter à savoir un prédicat adjectival, nominal, adverbial ou prépositionnel. Les prédicats nominaux ne fournissent pas de relations sémantiques. Nous nous sommes limités également aux prédicats non adverbiaux vu la difficulté de leur détection. Nous avons ensuite réalisé les tests nécessaires pour détecter les concepts en relation et les mots (ou groupes nominaux) responsables de la présence de la relation. Afin de traiter les prédicats prépositionnel, nous avons utilisé une classification des prépositions selon le sens qu'elles expriment en particulier les prépositions utilisées pour la présentation d'un complément de temps, d'espace et les prépositions en commun

ou présentant un autre sens. Remarquons qu'un certain nombre de préposition porte une ambiguïté par rapport au sens souhaité car elles sont utilisés dans différents compléments et peuvent déclencher plusieurs sens. De ce fait, notre base de règles de transformation agit non seulement sur la préposition de ce type mais également sur le contexte du prédicat. Cette étape reste encore un problème non décidable. En ce qui concerne les prédicats adjectivaux et nominaux, l'analyse diffère puisque c'est la lemmatisation qui est la tâche principale responsable de l'extraction des relations. Elle permet de déduire à partir du groupe adjectival des concepts importants tout en gardant la trace de leurs anciennes fonctions dans les phrases nominales. Toutes les étapes précédentes ont permis d'acquérir les connaissances nécessaires à la construction du module ontologique. L'étape suivante est la formalisation des connaissances acquises.

3.5 Représentation multi-dimensionnelle des connaissances

Force est de constater que les connaissances extraites sont de différents niveaux de conceptualisation. Nous retenons dans nos travaux trois niveaux : bas niveau, niveau intermédiaire et haut niveau de conceptualisation. À chaque niveaU correspond un formalisme approprié. Le bas-niveau représente le formalisme de représentation des descripteurs de bas niveau extraits depuis les données. Cela représente en particulier les descripteurs des objets visuels. Le niveau intermédiaire représente principalement les relations contextuelles extraites à partir des objets des images, notamment le graphe de similarité visuelle, les relations spatiales, les relations d'ordres, les alignements, etc. Enfin, le haut niveau vise à représenter principalement les relations sémantiques et enrichies. Les trois niveaux de représentation sont réunis dans une seule entité que nous appelons « patron » par analogie avec le terme anglophone « design pattern » (*cf.* Figure 3.15).

Un patron est une entité composée de trois éléments : un graphe visuel, un graphe contextuel et un module ontologique. Pour toute image I_i, un ensemble d'objets $\mathcal{O}_i = o_{i,1}, o_{i,2}, .., o_{i,K} \subset \mathcal{O}$ est formalisé par :

1. un vecteur de descripteurs $\mathcal{F}_i$ visuels de taille k ;

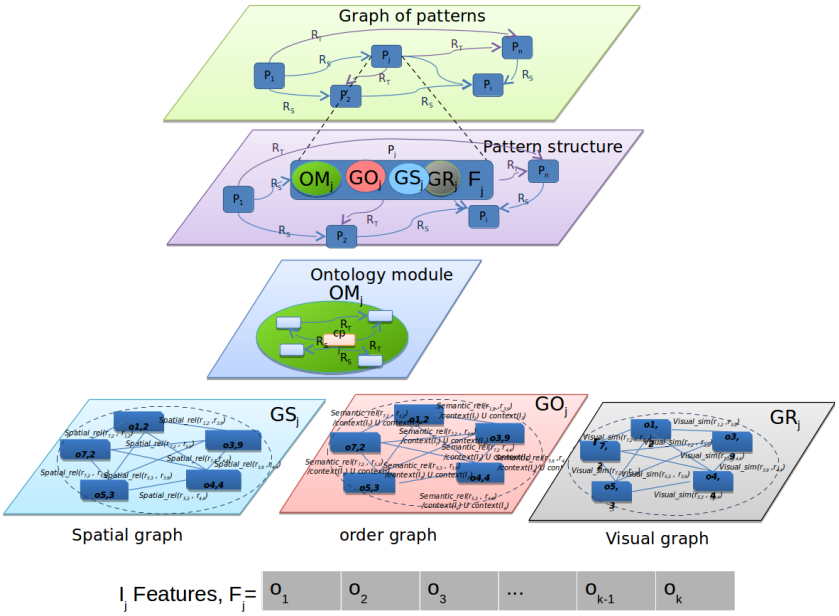


FIGURE 3.15 – Exemple de patron pour la représentation des différents niveaux conceptuels des connaissances

2. un sous-graphe visuels (GR dans la figure 3.15) de taille $K \times M$, avec $M <= N - K$, la cardinalité du sous-ensemble d'objets de $\mathcal{O}$ qui sont similaires à tous les objets de $\mathcal{O}_i$;

3. un sous-graphe des relations spatiales, des relations d'ordre, etc. (GS et GO dans 3.15) ;

4. un module ontologique (OM dans 3.15).

Le patron dispose d'une structure ouverte et pourrait très facilement s'étendre en incluant d'autres connaissances comme, par exemple, un graphe d'émotion, un graphe d'intention ou encore un graphe d'interprétation et d'explication des données à prédire ou à indexer.

Par ailleurs, la formalisation du module ontologique permet de transformer l'ensemble des concepts et des relations taxonomiques et sémantiques représentant le domaine au format adéquat au langage de modélisation d'ontologie. Pour ceci, la première tâche est le choix du langage de modélisation adapté. Les deux standards qui ont été proposés par le W3C sont RDF [9] OWL (*Ontology Web Langage*). OWL étant plus riche que RDF, il est alors choisi comme langage de formalisation par la suite. De plus, OWL a l'avantage d'offrir

9. https://www.w3.org/RDF/

une visualisation des concepts sous la forme d'un graphe comme le montre la figure 3.16 qui affiche les relations sémantiques entre les concepts par le biais de Protégé [10], éditeur libre capable de lire et sauvegarder des ontologies dans la plupart des formats et largement utilisé par la communauté scientifique. La visualisation du détail du graphe est présentée par la figure 3.17.

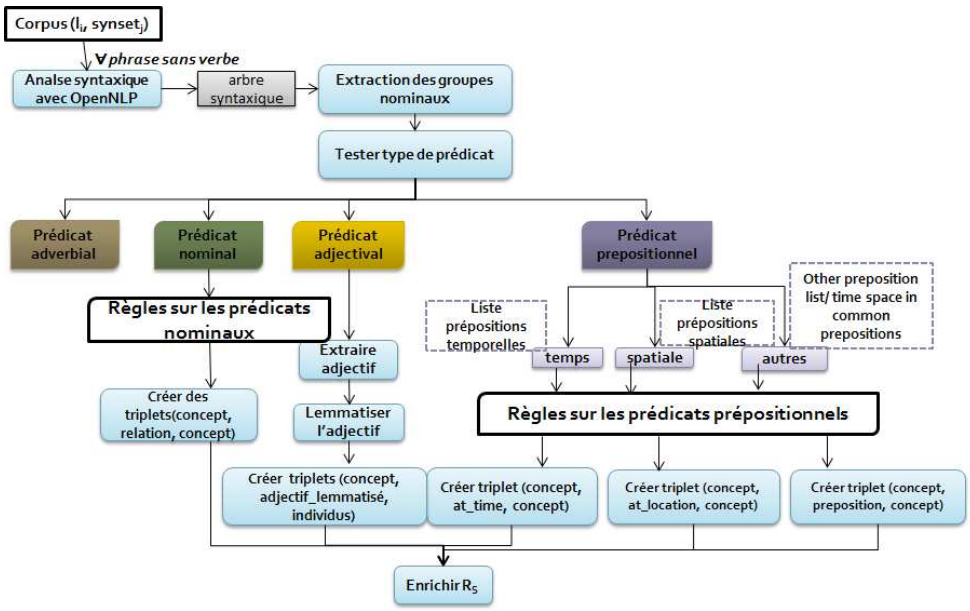


FIGURE 3.16 – Extrait d'une visualisation d'un module ontologique

3.5.1 Raisonnement

Le raisonnement sur des ontologies via les raisonneurs, permet entre autres de déduire de nouveaux faits à partir des faits existants et de garantir leur cohérence. On peut faire appel au raisonneur durant les différentes étapes du cycle de vie d'une ontologie si le besoin se présente. Cependant, nous nous limitons dans notre cas au rôle d'évaluation et durant la phase de déploiement de celle-ci pour déterminer la satisfiabilité des faits utilisés ou alors inférer des relations entre les concepts ou les instances. Dans ce travail, nous nous limitons au niveau terminologique où le raisonnement couvre seulement les concepts et les propriétés et ne s'étends pas aux instances de concepts. Plusieurs raisonneurs ont été développés ; par exemple, Pellet, FaCT++, HerMiT, ELK, etc. Comme nous avons travaillé avec OWL API pour le développement de notre ontologie, il fallait chercher un raisonneur

10. protege.stanford.edu

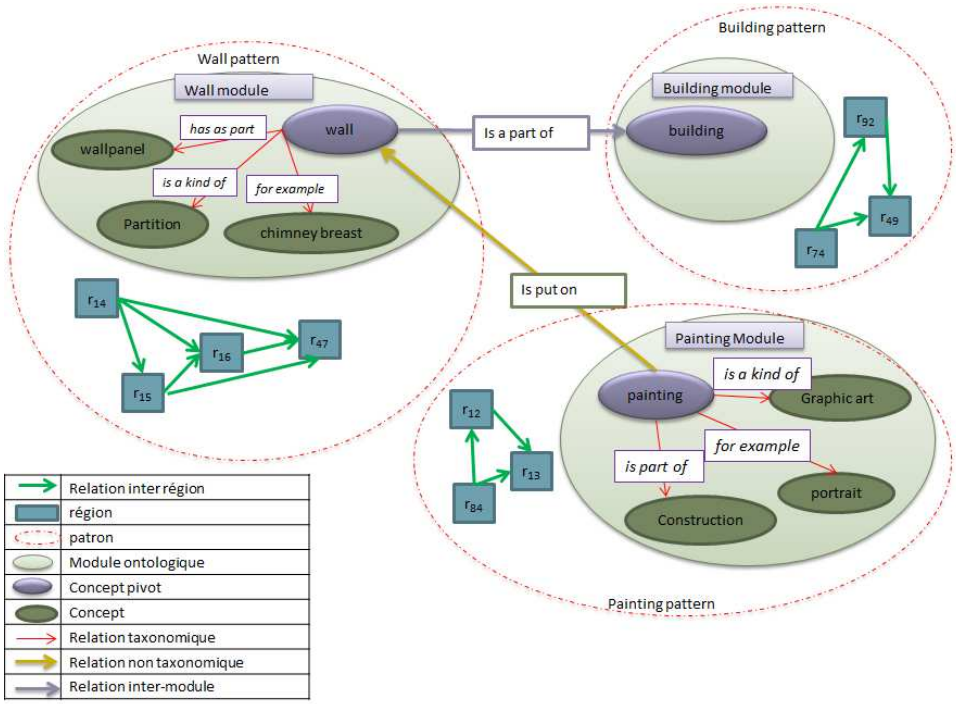


FIGURE 3.17 – Extrait détaillée d'une visualisation modulaire

qui soit compatible avec cette bibliothèque. Nous avons donc sélectionnée le raisonneur HerMiT [11].

3.6 Conclusions

Les descripteurs sémantiques visent en premier lieux à élaborer une ou plusieurs collections de métadonnées (littéralement des données à propos des données). En premier lieux, les descripteurs sémantiques permettent d'indexer sémantiquement les données, ce qui permet d'améliorer considérablement le processus de recherche d'information. En second lieu, comme indiqué par [GR02], ces connaissances permettent une réutilisation facilitée des contenus. Ensuite, elles revêtent un intérêt crucial dans le cadre des entrepôts de données et plus largement pour le web sémantique, qui est supposé fournir aux machines les moyens de comprendre même partiellement le sens des données qu'elles synthétisent [BLHL01]. On peut également considérer que l'enrichissement sémantique des ressources permet de mieux décrire les comportements, les préférences et les besoins d'information recherchés par les utilisateurs, à l'image de ce qui a été initié dans par exemple dans la thèse de

11. hermit-reasoner.com

Olfa Allani. Dans ce chapitre, la phase d'indexation et sa réutilisation dans le cadre de la recherche d'information ont été brièvement abordées mais ont été publiées dans un article de journal [AMBA17]. Nous avons cependant orienté ce chapitre sur la description de la méthodologie d'enrichissement des descripteurs. Notre méthodologie s'appuie sur des ressources ontologiques et sur le traitement automatique de la langue afin d'enrichir les graphes de régions indexés par des descripteurs visuels. Cette correspondance entre bas et haut niveau de description est représentée par un graphe de patrons. Chaque base d'images sur laquelle la recherche est faite, est pré-traitée et le graphe de patrons correspondant est construit pour inclure de manière structurée les concepts sémantiques et perceptifs des images. Ainsi l'étape d'exploration de ces graphes fera l'objet du chapitre 4. Nous tentons, en effet, de généraliser le processus d'exploration en menant une réflexion sur les modèles mathématiques de prédiction. Les modèles de prédiction peuvent se scinder en deux étapes. Une étape de modélisation/apprentissage menée hors ligne et une étape d'interrogation du modèle élaborée en ligne. Si nous prenons par exemple la problématique de recherche sémantique d'images, on peut le voir comme un problème de prédiction dans le but de trouver des images qui ressemblent autant que possible aux images requêtes de l'utilisateur. Dans ce contexte spécifique, la fonction de prédiction vise à trouver les images de la base tel que leur écart avec l'image de la requête calculée sur les propriétés perceptives et sémantiques soit le plus faible possible. Dans cette optique, nous menons actuellement une étude pluridisciplinaire sur la définition des concepts signifiants et signifiés dans une image comme dans un discours transcrit, ou dans du texte orienté utilisateur (en terme d'interprète du concept ou le consommateur du concept). C'est une étude qui croise un point de vue sociologique, psychologique, sémiotique et neurocognitive en s'appuyant sur des données expérimentales du terrain. Toutefois, il reste de nombreuses études qui peuvent être conduites notamment sur l'annotation des images à base des connaissances extraites et sous la forme de résumés textuels et/ou d'images pertinentes. Ces études sont certes pertinentes pour la construction des indexes sémantiques comme par exemple la notion de synthèse d'information [TLD18] et la prise en compte du contexte social et autres. Comme ces types d'information sont orientés utilisateur, ils sont alors décrits avec subjectivité et émotion. C'est donc impératif de faire évoluer notre modèle de représentation pour prendre en compte ces deux aspects et qui sont largement étudiés par [YRMP18]. Une autre

perspective intéressante à mener sur la réalisation de méthodes incrémentales mettant à jour automatiquement les relations entre graphe visuels et graphe sémantique. Enfin, d'un point de vue analyse de traces, il serait également pertinent d'enrichir les parcours avec les métadonnées extraites.

Deuxième partie

Apprentissage supervisé et prédiction sur des données massives, spatiales et temporelles

Chapitre 4

Analyse prédictive : données multimodales massives et temporelles

Sommaire

4.1 Introduction 106

4.2 Prédiction et apprentissage supervisé 106

4.2.1 Cadre formel de l'apprentissage supervisé 107

4.2.2 Modèles prédictifs induits par l'espace des données 111

Modèles prédictifs à base de dictionnaire 111

Modèles prédictifs à partir d'exemples 113

4.3 Modèles prédictifs pour des données massives 114

Formalisation et terminologie de la prédiction 115

Contributions et collaborations 116

4.4 Modèles prédictifs et images spatio-temporelles 118

4.4.1 Données spatio-temporelles simulées 121

4.4.2 Description du simulateur 124

4.4.3 Prédiction de labels pour images satellitaires 127

4.5 Prédiction de séries temporelles issues de capteurs 133

4.6 Prédiction à partir des données de réseaux sociaux 146

4.6.1 Word2Vec et Doc2Vec pour la représentation textuelle 148

4.6.2 Prédiction de séquences symboliques 152
Séquences symboliques par agrégation : la méthode SAX 153
Séquences symboliques par PMVQ 155
4.7 Conclusions et premières perspectives 157

4.1 Introduction

L'objectif principal de l'apprentissage est l'extraction de la connaissance depuis des données grâce à des algorithmes adéquats. Cette démarche est fondamentale à toutes les problématiques scientifiques. Cependant, les données et les requêtes sont généralement propres au domaine d'étude considéré tels que la biologie, la sociologie, la musicologie, l'environnement, etc. Le but de ce chapitre sera de comprendre dans un premier temps quels sont les outils mathématiques génériques mis en commun pour cette démarche d'apprentissage et d'analyser leur implantation algorithmique à des fins de prédiction. Les données s'accompagnent le plus souvent d'informations *a priori* sur leur structure indispensables à l'élaboration des algorithmes. Une dimension fondamentale que nous prenons en considération est le fait de comprendre comment incorporer cette information *a priori* dans les algorithmes d'apprentissage. Ces derniers utilisent davantage différentes branches de l'informatique (intelligence artificielle, base de données, calcul distribué, etc.) et des mathématiques (statistiques, probabilités, analyse harmonique, transformée de Fourier, géométrie etc.). Mais comment en arrive-t-on à la prédiction ?

4.2 Prédiction et apprentissage supervisé

La prédiction consiste à calculer une estimation $\tilde{y}$ de la réponse y à une requête à partir d'une donnée x. Cette réponse appartient à un alphabet $\mathcal{A}$. Cet alphabet peut être un réel qui appartient à un intervalle $[a, b]$ de $\mathbb{R}$. Cette prédiction est alors un problème de régression. Derrière le terme régression se cache un algorithme d'apprentissage basique visant à trouver la meilleure réponse y à l'aide d'une seule variable explicative (modèle univarié) ou bien plusieurs variables explicatives (modèle multivarié). Cette ou ces variables x sont les entrées du problème et ce sont des données que nous connaissons. Dans

un problème de classification, $\mathcal{A}$ est l'ensemble de toutes les classes possibles, par exemple, les chiffres de 0 à 9 pour la reconnaissance d'un chiffre y dans une image x. L'estimation se fait à l'aide de n exemples de données $x_i \in \mathbb{R}^d$ pour lesquelles nous connaissons la réponse y_i. C'est la forme supervisée de l'apprentissage car la réponse y_i est fournie avec la donnée x_i associée. Les applications de l'apprentissage supervisé sont considérables. Cela concerne tous les problèmes de perception (image, signal), de diagnostic médical et plus généralement, d'analyse de données dans toutes les sciences *dures*, sociales ou humaines. L'apprentissage supervisé est à l'origine du renouveau de l'intelligence artificielle (IA) actuelle, propulsée par l'apparition des données massives. Comme il l'a clairement dit, Stéphane Mallat [ZTAM20] dans l'une de ses nombreuses conférences au collège de France en 2019, l'IA est un terme flou, mal défini mais de nature essentiellement « anthropomorphique ». C'est donc l'art de percevoir les valeurs intellectuelles humaines sur les machines même si l'intelligence humaine est quelquefois mal définie. À partir des années 70 et pendant 20 ans, l'IA s'est principalement développée en émulant le fonctionnement du cerveau conscient et en reproduisant ce qui est perceptible humainement. Depuis une dizaine d'année, l'IA connaît un nouvel essor cherchant à imiter des fonctions cognitives humaine relevant du cerveau inconscient et, plus précisément, sur la perception. Les données massives sont les principaux acteurs et propulseurs de l'IA actuelle car les modèles d'apprentissage statistiques et probabilistes classiques ont très vite été dépassés. Ce qui amène à l'usage de l'IA sans effort de compréhension ou d'explication des résultats que l'on prédit.

4.2.1 Cadre formel de l'apprentissage supervisé

Dans le cadre de nos recherches, nous nous sommes focalisés sur la prédiction et l'apprentissage supervisé. Mais, commençons tout d'abord par définir les formalisations mathématiques générales de la prédiction à base de l'apprentissage supervisé que nous emploierons dans le reste de ce manuscrit.

On note $y \in \mathcal{A}$, la réponse ou bien le label à une requête posée par $x \in \mathbb{R}^d$, où $\mathcal{A}$ peut être continu dans le cas d'une régression et discret dans le cas de la classification et x peut représenter une image, un document textuel ou encore une série temporelle. Supposons

que l'on dispose de n échantillons de données aléatoires d'entraînement, $\mathcal{T} = \{(x_i, y_i)\}_{i \leq n}$, indépendants et identiquement distribués de même loi que (X, Y) avec $x_i \in \mathbb{R}^d$, $y_i \in \mathcal{A}$ ((X, Y) est une observation aléatoire et (x, y) une réalisation quelconque parmi les n couples aléatoires). On distingue deux types de problèmes d'apprentissage supervisé. Le premier problème est la classification où les y_i correspondent à des étiquettes telles que le nom des objets dans une image (chat, chien, ciel, voiture, etc.), ou bien une étiquette faisant référence à un groupe ou une structure homogène d'éléments ; ce sont les invariants discriminants les différentes classes. Le second type de problème est la régression où y_i est une fonction $f(x)$ définie sur $\mathbb{R}$. Un exemple de ce type de problème est la prédiction de la consommation énergétique sans connaître les lois physiques sous-jacentes mais plutôt à partir des données. Un algorithme d'apprentissage prend en entrée une donnée x que nous connaissons et pour laquelle il prédit une approximation $\tilde{y}$ correspondant à la réponse réelle y. On cherche alors à trouver la meilleure fonction hypothèse, nommée f, et qui a pour objectif d'approcher les valeurs de sortie Y pour une entrée X. Si la réponse y est unique pour une donnée x alors nous pouvons l'écrire comme une fonction de x, par $y = f(x)$. Estimer $\tilde{y}$ revient à estimer la fonction f qui dépend de d variables (la dimension de la variable d'entrée). Dans le cas de la régression linéaire univariée, la fonction hypothèse f peut s'écrire sous la forme $f(x) = \omega_1 x + \omega_0$. De manière similaire, dans le cas d'une régression multivariée , f prend la forme générale suivante $f(x) = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + .. + \omega_d x_d$.

Le but de l'algorithme d'apprentissage est de trouver les meilleurs paramètres internes tels que $\|\tilde{y}_i - y_i\| \simeq \epsilon$ pour tout $i \leq n$ et ϵ petit. Il faut donc minimiser cette erreur de telle sorte que la précision de la prédiction se généralise. Nous parlons alors d'erreur de généralisation. L'algorithme d'apprentissage supervisé calcule $\tilde{y} = \tilde{f}(x)$ à chaque itération, où la fonction $\tilde{f}$ est choisie parmi une classe $\mathcal{H}$ de fonctions hypothèses possibles. Afin d'illustrer ceci, nous considérons le cas des classificateurs linéaires pour un problème de classification binaire. Nous cherchons à définir un hyperplan $\{x \in \mathbb{R}^d | < w, x >= b\}$ frontière entre deux classes $\mathcal{A} = \{-1, 1\}$ dont la fonction $\tilde{f}$ est sélectionnée dans $\mathcal{H} = \{\tilde{f} : \mathbb{R}^d \to \mathcal{A} = \{-1, 1\} | \tilde{f} : x \mapsto signe(\langle w, x \rangle - b)\}$. Le choix de $\tilde{f}$ est obtenu par la sélection de la valeur optimale des paramètres w et b qui minimise l'erreur sur les n échantillons

d'entraînement. Cependant, deux problèmes apparaissent à ce niveau :

1. la valeur optimale n'est pas toujours unique, et pour certains algorithme, cet optimal peut correspondre à un optimal local et non pas global ;
2. la séparation entre classes n'est pas toujours linéaire ; elle peut être une frontière courbe.

Pour remédier à ces problèmes, certains algorithmes passent par une transformation de l'espace des données par un changement de représentation au lieu d'utiliser directement une classe de famille $\mathcal{H}$ de frontière courbe, ce qui est compliqué dans le plus souvent d'application. La transformation consiste à adapter les données d'origine en passant par un changement de variable défini par une fonction $\phi(.)$ de l'espace d'origine à un espace d'arrivée de dimension d', $\phi : x \in \mathbb{R}^d \mapsto \phi(x) = (\phi(x)_1, ..., \phi(x)_{d'}) \in \mathbb{R}^{d'}$ dans lequel la séparation entre les transformés de x soit linéaire (par un hyperplan). La famille des fonctions s'écrit, suite au changement de variable : $\mathcal{H} = \{\tilde{f} : \mathbb{R}^d \to \mathcal{A} | \tilde{f} : x \mapsto signe(\langle w, \phi(x)\rangle - b) = signe(\sum_{k=1}^{d'} w_k \phi(x)_k - b)\}$. En sachant la fonction de transformation $\phi(.)$, le classificateur est caractérisé par un hyperplan de vecteur normal w sur lequel, ϕ est projeté pour en déduire la classe de x. $\phi(x)_k$ est vu comme un attribut discriminant faible de chaque classe. Leur combinaison linéaire pondérée correspond à un vote majoritaire donnant lieu à la confiance accordée à la prédiction de la classe.

Les algorithmes tels que *Support Vector Machine* (SVM), *Boosting*, les arbres de décision vérifient ces cadres de formalisation. De plus, ils ont tous l'objectif d'aboutir à un algorithme optimal en trouvant les paramètres minimisant l'erreur ou le risque empirique (appelée aussi la fonction de perte) et donc une erreur faible de généralisation. L'erreur empirique aléatoire est calculée par estimation de l'erreur sur la distribution de l'échantillon d'entraînement $\tilde{loss}(\tilde{f}) = \frac{1}{n} \sum_{i=1}^{n} loss(\tilde{f}(x_i), y_i)$. Il faut savoir que la fonction $loss$ peut être symétrique, par exemple $loss(f(x), y) = 1$ si $f(x) \neq y$ et 0 sinon ; comme elle peut être asymétrique, par exemple dans le cas d'applications médicales. Par ailleurs, l'erreur moyenne de généralisation est estimée à partir de la distribution jointe (X, Y) par $R(f) = \mathbb{E}[loss(\tilde{f}(X), Y)]$. En pratique, nous cherchons à optimiser le choix de la fonction $\tilde{f} \in \mathcal{H}$ en minimisant l'erreur $\tilde{loss}$ directement à partir des échantillons d'entraînement en considérant $\tilde{f} = \underset{f \in \mathcal{H}}{argmin}\ \tilde{loss}(f)$. Comme le calcul le montre, cette erreur ou risque

empirique est dans bien souvent des cas biaisée par sa dépendance aux données X_i, Y_i. Elle ne permet pas une estimation *idéale* de l'erreur de généralisation. Par conséquent, le minimiseur du risque empirique a tendance à sélectionner un modèle spécifique très proche des données d'entraînement sans capturer la régularité du phénomène sous-jacent à l'invariance des classes permettant la généralisation du résultat. De surcroît, l'erreur de généralisation est forte alors que l'erreur empirique est faible, c'est le phénomène connu par le sur-apprentissage. De ce problème vont surgir deux contraintes supplémentaires liées à la grande dimension. Quand la dimension des données augmente (par exemple, les images, le texte ou la parole et la vidéo), l'injection de la fonction *a priori* ainsi que la fonction de transformation deviennent très difficiles à élaborer. Or, si nous considérons les modèles tels que SVM, les arbres de décisions, les forêts aléatoires et bien d'autres, ces algorithmes nécessitent de leur adjoindre la définition de connaissance *a priori* pour qu'ils soient optimaux. L'efficacité de ces algorithmes est ralentie par la taille de l'échantillon. En effet, il a été montré dans la littérature [1] que ces algorithmes continuent à être efficaces tant que $log(\mathcal{H})/n \ll \epsilon^2$ (avec ϵ petit) et par conséquent $d'/n \ll \epsilon^2$, c'est-à-dire que le nombre de variables d' doit être largement inférieur au nombre d'échantillons. Dès l'instant où les deux paramètres d' et n peuvent être très élevés alors la démarche de choix de l'algorithme doit être clairement posée. Les algorithmes de *machine learning* tels que les arbres de décision, XGboost, apprentissage à noyau (SVM) nécessitent moins de données mais utilisent de l'information *a priori*. À l'inverse, les réseaux de neurones profonds nécessitent beaucoup de données mais sont capables d'apprendre et de déterminer $\phi(.)$ et w en même temps. En effet, l'ère du *big data* se caractérise en partie par son pragmatisme, où la démarche de modélisation adopte la minimisation de l'usage *a priori* pour la construction des modèles et leur qualité est mesurée par leur pouvoir prédictif. Or, on peut observer des corrélations fallacieuses [CL17] quand le jeu de données d'apprentissage n'est pas suffisamment représentatif de son contexte d'exploitation ou bien quand les données sont multipliées. Ceux-ci augmentent la probabilité d'apprendre des relations qui ne sont que du bruit. Il est à la fois indispensable et nécessaire de créer des systèmes interprétables permettant de comprendre la décision générée. La démarche de modélisation prédictive

1. « The Elements of Statistical Learning », livre de référence sur le *Data Mining*, l'inférence, et la prédiction

se fait généralement en trois étapes. D'abord, les données sont accumulées et des caractéristiques mesurables ou bien issues directement de mesures (depuis des capteurs, ou bien des satellites, des sites web, etc.) sont définies à partir de ces premières. À partir de ces caractéristiques, les variables à prédire sont choisies. Dans un second temps, vient l'étape du choix de l'algorithme d'apprentissage permettant de modéliser les relations statistiques entre les caractéristiques et la/les variable(s) à prédire ; c'est donc l'élaboration du modèle prédictif selon l'espace d'hypothèses. Enfin, en phase d'exploitation, ce modèle est utilisé sur de nouvelles caractéristiques pour inférer la variable à prédire. Les deux premières étapes sont des tâches élaborées hors ligne alors que la dernière est une tâche réalisée en ligne.

4.2.2 Modèles prédictifs induits par l'espace des données

Dans cette section, nous allons nous restreindre à la description de la phase hors ligne. Elle aborde principalement les méthodes de choix de l'algorithme d'apprentissage et par conséquent du modèle de représentation de l'information *a priori*. Deux familles d'apprentissage supervisé peuvent se distinguer par les modèles sous-jacents employés quant à l'explicitaion de la famille des fonctions d'hypothèses $\mathcal{H}$. La première famille cherche en effet à modéliser les régularités et les invariants de l'espace des données du monde réel à partir de combinaisons de fonctions de référence. Les hypothèses, dans ce cas, sont complètement théoriques et n'utilisent jamais les exemples de l'échantillon d'entraînement. Tandis que la seconde famille explicite les hypothèses en s'appuyant sur les exemples de l'échantillon d'entraînement ou bien à un sous-ensemble de celui-ci. C'est donc à travers des fonctions de similarités avec des pondérations que seront calculées les réponses à une observation.

Modèles prédictifs à base de dictionnaire

Il s'agit par exemple de familles $\mathcal{H}$ de distribution de probabilité où la forme de la fonction $f(x) \in \mathcal{H}$ est un mélange de m fonctions de référence de type gaussien. D'une manière générale, $f(x)$ serait une combinaison linéaire de fonctions de référence :

$$f(x) = \sum_{i=1}^{m} w_i \phi_i(x) + w_0 \tag{4.1}$$

où les fonctions ϕ_i sont les fonctions de base et les w_i les coefficients ou paramètres. Comme nous l'avons évoqué dans la section précédente, ces fonctions de base ϕ_i permettent en un sens la re-description des entrées (transformation ou changement de variable) ; et, les paramètres w_i servent alors à sélectionner ou à pondérer l'importance des éléments de cette re-description. Ces méthodes utilisent un ensemble de fonctions de base souvent données *a priori*, telles que les séries de Fourier. Par conséquent elles sont appelées des méthodes à base de dictionnaire. Le nombre de fonctions de base utilisées peut servir à contrôler la capacité de l'espace d'hypothèses et donc à régulariser et converger le risque empirique. Par ailleurs, les fonctions de base utilisées dans ces méthodes ne sont pas nécessairement orthogonales, contrairement à l'analyse de Fourier et à l'analyse fonctionnelle en général[Aze18]. Cela peut poser certains problèmes comme l'unicité et la stabilité des solutions ; de surcroît les coefficients w_i peuvent varier fortement quand l'échantillon d'apprentissage est modifié légèrement. Dans les modèles de mélange, $\phi_i(x)$ est une fonction de densité de probabilité et :

$$f(x) = \sum_{i=1}^{m} \pi_i \mathbb{P}_i(x|\theta_i). \tag{4.2}$$

Dans un réseau de neurones à perceptrons multicouches, les fonctions de base sont typiquement des sigmoïdes où l'apprentissage consiste à apprendre les poids des connexions entre les neurones des couches consécutives qui ne sont rien d'autres que les coefficients de la fonction d'activation choisie :

$$f(x) = f_a(\sum_{i=1}^{m} w_i x_i + w_0) = f_a(w \cdot x) \tag{4.3}$$

avec f_a la fonction d'activation qui peut prendre la forme suivante $f_a = \frac{1}{1+exp(-a)}$.

Dans les modèles graphiques (comme les réseaux bayésiens ou les modèles markoviens), les fonctions sont des dépendances conditionnelles entre les variables où le rôle de l'apprentissage consiste à apprendre les probabilités conditionnelles en chacun des noeuds du réseau. Il faut souligner que cette famille utilise des modèles pour la plupart linéaires. S'ils sont bien adaptés au problème, ces modèles peuvent approcher n'importe quelle régularité cible et sont faciles à interpréter. Cependant, contraindre l'expression du modèle du monde à être linéaire peut aussi conduire à un modèle artificiel qui ne rend pas compte de

la structure des dépendances réelles entre propriétés des données.

Modèles prédictifs à partir d'exemples

Si l'on se place toujours dans le cadre des modèles linéaires, les méthodes à base d'exemples par calcul de voisinage proposent une toute autre représentation des fonctions de décision ayant une forme générique suivante :

$$f(x) = \sum_{i=1}^{m} \alpha_i sim(x, x_i) y_i \tag{4.4}$$

où $sim(x, x_i)$ est une fonction de similarité entre l'observation x et l'exemple x_i. L'hypothèse est en effet explicitée par une combinaison linéaire des étiquettes y_i des points d'apprentissage x_i, pondérée par les similarités. L'apprentissage dans ce contexte, consiste donc à sélectionner les *meilleurs* exemples de référence x_i. Ce choix est fortement dépendant du choix de la mesure de similarité. C'est le problème essentiel des méthodes à base d'exemples où l'effort doit être employé dans le choix des mesures adéquates par rapport à la sémantique des données. Même si la bonne adéquation de la fonction de similarité est bien considérée vis-à-vis de la sémantique du domaine, le problème de sa définition formelle se pose. Malgré l'existence de nombreuses mesures adaptées aux données vectorielles et numériques, comme par exemple la distance, le problème est bien plus ouvert lorsque les données font intervenir des descripteurs symboliques et/ou sont définies dans des espaces non vectoriels, comme le cas des textes. C'est la raison pour laquelle une partie importante des contributions actuelles en apprentissage artificiel concerne la définition et le test de nouvelles mesures de similarité appropriées pour des types de données spécifiques telles que les séquences temporelles, les données semi-structurées, et les données non structurées, etc. Une fonction de similarité est généralement une métrique comme la distance entre deux objets, ayant un certains nombre de propriétés telles que la symétrie, l'inégalité triangulaire. Elle est nulle quand les deux objets sont identiques. Un exemple classique de cette métrique est la norme euclidienne L_2,

$$\|x - y\| = \sqrt{\langle x - y, x - y \rangle} \tag{4.5}$$

explicitée par le produit scalaire entre deux vecteurs. Si on remplace $sim(x, x_i)$ par le produit cartésien $\langle x - x_i, x - x_i \rangle$, la fonction hypothèse

$$f(x) = \sum_{i=1}^{m} \alpha_i \langle x - x_i, x - x_i \rangle y_i \tag{4.6}$$

devient l'hypothèse d'un séparateur linéaire entre n nuages de points tel qu'il a été défini par Vapnick [Vap95] au sens où on cherche une marge maximale pour des points linéairement séparables. Cette observation est très intéressante puisque l'hypothèse dépend directement des données de l'échantillon d'apprentissage et non pas sur un dictionnaire. De plus, quand la forme du produit cartésien est remplacée par une fonction vérifiant les propriétés spécifiques d'une fonction noyau $\mathcal{K}(x, x_i)$, la fonction f devient générique pour un problème de séparation linéaire ou non. Parmi les formes usuelles de la fonction $\mathcal{K}(x, x')$, on trouve le noyau polynomial $(1 + x^T x')^p$, le noyau Gaussien $e^{\alpha x - x'^2}$ ou le noyau sigmoid $tanh(\alpha x x' - \beta)$. Par exemple, le noyau Gaussien est choisi par défaut pour le SVM dans Scikit-learn.

C'est un noyau qui permet de mesurer l'interaction entre les observations et où la surface de décision est fortement influencée par le paramètre α. La complexité de ces algorithmes à base d'exemples est principalement dominée par la taille de l'échantillon d'entraînement et non pas par la dimension de l'espace considéré, comme c'est généralement le cas des algorithmes à base de dictionnaire. Cependant, la qualité de la fonction repose essentiellement sur le choix de la fonction sim ou de la fonction noyau.

4.3 Modèles prédictifs pour des données massives

Dans le cas des méthodes à base de dictionnaires, le choix du bon espace d'hypothèses (le bon dictionnaire), se traduit par le problème de sélection de modèles. Cette sélection peut se faire à la main, par l'expert, ou par des méthodes automatiques, mais l'espace des hypothèses $\mathcal{H}$ est donné *a priori*. Les méthodes à base d'exemples introduisent une souplesse supplémentaire quand l'expression des hypothèses candidates dépend désormais seulement des exemples d'apprentissage. L'adaptation au problème se fait ainsi plus naturellement mais fortement dépendant du bon choix de la fonction de similarité ou de la

fonction noyau permettant la comparaison entre les exemples. Dans tous les cas, l'apprentissage consiste essentiellement à explorer un espace d'hypothèses que l'on peut qualifier de paramétré. Quelle que soit la famille de modèles, la recherche du ou des meilleure(s) hypothèse(s) devra systématiquement contrôler le sur-apprentissage. Or, quand les données sont massives comme, par exemple, les images de grande résolution, les textes ou les séries temporelles, la complexité du problème liée au degré de liberté des modèles, devient très élevée. Par conséquent, les modèles à base de dictionnaires ne sont pas du tout adaptés et les algorithmes classiques d'apprentissage associés deviennent inadaptés. L'alternative à ce problème est de pouvoir réduire la dimension des données en faisant appel à des algorithmes de réduction qui peuvent dégrader la qualité des données. Quant aux modèles à base d'exemples, les algorithmes d'apprentissage sont fortement influencés par la qualité de l'échantillon d'entraînement en terme de représentativité. En d'autre terme, si l'échantillon représente qu'une partie de la réalité du monde, le modèle de prédiction ne sera jamais capable de généraliser face à des observations qui s'écartent de l'échantillon d'entraînement. Alors, est ce déraisonnable de prédire à partir des données massives et dans notre contexte des données complexes ? Nous allons montrer à travers les différentes applications auxquelles nous étions confrontées que la réponse à cette question en deux parties n'est pas évidente et nous montrons que le choix d'un modèle idoine devra être guidé à la fois par les données et par les connaissances sous-jacentes. Nous allons aborder trois types d'application phare qui ont motivées nos travaux de recherche : la prédiction à partir d'images, la prédiction à partir de données textuelles et la prédiction à partir de données de capteurs. Le dénominateur commun à ces trois applications est la dimension spatiale et temporelle des données. Ce sont des contextes où la notion de régularité dans les échantillons d'entraînement n'est pas toujours vérifiée.

Formalisation et terminologie de la prédiction

L'analyse prédictive ou la prédiction est un processus d'optimisation d'une fonction à partir d'un ensemble d'échantillons étiquetés (ensemble de données), de telle sorte que, pour un échantillon donné, la fonction renvoie une valeur qui se rapproche de l'étiquette observée. On suppose que l'ensemble des données ainsi que les autres échantillons non observés sont échantillonnés à partir de la même distribution de probabilité. Nous désignons

les opérateurs $\mathbb{P}$ et $\mathbb{E}$ comme étant la probabilité et l'espérance des variables aléatoires. On désigne par $x \sim \mathcal{D}$ indiquant qu'une variable aléatoire x est échantillonnée à partir d'une distribution de probabilité $\mathcal{D}$ et $\mathbb{E}_{x\sim\mathcal{D}}[f(x)]$ indique la valeur attendue de $f(x)$ pour une variable aléatoire x. Quand la variable aléatoire x dépend du temps et les données sont observées en un nombre fini de points temporels connus, nous nous référons donc à des séries. Plus précisément, en présence de séries temporelles (ou chronologiques), nous considérons qu'elles sont explicitement constituées à la fois (i) des points temporels des observations et (ii) d'observations à ces points temporels. Autrement dit, on considère $x = x(t_1), x(t_2), ..., x(t_N)$, une série temporelle composée d'observations aléatoires indexées par les points temporels $t_1, t_2, ..., t_N$.

Formellement, étant donné une distribution de probabilité des données $\mathcal{D}$, une variable aléatoire $x \sim \mathcal{D}$, un domaine X à partir duquel nous construisons des échantillons, un domaine d'étiquetage Y, et une classe d'hypothèses $\mathcal{H}$ contenant les fonctions $f : X \to Y$, la prédiction revient à un problème d'optimisation, où nous cherchons à minimiser l'erreur de généralisation, définie par la fonction de perte :

$$(Loss) L_{\mathcal{D}}(f) \equiv \mathbb{P}[f(x) \neq h(x)] \tag{4.7}$$

où $h(x)$ représente le véritable label de x. La prédiction temporelle, est une généralisation d'un problème de prédiction classique où le problème consiste à apprendre un prédicteur f capable de prédire la ou les valeurs observées de y_{T_k} futures sur un horizon temporel $T_1, ..T_K$ compte tenu les observations passées $(x, y) = ((x(t_1), y(t_1)), (x(t_2), y(t_2)), ..., (x(t_T)y(t_T))$.

Contributions et collaborations

Depuis ces cinq dernières années, la disponibilité accrue des données a accentué nos collaborations avec des partenaires industriels et académiques. Nous avons mené des travaux sur la prédiction en imagerie médicale et satellitaires dont l'objectif est de mettre en place des modèles prédictifs de caractéristiques spatiales et temporelles. La prédiction de l'ostéoporose à partir d'images 2D enrichie par des images 3D semi-réalistes issues de simulation fut un projet très complexe mené de pair avec l'Institut de recherche en ostéoporose (l'IPROS) pendant plus de trois ans. Nous avons proposé un nouveau simulateur réaliste

permettant de générer des images 3D réalistes à la demande. Ces images permettent de rendre compte de l'évolution spatiale et temporelle de la maladie en fonction du profil du malade puisque la simulation s'appuie sur l'architecture osseuse de ce dernier. De plus, le simulateur pourrait être utilisé comme outil de prédiction de la maladie à partir d'une architecture osseuse saine du patient. La dimension temporelle permet également de prédire l'âge de la présence de la maladie ainsi que son degré de sévérité. Ce travail est détaillé dans les sections 4.4.1 et 4.4.2.

Quant aux images satellitaires, nous avons été confrontés d'abord à des problèmes de prédiction des classes des objets géographiques telles que les plans d'eau, les forêts, les zones urbaines, les bâtiments, etc. Étant donné que la prédiction de la classe d'un objet peut se faire à partir d'images hétérogènes (en format, en taille et en caractéristiques), nous avons proposé une architecture de prédiction distribuée permettant à la fois de répondre au problème de passage à l'échelle et d'accélérer les temps de traitement lié à la prédiction par des réseaux de neurones profonds. Ce travail a été réalisé par Imen Chebbi dans le cadre de sa thèse en cotutelle avec l'université de la Manouba (Tunisie). Une étude comparative entre une architecture classique et des architectures distribuées ont fait le sujet du stage de fin d'étude de Hanen Balti en co-encadrement avec l'Université de la Manouba. Suite à ces travaux, de nouvelles collaborations ont démarrées fin 2019 avec l'Institut de recherche en sciences géographiques et ressources naturelles de l'université de Beijing (Chine) et l'Université de la Manouba. Dans cette perspective, une thèse sur la prédiction temporelle et spatiale de la sécheresse en Chine et en Tunisie sont les sujets d'une nouvelle thèse qui s'appuie sur les résultats déjà approuvés dans le cadre des travaux précédents (le détail est abordé dans la section 4.4.3.

De nouvelles données issues de capteurs sont traitées par cette collaboration. Les capteurs présentent l'avantage de la production quasi-continue des données mesurées. Toutefois, les précisions, et surtout les échelles d'hétérogénité de ces données sont une d'une grande complexité de part leur origine mais aussi de part la complexité du système qu'elles représentent. En collaboration avec l'INRAE, nous avons déjà mis en place des modèles prédictifs à partir de données de capteurs enrichies par des données issues de simulation de modèles physiques complexes. Ces travaux sont décrits dans la section 4.5. Enfin, des

modèles prédictifs à partir de données de réseaux sociaux multilingues et hétérogènes ont été largement étudiés dans le cadre d'un projet FUI mené par Sidahmed Benabderrahmane (post-doctorant). Lors de ce projet, nous avons proposé une architecture de prédiction distribuée alliant des données textuelles et des données spatio-temporelles massives. Le modèle de prédiction a été contraint à la réduction de la dimension des données sans dégrader les résultats de la prédiction. Ce travail a permis de soulever une nouvelle problématique liée au traitement des données textuelles et principalement la question de la détection de l'incohérence qui se révèle un défi scientifique pour de multiples application et constitue le point central de nos travaux de recherche en collaboration avec Aurélien Bossard (MCF), coordinateur du projet JCJC ASADERA, depuis 2017. Ces travaux sont présentés dans la section 4.6. Le récapitulatif de l'ensemble de ces travaux et collaborations sont synthétisés par le tableau 4.1.

Master/Thèse/Post-Doctorat	Thème	Collaborations
	Prédiction à partir d'images	
projet ANR09-BLAN-0029-01 Mataim (2010-2014)	Production massives d'images temporelles pour la prédiction de l'ostéoporose	Frédéric Richard, Anne Ricordeau Map5, Univ.ParisDescartes, IPROS (CHU Orléans)
Imen Chebbi (thèse : 2016-)	Analyse d'images satellitaires massives et hétérogènes pour la prédiction du changement	Thèse en co-tutelle avec Myriam Lamolle (Liasd, P8) et Riadh Farah (La Manouba, Tunis)
Hanen Balti (thèse : 2020-)	prédiction de la sécheresse et des précipitations en Chine	thèse en co-tutelle avec Myriam Lamolle (Liasd, P8), Riadh Farah (La Manouba, Tunis) et Song (Yanfang Sang, Institut de Géographie Chine)

4.4 Modèles prédictifs et images spatio-temporelles

Des objets géographiques comme les lacs, les rivières, les fronts pluviaux, les zones forestières possèdent des propriétés dynamiques très spatiales et présentent un comportement

Master/Thèse/Post-Doctorat	Thème	Collaborations
Prédiction à partir de données textuelles		
Sidahmed Benabderrahmane (post-doc. 2015-2017), projet FUI SONAR	analyse prédictive des comportements des jobs-board	Myriam Lamolle, Mario Cataldi (Liasd, P8) et l'entreprise Multiposting
Faiza Aziz (2018-2019, 6 mois de stage Master2, Tunis) Manel Ben Youssef (2019, 4 mois de stage Master1 ENSTA)	Détection de l'incohérence dans des documents multilingues et cross-lingues, Projet ANR JCJC ASADERA (2017-2020)	Aurélien Bossard (Liasd, P8)

Master/Thèse/Post-Doctorat	Thème	Collaborations
Prédiction à partir de capteurs		
Mahdjouba Kerma (Thèse : 2017-2020), projet Adem Plexifroid	prédiction de l'effacement électrique dans les entrepôts frigorifiques	Minh Hoang (CR INRAE) et Antony Delahay (DR INRAE)

Table 4.1 – Récapitulatif des travaux et collaborations sur la prédiction à partir de données massives et hétérogènes

spatiale et temporel par nature. Il est possible qu'un objet change ses attributs et donc de propriété spatiale tout au long du temps. Par exemple, les objets avec la représentation spatiale pourraient croître, rétrécir (cas de l'espace urbain), changer leur forme (par exemple, quand une ville change ses frontières), se diviser, se fusionner en un nouvel objet ou alors la disparition d'un objet au fil du temps. Un exemple de division peut avoir lieu dans le cas où une forêt est divisée en zone urbaine et zone forestière. Inversement, il existe le phénomène de fusion où des sous-parcelles fusionnent en une seule, comme par exemple suite à une disparition d'un plan d'eau. Ces types d'événement sont réalisés suite à un ou plusieurs phénomènes qui sont liés à des processus géographiques tels que la déforestation, l'urbanisation, la désertification, etc. Typiquement, la disparition d'une forêt est supposée causée par un processus d'urbanisation. Le changement est ainsi constitué par l'altération des propriétés d'un objet persistant, d'une région ou d'une zone de l'espace. En télédé-

tection, le changement est défini comme étant le processus d'identification des différences dans l'état d'évolution d'un objet ou d'un phénomène dans le temps, en l'observant à des moments différents et en différentes positions. Les changements dans les états des objets sont le résultat d'événements et/ou de processus. L'état d'une entité spatiale représente une valeur stable de cette entité dans un intervalle de temps.

Dans un autre contexte, plutôt médical, nous retrouvons ces mêmes observations liées au changement de la structure osseuse à travers le processus dynamique naturel de vieillesse. Les changements peuvent être de l'ordre d'un voxel ou de l'ordre d'une entité topologique. Rappelons d'abord que les entités topologiques sont des plaques, des poutres et des zones de connexions. Comme impact de la vieillesse, un structure labellisée *plaque* à des états précédents, peut par la suite changer de propriétés topologiques pour se réduire à une ou plusieurs poutres. Les poutres, quant à elles, peuvent s'affiner et finir par disparaître. Le processus de vieillesse peut engendrer des perforations aux niveaux des plaques et quand deux perforations voisines s'élargissent au fil du temps, elles fusionnent en donnant lieu à des poutres ou bien à des déconnexions d'entités.

De manière générale, l'analyse et la modélisation des phénomènes topologiques impliquent de connaître |

(i) quels sont les éléments consécutifs qui les caractérisent,

(ii) quelle est la répartition spatiale de ces éléments,

(iii) à quel moment ces phénomènes surviennent.

La composante (i) consiste généralement à identifier les différents types d'entités rencontrées, leurs propriétés géométriques (*i.e.* leur forme) et les attributs permettant d'en calculer la sémantique. L'identification du concept lié à l'objet permet de donner une signification sémantique à un objet (par exemple, une forêt, une zone urbaine, un lac, une plaque, une poutre, une perforation, etc.). Cette identification par le concept permet de distinguer un objet des autres.

Pour la question (ii), le phénomène renvoi à des informations spatiales de localisation, explicitée, par exemple, par la distance ou la position relative d'un objet par rapport à un autre, ou bien son orientation, etc. Cette composante est étroitement liée à la caractérisa-

tion des relations spatiales entre les entités.

La troisième composante fait naturellement référence à la dimension temporelle dans laquelle s'inscrit le phénomène. L'étude peut être statique s'il s'agit d'un *arrêt sur image* limité à un instant ou une période en particulier, ou dynamique si elle prend en compte des changements se produisant sur un ensemble d'instants ou de périodes. Dans les chapitres précédents, nous avons largement abordé les deux premiers points. Le troisième point fait l'objet de ce chapitre. Nous allons l'aborder essentiellement sous les facettes suivantes :

1. le recueil des séries temporelles par simulation quand les données se font rares voire même difficiles à obtenir. Par simulation, les données obtenues sont en général à moindre coût, moins bruitées, et peuvent être générées à la demande. Plus le simulateur est réaliste, plus les données générées s'approchent des données réelles en terme de comportement et de valeurs ;
2. le pré-traitement des séries temporelles est une étape indispensable pendant laquelle se pose la question du dé-bruitage, l'analyse du comportement des séries et, dans certains contexte, la réduction de leur dimension ;
3. l'analyse prédictive sur des séries temporelles massives.

.

4.4.1 Données spatio-temporelles simulées

Comme nous l'avons évoqué dans l'introduction de ce manuscrit, la performance de l'analyse prédictive, en particulier sur des séries temporelles, croît avec l'échantillon des données. Or, dans certains domaines, il est très difficile de recueillir des données réelles. Par exemple, en imagerie médicale, il est inconcevable d'exposer des patients à des radiations fréquentes, même à faible dose, dans le but de suivre l'évolution d'une maladie. La simulation est une approche non-destructive de production de données temporelles. C'est notamment le cas de l'ostéoporose où une radiographie ne représente qu'un état de la maladie à un instant t donné. Un modèle de prédiction de cette maladie basé seulement sur l'analyse 2D des radiographies n'est pas une solution fiable et loin d'être généralisable. C'est, en effet, par le biais de nos études détaillées dans [SBR+04], que nous avons procédé à une étude comparative de plusieurs modèles de prédiction supervisés. Nous avons choisi principalement

deux méthodes d'analyse discriminante à noyau (l'analyse linéaire discriminante (LDA) de l'analyse quadratique discriminante (QDA)) et un arbre de décision. Les deux modèles LDA et QDA sont caractérisés par des règles de Bayes avec des distributions normales pour les densités conditionnelles. La variable Y, celle que nous cherchons à expliquer par les modèles, prend ses valeurs dans $\{0, 1\}$ (classification binaire), 1 pour un sujet fracturé et 0 pour un sujet témoin. L'ensemble $X \subset \mathcal{X}$ représente l'ensemble des variables explicatives ou les descripteurs pertinents parmi tous les descripteurs détaillés dans le chapitre 3 en y ajoutant la densité minérale osseuse (BMD) et l'âge du patient (age), seuls paramètres utilisés dans la routine clinique du diagnostic de l'ostéoporose. Nous disposons également de l'ensemble des données d'étude $D = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ sur lequel, nos classifieurs ont été conçus, d'un ensemble d'entraînement $D_{train} \subset D$ et d'un sous-ensemble d'index de descripteurs $F \subset \{1, .., dim(\mathcal{X})\}$ y compris la BMD et age. La règle de classification

$$\widehat{Y} = R_{(D_{train},F)}(X) \tag{4.8}$$

est utilisée pour l'estimation de y, avec

$$R(x) = \arg\min(x - \mu_y)^t \textstyle\sum_y^{-1}(x - \mu_y) + ln(|\textstyle\sum_y|) - 2ln(\pi_y) \tag{4.9}$$

avec la probabilité *a priori* π_y, la moyenne μ_y et la matrice de covariance $\sum_y$ pour la classe y. L'erreur de chacun des classifieurs est mesurée par

$$\epsilon = P(R(X) \neq Y) = E(|Y - R(X)|). \tag{4.10}$$

Bien que l'hypothèse *a priori* utilisée par chacun des modèles soit différente, les résultats des meilleurs classifieurs obtenus sont similaires et peu concluants. Les raisons de l'échec de ces modèles sont étroitement liées à la taille très insuffisante de l'échantillon d'entraînement malgré l'usage de la technique de ré-échantillonnage (de bootstrap) classique d'Efron [ET97] pour réduire le biais de la taille. L'autre raison, et non pas la moindre, serait la complexité de la maladie qui se traduit par la multitude de configurations imperceptibles sur l'image radiographique 2D, en particulier la détection du changement de l'anisotropie. Toutes ces raisons nous ont conduit à proposer une étude unique et originale qui a suscité

l'intérêt des médecins. Son objectif est avant tout de permettre le suivi de l'évolution de la structure de l'os en 3D dans le temps et dans l'espace et d'aider à la compréhension des facteurs complexes qui contribuent à l'installation de l'ostéoporose. Pour réaliser ce travail, nous avons mis en place, dans le cadre de l'ANR *Mataim*, un simulateur réaliste capable de mimer le processus biologique du remodelage osseux en utilisant peu de paramètres. Des scénarios réalistes de simulation du remodelage osseux à un niveau microscopique fondée sur un modèle stochastique de type *« germ-grain »* ont été élaborés dont le paramétrage a été validé à la fois par des mesures de la littérature mais aussi par les médecins. En effet, selon les théories de Frost [Fro90], un tel processus de remodelage a lieu dans des zones circonscrites appelées *« Bone Multicellular Unit »* (BMU). Un processus au niveau de la BMU est décrit en tenant compte du nombre de BMU travaillant à la surface de l'os trabéculaire ; mais il diffère des autres travaux [LHG+98] par le fait que la probabilité d'activation d'une BMU n'est pas purement aléatoire. Plus précisément, la localisation des BMU à activer sont choisis aléatoirement à la surface de l'os trabéculaire où les endroits sont ceux qui expriment une force d'activation locale dépassant un seuil fixé. Une telle contrainte, inspirée d'un modèle décrit dans [RVHH05], permet de prendre en compte la présence de microfissures et de dommages, supposés aléatoires, ainsi que l'âge de la dernière formation osseuse. Des implémentations ont été effectuées pour des images de synthèses simples en deux dimensions comme première étape de validation, puis étendu par la suite à des images tridimensionnelles réalistes de la base décrite au chapitre précédent. Le but est de pouvoir créer de nouvelles images avec des altérations spatio-temporelle variées simulant divers scénarii d'évolution de la maladie. Sur ces images massives et temporelles, des paramètres micro-architecturaux 3D et 2D sont calculés en fonction du temps. En résumé, via le simulateur, nous avons proposé un outil simple avec peu de paramètres, capable de :

1. générer en masse de nouvelles images 3D et les images projetées 2D associées (fausses radiographie) visualisant l'évolution spatiale et temporelle de la maladie à la fois en deux et trois dimensions ;

2. suivre l'évolution des descripteurs 3D et 2D déjà décrits dans le chapitre précédent et analyser leurs corrélations possibles ;

3. pouvoir simuler différents scénarii à différentes échelles temporelles de la maladie

À chaque pas de temps t_k du simulateur, un ensemble de nouvelles images est obtenu venant s'ajouter aux ensembles d'images précédents. Rappelons que nous disposons d'une centaine d'échantillons d'os au départ qui ont été prélevés sur des cadavres dont on connaît l'âge, le sexe ainsi que l'état initial de la structure osseuse. De plus, le label ostéoporotique ou non est connu. Sur ces échantillons, les images 3D, ainsi que leur radiographie 2D, ont été soigneusement préparées et constituent notre échantillon de données d'étude. Nous notons par V_k, I_k, $\mathcal{F}3d_k$, $\mathcal{F}2d_k$ et $label_k$, respectivement l'image 3D, l'image 2D, le vecteur des descripteurs 3D, le vecteur des descripteurs 2D et le label à l'instant t_K. Notons par $\mathcal{E}_k$ l'ensemble des images obtenues à l'instant t_k. $\mathcal{E}$ est composé de l'union de tous les ensembles obtenus précédemment $\bigcup_{t=t_0..t_{k-1}} \mathcal{E}_t$. Cet ensemble est décrit par les paramètres $\mathcal{F}3d_k$ et $\mathcal{F}2d_k$ que nous avons mis place en s'inspirant de ceux utilisés dans la littérature (*cf.* chapitre 3).

4.4.2 Description du simulateur

Dans l'os trabéculaire, les ostéoclastes résorbent les cavités en forme de soucoupe se déplaçant à la surface en creusant une tranchée d'une profondeur de 40 à $60\mu m$ et couvrant des superficies de tailles diverses allant de $50 \times 20\mu m$ jusqu'à $1000 \times 1000\mu m$ comme mentionné dans [HHM99]. Les paramètres qui caractérisent une forme BMU en 3D sont la profondeur (D), la longueur (L) et la largeur (W). Ils sont considérés comme des variables aléatoires. Pour des raisons de commodité, la longueur est choisie comme $L = 1 + 2(a_L + b_L.U)$, la profondeur est choisie comme $D = a_D + b_D.U$ avec U une variable aléatoire uniforme sur $[0, 1]$. La fréquence d'activation (FA) est souvent utilisée pour mesurer l'activité de la BMU dans l'os trabéculaire. Cependant, FA exprime le taux d'apparition de la BMU sur une lame histologique et non le taux de régénération. Le nombre N de BMU créé par unité de surface S et par unité de temps t est calculé en fonction de AF par la formule suivante :

$$N(t) = \frac{AF \times |S(t)|}{W(t) \times R(t)} \tag{4.11}$$

où $|S|$ le total d'os à la surface, W est la largeur moyenne d'une BMU, R la proportion moyenne de la propagation sur la surface ; bien sûr, tous dépendant du temps t. Cette

même formule peut être exprimée en fonction de la fréquence de régénération OF par :

$$N(t) = OF \times |S(t)| \times \sigma(t) \tag{4.12}$$

avec σ étant la durée de vie moyenne d'une BMU. Connaissant la surface S et son volume V, chaque i^e BMU est localisée à la surface S par son centre de gravité $\mathcal{X}_i = (x_i, y_i, z_i)$ et ses dimensions que l'on note par $E_i = (D_i,\ L_i$ et $W_i)$ à un pas de temps t donné. Par conséquent, l'ensemble de toutes les BMU généré à un pas de temps t peut être vu comme un processus de points marqués (*Marked point process*) $\{\mathcal{X}_i, E_i\}$ et infère un modèle *germ-grain* associé comme suit :

$$\bigcup_i (\mathcal{X}_i(t) \oplus G_i(t)). \tag{4.13}$$

Les $\mathcal{X}_i$ sont les points marqués ou les germes du processus, alors que les G_i sont les grains localisés à la surface S et marqués par leur états E_i. Un processus de Poisson a été retenu comme processus de comptage nous permettant une approximation acceptable de la répartition spatiale des BMU à la surface. Ce choix de processus est justifié par son comportement indépendant par rapport au :

i le nombre de points dans une région finie S est une variable aléatoire $N(t)$ suivant une distribution de Poisson avec une moyenne $\lambda(t)x|S(t)|$ pour une certaine intensité λ,

ii au choix des N points $X_1, \ldots, X_N$, en outre choisis indépendamment et de manière purement aléatoire à la surface S et

iii pour tout $t_0 = 0 \leq t_1 < \cdots < t_k$, les variables aléatoires $(N(t_k) - N(t_{k-1})) \ldots (N(t_1) - N(t_0))$ sont indépendantes.

À chaque tirage des N points, il est indispensable de savoir si un site $\mathcal{X}_i$ pouvait être candidat pour être choisi comme site de régénération de la BMU. L'hypothèse proposée et utilisée par ailleurs dans la littérature [HHM99] suppose que l'absence de signaux ostéocytaires à la surface de l'os attire les ostéoclastes pour la résorption osseuse. Un tel manque peut être dû à une défaillance mécanique, à la mort des ostéocytes (apoptose) ou à des micro-dommages de fatigue (micro-fissures) qui se produisent de manière aléatoire

à la surface de l'os trabéculaire. Par conséquent, la répartition spatiale des positions des BMU dans le temps est limitée à des sites qui expriment une énergie $\|FA\|$ de la fréquence d'activation FA supérieure à un seuil α. Ce seuil est un paramètre défini en fonction du scénario de simulation et qui peut être fixé par les spécialistes. L'énergie est simplement exprimée comme :

$$\|FA(\mathcal{X}_i)\| = f(M(x)), \forall x \in S_{\mathcal{X}_i}, r. \tag{4.14}$$

Autrement dit, l'énergie d'activation est calculée dans une zone centrée en $\mathcal{X}_i$ et qui s'étend sur un voisinage de points de la surface S situés à une distance r. La matrice $M(x)$ est vue comme l'état de minéralisation du site en un point x depuis la dernière régénération (ou formation de l'os nouveau). La fonction f nous permet de paramétrer de différentes manières les scénarii d'altération (vieillesse jusqu'à ostéoporose) ou régénération plus ou moins fréquente pour maintenir un état stable de la matrice osseuse, etc. Ainsi, dans le cadre d'un processus fortement ciblé, l'activation d'un site $\mathcal{X}$ peut être validée, par exemple, lorsque l'énergie $\|FA(\mathcal{X}_i)\| > percentile(\|FA\|, \kappa)$. Si, par exemple, $\kappa = 85\%$ (*cf.* Figure 4.4), le scénario correspond à viser principalement à renouveler les sites les plus vieux. Enfin, les sites sélectionnés pour être régénérés subissent deux étapes successives (*cf.* figures 4.1, 4.2, 4.3).

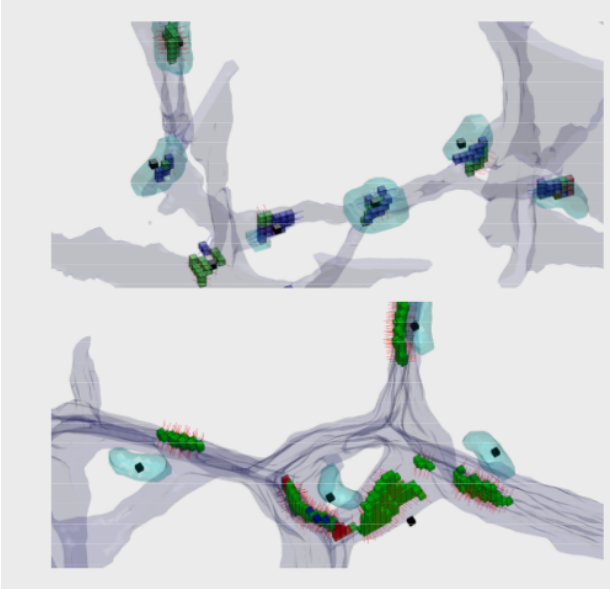


FIGURE 4.1 – Deux étapes illustrées par les images du simulateur 3D : Résorption et formation

Tout d'abord, la résorption de l'os endommagé a lieu au germe $\mathcal{X}_i$, marqué par le nombre

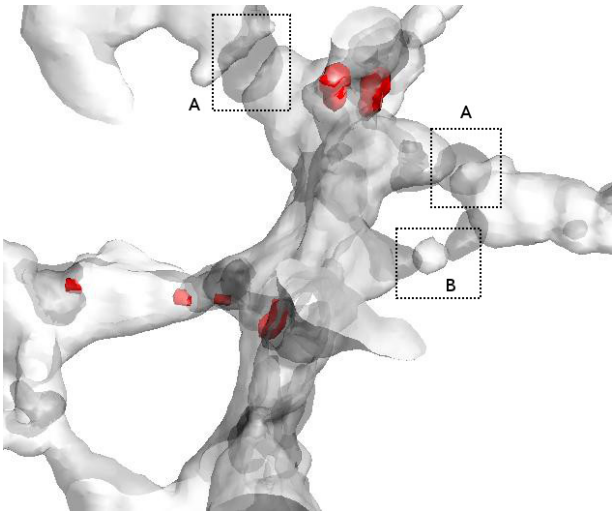


FIGURE 4.2 – Visualisation 3D par le simulateur. zones A : résorption est plus importante que la formation ; Zone B : résorption puis formation. Zones en rouges sont les formes de la zone de remodelage (BMU).

de voxels de la surface de S, $|G_i^r|$ à résorber et soit inférieure à $L \times W \times D$. Cette étape est suivie par le processus de formation sur le même site, paramétré par $|G_i^f|$, le nombre de voxels à former avec de l'os nouveau. La balance osseuse locale $(\mathcal{X}_i)$ est égale à la différence entre la quantité formée $|G_i^f|$ et la quantité résorbée $|G_i^r|$ avec $|G_i^f| = \phi \times |G_i^r|$, $\phi \in [0, 1]$. Si $(\mathcal{X}_i) = 0$ alors le site a été entièrement régénéré (la matrice minérale $M(\mathcal{X}_i)$ est entièrement nouvelle) alors que quand la balance est négative, la perte osseuse peut être irréversible avec l'apparition soit d'affinement des structures, soit des perforations. Le détail d'autres scénarii peut être consulté dans l'article [RM16] du journal CMBBE.

4.4.3 Prédiction de labels pour images satellitaires

Les images de télédétection fournissent des observations sur les objets acquis qui sont exploités pour expliquer la dynamique des processus de changements (forêt vers zone agricole ou bien transition d'une zone agricole vers zone urbaine, etc). Après identification des objets, nous pouvons identifier les objets en mutation qui participent à un tel processus ou bien quel processus a changé l'état d'un objet. Avant de suivre l'évolution des objets, il est essentiel de pré-traiter les images. Les opérations de pré-traitement nécessaires pour les images de télédétection sont le filtrage, la segmentation et l'extraction des descripteurs. Comme nous l'avons expliqué dans le chapitre 2 (section 2.3, p. 46), le filtrage et la

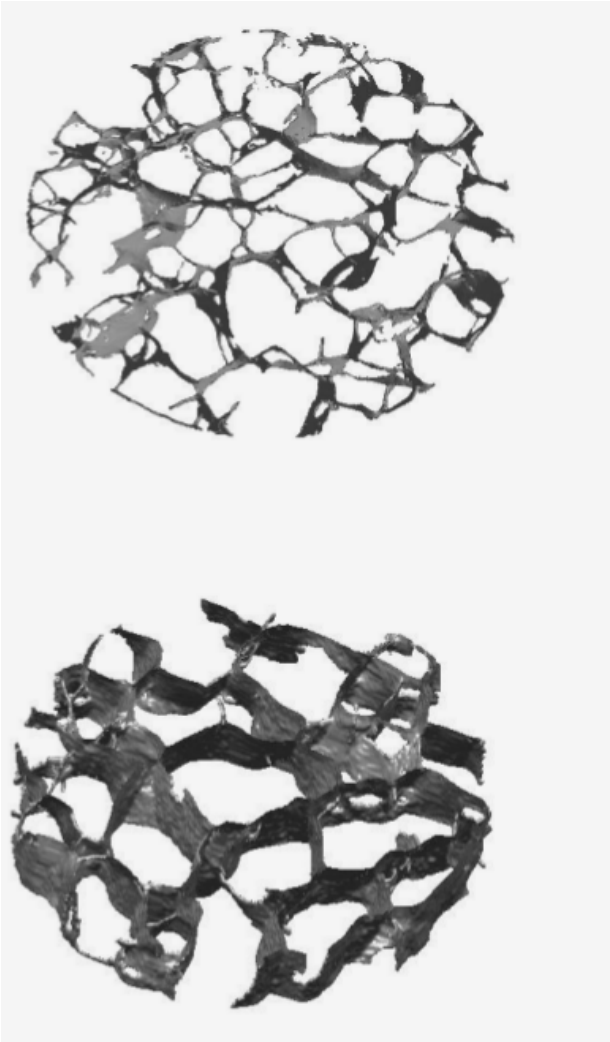


FIGURE 4.3 – Résultat du remodelage en 3D

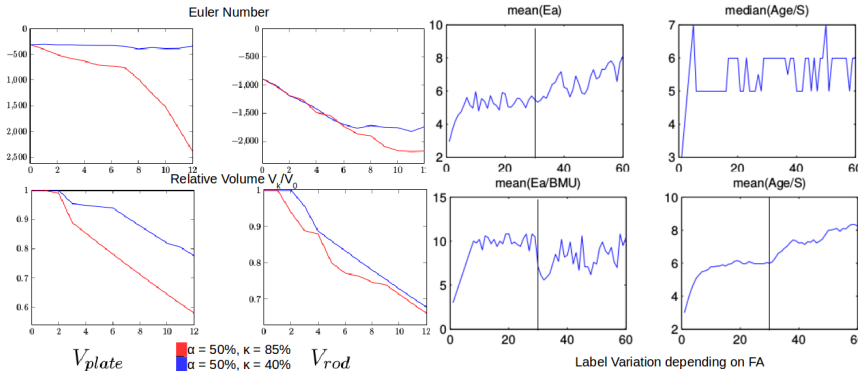


FIGURE 4.4 – Extrait des paramètres du simulateur de remodelage en 3D et dans le temps

segmentation sont deux opérations très coûteuses et nécessitent des ressources machines à la fois pour le traitement en mémoire vive mais aussi pour pouvoir stocker les images bien sûr brutes et traitées pour une exploitation rapide lors de la phase de prédiction.

Force est de constater que les images satellitaires sont de formats très divers. Comme elles sont obtenues par différents satellites d'observation de la terre, plusieurs types d'images sont acquis dont la résolution est étroitement liée à la résolution du capteur. En effet, le capteur possède en général quatre types de résolutions à savoir la résolution spectrale, la résolution spatiale, la résolution temporelle et radiométrique. Ainsi la distinction de deux points voisins sur une image sera plus ou moins précise en fonction de la résolution. Par exemple, les objets spatiaux de petite taille, tels qu'un bâtiment ou une petite parcelle, ne peut pas être distinguée par une résolution spatiale grossière, mais une résolution élevée est cependant nécessaire. La résolution spectrale est définie par le nombre de bandes ainsi que la longueur d'onde de chacune. Deux capteurs spectraux sont distingués donnant lieu à deux types d'images multi-spectrales et hyperspectrales. Avec une résolution temporelle, les capteurs sont dotés de périodes de temps orchestrées par la fréquence de passage du satellite leur permettant de réviser la même zone observée à partir de la même position. En résumé, les types d'images qui en découlent sont les images monospectrales ou panchromatiques (la valeur d'un pixel varie entre 0 et 255), les images multi-spectrales (image composée d'au plus 10 bandes d'images monospectrales, un pixel est donc un vecteur de taille au plus de 10. Si le nombre de bandes est égale à 3, on retrouve les images classiques RVB), et enfin les images hyperspectrales qui contiennent au moins une centaine de bandes de longueurs d'ondes différentes. Enfin, les images satellitaires sont représentées sous forme de matrices contenant les valeurs des pixels. De plus, elles contiennent des métadonnées fournissant des connaissances contextuelles telles que la localisation du satellite lors de la prise de vue, ou bien le système de coordonnées utilisé, ou encore l'éclairage solaire. Des indices, comme ceux de végétation ou ceux de variation d'eau, sont souvent utilisés pour étudier et interpréter la dynamique des phénomènes spatio-temporels. Par exemple, l'indice de végétation NDVI est utilisé pour étudier la structure de végétation, tandis que les indices NDWI et NDSI sont destinés pour étudier respectivement la structure d'eau et la structure des sols. Ces indices utilisés comme paramètres globaux en complémentarité avec les images et sont souvent disponibles en quantité massive et sur des périodes de temps qui varient de la semaine à un mois et ce sur plusieurs années de cohortes. D'ailleurs, certains indices tel que SPEI permettant de prédire les changements climatiques sont disponibles depuis 1900 à jusqu'à ce jour. Face à une présence accrue des

données, les modèles de prédiction orientés données sont à privilégier tels que les réseaux de neurones profonds (DNN pour *Deep Neural Network*). En effet, les DNN continuent à élargir les limites de leurs territoires d'application. Par exemple, les réseaux neuronaux convolutionnels (CNN) sont devenus *de facto* la méthode standard pour la reconnaissance d'images/objets en vision par ordinateur [SLJ+15]. D'autres types de DNN ont également montré des performances exceptionnelles dans divers problèmes d'apprentissage et en particulier la classification d'images. Cependant, le temps d'apprentissage des DNN sur de grands ensembles de données est le principal goulot d'étranglement du flux de travail dans un certain nombre d'applications importantes tels que dans les systèmes de monitoring et de prédiction à partir de gros volumes d'images satellitaires.

Pour minimiser ce temps, la tâche d'apprentissage d'un DNN doit être distribuée. Elle doit être étendue sur un plus grand nombre de machines possibles en distribuant la méthode d'optimisation utilisée lors de l'apprentissage. Bien qu'un certain nombre d'approches aient été proposées pour la descente de gradient stochastique distribuée (SGD), à l'heure actuelle, les approches synchrones de SGD distribuées semblent être les plus performantes à grande échelle. La mise à l'échelle synchrone de la SGD souffre de la nécessité de synchroniser tous les processeurs à chaque étape du gradient et manque de robustesse face à des processeurs défaillants ou en retard. Dans les approches asynchrones utilisant des serveurs de paramètres,l'apprentissage est ralenti par le décalage du serveur de paramètres [LZZ+17] [BH19]. Ces approches exploitent principalement le parallélisme des modèles et peuvent offrir une possibilité d'évolution. Dans ce contexte où le parallélisme des données et des modèles manquent, la thèse en co-tutelle d'Imen Chebbi propose des améliorations de l'architecture distribuée en posant deux hypothèses. La première hypothèse suppose que l'environnement Spark est en mesure d'apporter une solution robuste quant à la distribution des données. Or, la distribution des données n'est pas une solution optimale pouvant ralentir davantage la phase d'apprentissage puisque le coût de lecture/écriture s'élève avec la quantité des données. L'objectif de la seconde hypothèse est de proposer une architecture distribuée de l'apprentissage des DNN et des données. Pour assurer les opérations parallèles, Spark distribue automatiquement les charges de travail, les paramètres aux nœuds Tensorflow et agrège d'une manière itérative les résultats de l'appren-

tissage par les approches synchrones ou asynchrones. Le processus distribué dédié à la prédiction des objets dans les images satellitaires est illustré par la figure 4.5 et se décline en plusieurs étapes. Tout d'abord, les images sont divisés en images de taille optimale

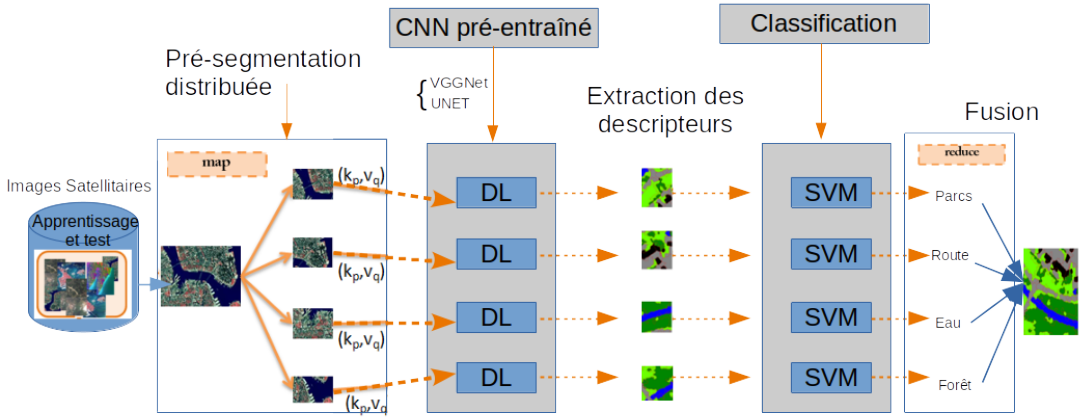


FIGURE 4.5 – Modèle de prédiction distribué d'objets à partir d'images satellitaires[BMC+19]

(144x144). Cette taille est élaborée par une stratégie heuristique alliant le temps de calcul et la performance de prédiction. Les données sont, par la suite, réparties sur un ensemble de noeuds de Spark. L'algorithme de segmentation décrit dans le chapitre 2, section 2.3, à la page 46. Après avoir fractionné les données, nous obtenons un certain nombre d'images ayant une résolution homogène. Ces données sont stockées dans des ensembles de données distribuées résilientes (RDD) en utilisant la technique MapReduce. En utilisant la fonction Map, chaque bande (de taille 144x144) est traitée pour produire des paires de clés-valeurs (kp, vq) comme sorties intermédiaires que nous nommerons par vecteurs RDDs. Lorsque les vecteurs RDDs sont établis, ils sont transmis aux différents réseaux neuronaux profonds pour l'étape d'extraction des descripteurs. Le nombre de réseaux est égal au nombre de noeuds correspondant aux nombre de *workers* que nous avons fixé de manière heuristique. Deux architectures pré-entraînées sont choisies pour l'étape d'extraction des descripteurs à partir des images. Dès lors que les architectures sont déjà pré-entraînées, le temps nécessaire à leur paramétrage se réduit considérablement. De plus, grâce à la technique de transfert de l'apprentissage, le biais de sur-apprentissage est faible.

Concrètement, les paramètres des deux architectures déjà apprise avec Unet et VGGNet, sont réutilisés et ajustés à nos vecteurs RDD pour l'extraction des descripteurs. La der-

	SIRI-WHU			AID			Our database			16- band database		
	Accuracy	Recall	F1-score	Accuracy	Recall	F1-score	Accuracy	Recall	F1-score	Accuracy	Rappel	F1-score
UNET	-	-	-	-	-	-	-	-	-	0.92	0.57	0.70
UNET-SVM	0.73	0.72	0.72	0.78	0.78	0.78	0.84	0.82	0.82	0.94	0.92	0.92
VGGNET-SVM	0.73	0.68	0.70	0.67	0.50	0.55	0.79	0.80	0.79	-	-	-
Seg-UNET-SVM	0.91	0.89	0.90	0.93	0.92	0.92	0.95	0.93	0.94	0.94	0.92	0.92

TABLE 4.2 – Comparaison des résultats sur des bases d'images à 3 et 16 bandes.

nière couche de sortie de chacun de ces deux réseaux est remplacée par une classification partielle (ou locale) SVM via la bibliothèque MLib de spark. Afin d'adapter l'algorithme SVM à notre classification, nous avons utilisé l'approche *un contre un*, qui consiste en une série de classifieurs appliqués à chaque paire de classes. Ainsi, $n(n-1)/2$ threads sont appliqués à chaque paire de classes. La stratégie *max-wins* a été utilisée lorsque tous les classificateurs ont été balayés en ajustant les hyper-paramètres C de régularisation et gamma de pénalité. Le paramètre de régularisation indique à l'optimiseur notre tolérance aux erreurs de classification et le paramètre gamma décrit l'impact d'un exemple mal classé. En d'autres termes, avec un gamma faible, les points éloignés de la ligne de séparation plausible sont pris en compte dans le calcul de marge de séparation. Pour la sélection des hyper-paramètres, nous avons utilisé la méthode de validation croisée *k-fold*. Enfin, une étape de fusion de l'ensemble des classifications partielles est effectué pour étiqueter les objets sur l'image source. Force est de constater que cette dernière étape consiste en la phase « Reduce » de l'algorithme MapReduce augmentée par une tâche spécifique de gestion des étiquettes en conflit pour un même objet. Pour évaluer ce processus distribué de traitement, nous avons comparé trois modèles d'extraction et de classification : (1) un modèle Unet seul, (2) Unet suivi d'un SVM et (3) VGGNet suivi d'un SVM. Les résultats obtenus sont résumés par le tableau 4.2 où nous constatons que le modèle correspondant au processus que nous avons mis place surpasse largement les autres modèles. Par ailleurs, ce modèle a été également comparé à des modèles de la littérature et permet d'atteindre les mêmes performances que le modèle de [KLSS17] (93,6% en précision) et dépasse largement le modèle de [CYY+18] (83% précision), ainsi que celui de [NHP13] (80% en précision). La figure 4.6 illustre le résultat de la prédiction des classes à partir d'images satellitaires 3 bandes, et la figure 4.7 à partir d'images 16 bandes.

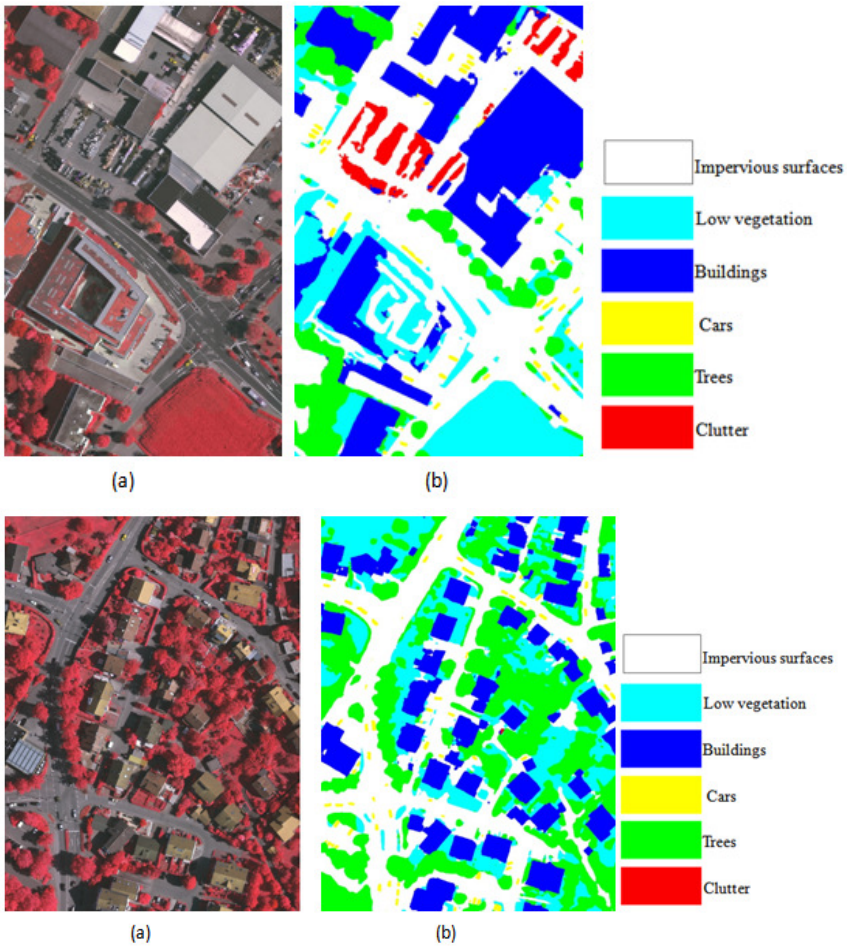


FIGURE 4.6 – Images satellitaires 3 bandes et exemples de prédiction[Mel19b]

4.5 Prédiction de séries temporelles issues de capteurs

Dans le cadre d'un appel à projet ADEME[2], l'INRAE (ex-IRSTEA, Anthony) propose d'apporter une solution écologique pour réduire efficacement la consommation électrique. Dans ce cadre, nous avons été sollicitées pour collaborer avec l'équipe Génie des procédés frigorifiques afin de contribuer à la mise en place de modèles de prédiction de la surconsommation électrique pour la recommandation d'un effacement idoine.

Avant de détailles des modèles mathématiques, nous allons tout d'abord expliquer très rapidement le contexte applicatif (ce qu'est un effacement électrique dans un usage particulier qui est la chambre frigorifique) justifiant l'intérêt de ce projet, nommé *Flexifroid*. L'énergie

2. ademe.fr

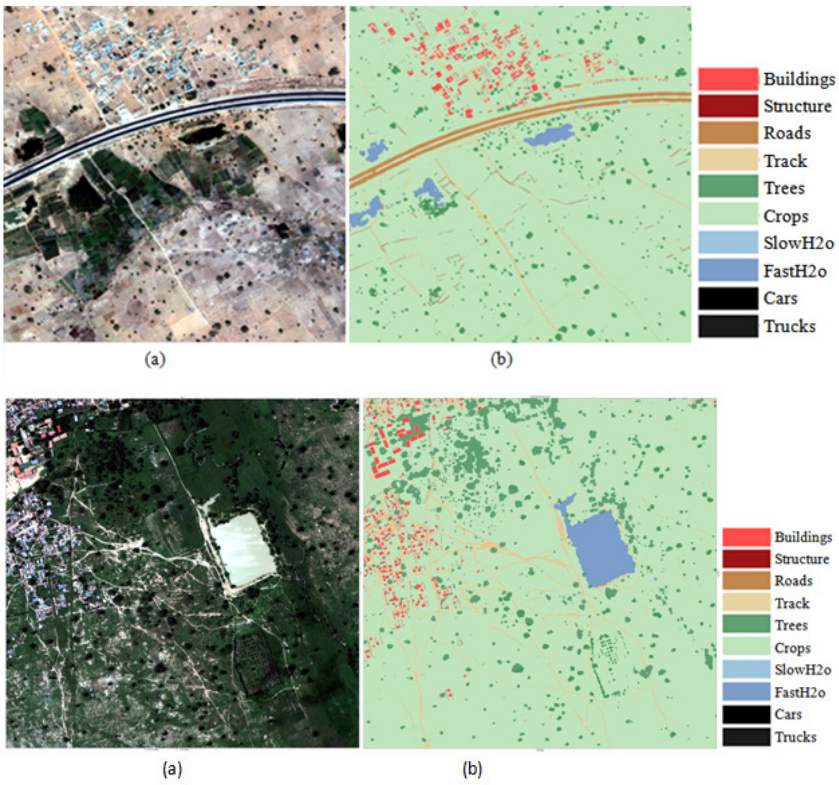


FIGURE 4.7 – Images satellitaires 16 bandes et exemples de prédiction

électrique représente une part importante de la consommation finale à l'échelle mondiale (environ 20%). Sa production constitue un enjeu fondamental de son développement mais doit tenir compte des nouveaux challenges environnementaux. Dans ce contexte, et compte tenu des moyens humains et matériels importants mis en jeu, l'industrie électrique doit proposer des solutions pour optimiser la gestion énergétique des réseaux afin d'assurer un équilibre permanent entre production et consommation d'électricité (entre l'offre et la demande). De plus, la production de froid représente, quant à elle, une part importante de la consommation électrique mondiale, soit 17%, et correspond à 10% des émissions de gaz à effet de serre [CDP15]. L'explosion industrielle aggravée par le réchauffement climatique provoquent à eux seul une augmentation exponentielle de la consommation électrique ces dernières décennies. Rien qu'en Europe, il existe 1,7 millions de chambres froides et d'entrepôts frigorifiques qui représentent la part principale du secteur du froid industriel. Ces équipements et infrastructures ont une capacité de 60 à 70 millions de m^3 de volume de stockage avec une consommation d'énergie spécifique entre 6 et 240 $kWh.m^{-3}.année^{-1}$ pour le stockage de produits surgelés [EFH+14]. Par ailleurs, il faut noter que les produits

surgelés jouent le rôle de batteries électriques permettant le stockage de l'énergie sous forme thermique (stockage d'énergie par inertie thermique). Il convient donc de procéder à un effacement électrique, c'est-à-dire à couper l'électricité pendant les périodes de surconsommation. Cette thèse, fort intéressante d'un point de vue écologique, s'avère très difficile à déployer à grande échelle dans les entrepôts frigorifiques et les chambres froides industriels pour des raisons financières. En effet, une coupure de l'électricité signifie pour les industriels une remontée potentielle de la chaleur et, par conséquent, la dégradation de la qualité des articles surgelés. Cet été de fait conduit, tout d'abord, à l'exposition des clients à des risques sanitaires conséquents et, avant tout, à une grosse perte de leur chiffre d'affaire. Il convient alors d'avoir une bonne connaissance du comportement du système, en termes de fluctuation de température et de consommation électrique, sans l'appui des données réelles des industriels, afin d'approuver les modèles d'effacement et d'envisager, par la suite, leur déploiement à petite puis grande échelle. Le défi de ce projet a été double. Il a été indispensable de substituer l'environnement frigorifique réel par un dispositif réaliste permettant non seulement de produire des données dans un contexte réaliste mais aussi et surtout de pouvoir couvrir les différents comportements physique, thermodynamique et électrique face à des paramètres environnementaux (tels que la température extérieure, la saison, le week-end, les vacances, la localisation, etc.), des paramètres de chargement (le taux d'occupation de la pièce, la répartition des produits, etc.), des paramètres géométriques (la dimension de la pièce), des paramètres techniques liés aux composants pour la production du froid et, enfin, des paramètres aléatoires (fréquence de l'ouverture des portes, la durée d'ouverture, etc.). La première collaboration sur ce projet date de fin 2018 où les premiers jeux de données issues d'une vraie chambre froide expérimentale (installée à l'INRAE) sont collectés grâce aux différentes manipulations menées par la doctorante Mahdjouba Kerma [3]. Ces données ont été appareillées au peu de données recueillies depuis des chambres froides réelles appartenant à l'un des partenaires industriel de ce projet. Cette étape a permis de calibrer certains dispositifs et paramètres liés aux appareils de la chambre froide expérimentale. Une fois le dispositif largement calibré, nous avons mis en place le modèle de prédiction de l'effacement sur la chambre froide expérimentale pour différents scenarii. Une chambre froide est modélisée par un système dynamique multiva-

3. en 2ème année de thèse au moment de l'écriture de ce mémoire

	Nom	Description
Sortie	$Temp_{In}$	Température intérieure
	$Temp_{Air}$	Température de l'air
	$Temp_{Prod}$	Température du coeur des produits
	Ep	Consommation
Entrée	$Temp_{out}$	Température extérieure
	DR	période d'effacement
	Def	période de dégivrage
	$\delta(DR)$	Le temps écoulé depuis le dernier effacement
	$\delta(Def)$	Le temps écoulé depuis le dernier dégivrage
	$Compressor$	On/off du compresseur

TABLE 4.3 – Entrées/sorties du système dynamique d'une chambre froide

rié. Ses paramètres physiques se déclinent en paramètres intérieurs et extérieurs que nous résumons dans le tableau 4.3. Les fluctuations des valeurs des paramètres de sortie du système dynamique sont reliées non linéairement aux variations des valeurs des paramètres d'entrées. Les variations peuvent se résumer par un système à trois régimes :

1. état stationnaire caractérisé par des variations régulières de la température selon un ensemble de points limites (haut/bas) ;
2. réponse à l'effacement électrique qui se manifeste par une augmentation de la température sur une longue période (30 min. à 3h) ;
3. dégivrage par une augmentation soudaine de la température pendant une très courte temps (5 à 10min).

Les trois régimes sont illustrés par la figure 4.8 qui montre une grande différence comportementale et temporelle des trois régimes. Il est important de signaler que la distinction entre les trois régimes est bien plus facile quand la chambre froide est déchargée. En effet, plus la chambre est chargée, plus le dégivrage est fréquent (régime 2). A contrario, le temps d'augmentation de la température de l'air pendant le régime d'effacement est plus lent (comme on peut le constater sur les régimes 4.9).

Ce problème est abordé classiquement par des modèles de comportement dynamique des systèmes thermiques. Ces modèles nécessitent une très bonne connaissance d'un tel sys-

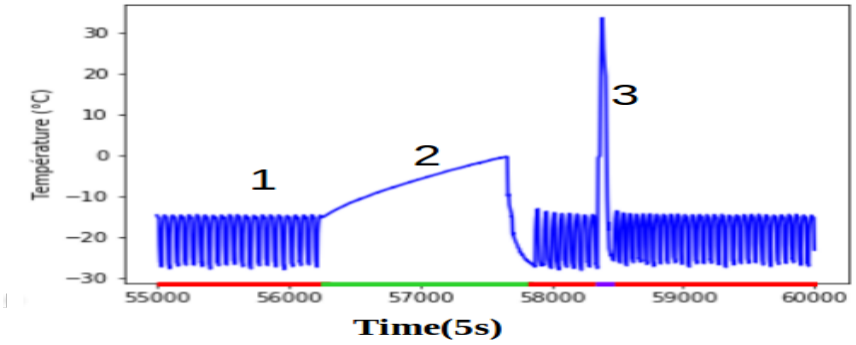


FIGURE 4.8 – Trois régimes : (1) état stationnaire ; (2) effacement électrique ; (3) dégivrage

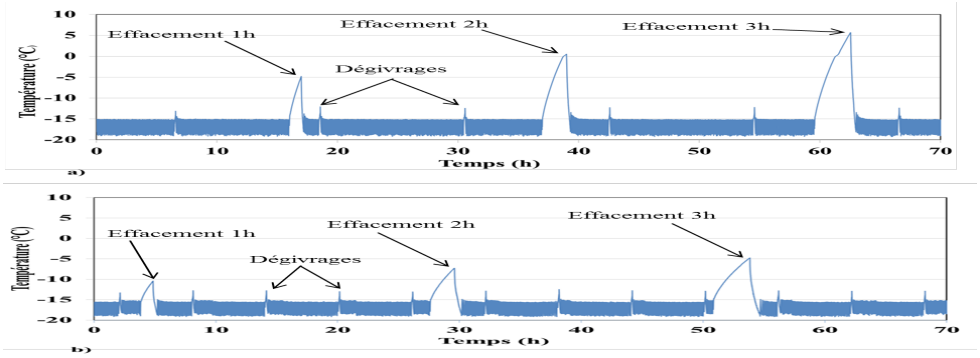


FIGURE 4.9 – Régime sans et avec charge : (a) sans charge, (b) avec charge[Mah20]

tème thermique et ne tolèrent pas l'incomplétude des équations différentielles, ce qui est souvent un compromis entre le volume des données traitées et le nombre d'équations à résoudre. Le passage à l'échelle de ces modèles n'est, pour l'instant, pas possible et ouvre des perspectives à d'autres types de modélisation orientés données (*end to end models*). En effet, lorsque l'objectif principal est le résultat final obtenu à la sortie d'un système, indépendamment du fonctionnement interne, il peut être intéressant d'exploiter des modèles non linéaires dont le but est uniquement de prédire les paramètres de sortie à partir des entrées. Cela présente, en outre, l'avantage d'être facilement généralisable, au moins en présence de données suffisantes et représentatives. Des modèles capables de capturer la structure interne et les liens entre les données tels que les réseaux neuronaux artificiels (ANN) pourraient constituer une approche alternative pour résoudre des problèmes complexes. Les ANN permettent de cartographier des relations non linéaires complexes internes aux données et arrivent à résoudre plusieurs problèmes tels que la planification, le contrôle, l'analyse, la prédiction et la reconnaissance. La littérature scientifique a démontré leur capacité de surpasser les méthodes conventionnelles avec un potentiel remarquable de modéliser des processus non linéaires. À l'heure actuelle, bien que des études [HXW17] [XWYC14] aient été menées dans le cadre de la consommation électrique, aucune méthode de ce type ne semble avoir été appliquée à la problématique de l'effacement électrique pour la réfrigération.

Prédiction de séries temporelles par des RNN

Les réseaux de neurones ont attiré beaucoup d'attention ces derniers temps et ont intégré un panel d'applications très large. Ils ont été utilisés avec beaucoup de succès dans différents domaines tels que, à titre d'exemples la reconnaissance de motifs en analyse et traitement de signaux et d'image, en finance, en diagnostic en détection de pannes. Certaines architectures de réseaux de neurones sont employées en adéquation avec les objectifs à la fois de l'application et du jeu de données. Typiquement, si l'objectif est la classification et les données en entrée sont des images alors les réseaux de neurones convolutionnels sont un meilleur choix d'architecture. Si les données en entrée sont des séries temporelles alors les réseaux de neurones récurrents, en particulier les modèles LSTM (*Long Short-Term Memory*), sont les plus appropriés. Nous constatons donc que, pour un même objectif,

l'architecture diffère en fonction des données en entrée. Dans notre contexte où les données en entrées sont des séries temporelles multivariées, deux types d'architectures peuvent être candidates : MLP (*MultiLayer Perceptron*) et LSTM. Les modèles MLP supposent une indépendance entre les entrées et les sorties avec une séquence de taille 1, c'est-à-dire une séquence de mémoire extrêmement courte ; tandis que les modèles LSTM disposent de cellules de mémoires capables de mémoriser les dépendances temporelles entre les sorties et les entrées sur une séquence longue. Théoriquement, les modèles LSTM mémorisent des séquences de longueurs arbitraires. Ils se composent de cellules d'états (C_t, pour « Cells ») et de portes qui décident les données à conserver ou à jeter pour chaque séquence. Il existe trois types de portes : une porte d'oubli (f_t, pour *forget*), une porte de mémoire pour stocker les nouvelles information entrantes (i_t, pour *input*) et, enfin, la porte de sortie (o_t, pour *output*). Chaque porte est une couche de réseau de neurones composée soit par une sigmoïde (σ, valeur $\in [0, 1]$), soit par une tangente hyperbolique ($tanh$, valeur $\in]-1, 1[$). Le but d'une sigmoïde est de permettre d'oublier (0 oubli) ou non une valeur (1 prendre la valeur telle qu'elle) alors que $tanh$ permet de normaliser les valeurs. Par conséquent, l'état d'une cellule C_t à un instant t (*équation* 4.18) est contrôlé par l'information qui doit être ignorée par rapport à l'état de la cellule C_{t-1} à l'instant précédent $t-1$ (*équation* 4.15) et ce qui doit être stocké comme nouvelle information entrante i_t (*équation* 4.16) à l'état actuel $\tilde{C}$ (*équation* 4.19). La sortie prédite h_t à l'instant t (*équation* 4.20) est le produit de la sortie o_t (*équation* 4.17) filtrée par l'état de la cellule à l'instant t. Les équations de contrôle de l'état d'une cellule à un instant t se résument par les équations suivantes :

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{4.15}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{4.16}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{4.17}$$

$$C_t = C_{t-1} \otimes f_t \oplus \tilde{C}_t \otimes i_t \tag{4.18}$$

$$\tilde{C}_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{4.19}$$

$$h_t = o_t \otimes tanh(C_t) \tag{4.20}$$

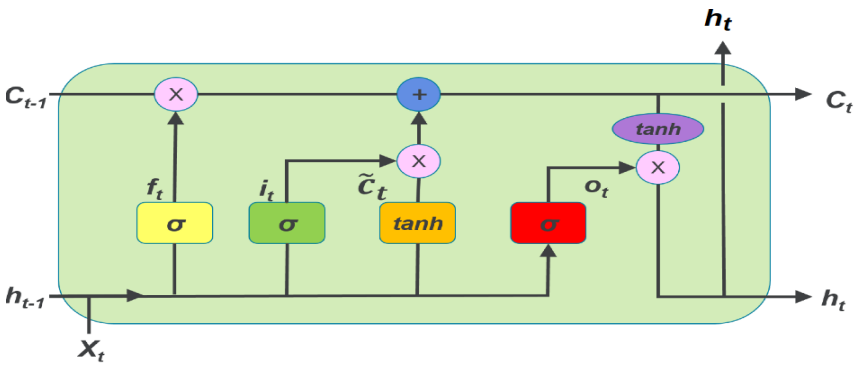


FIGURE 4.10 – Composition d'une cellule LSTM

où $\otimes$ et $\oplus$ sont respectivement l'opérateur ponctuel de multiplication et d'addition appliqué aux matrices (point à point) ; W et b sont respectivement la matrice des poids et le biais liées à chacune des portes (oubli, mémoire, sortie) et l'état (de la cellule). Plusieurs variantes de l'architecture LSTM présentée ici comme celle de référence, existent dans la littérature en proposant des structures de cellules légèrement différentes. Beaucoup d'entre elles sont utilisées pour l'analyse automatique de texte ou bien pour des problématiques de traduction. Les questions légitimes qui nécessitent d'être posées à ce stade sont quelle architecture serait la plus performante, pour quelle application et pour quelles données ?

Une étude empirique a été menée par Google [JZS15] pour apporter des éléments de réponse à la première question. Ils ont montré que les différentes architectures sont similaires en terme de performance sur trois domaines d'application. Inspiré par ces travaux, et face à la complexité de la dépendance temporelle qui existe entre les données, nous avons mené une étude comparative de cinq architectures différentes en partant de l'architecture de référence présentée ci-dessus. Nous notons par y_t et x_t des observations de séries chronologiques multivariées, représentant respectivement tout paramètre de sortie et d'entrée dans le tableau 4.3. Nous notons par f une fonction de prédiction non linéaire avec des entrées $m + p$ telle que

$$y_{t+1} = f(y_t, y_{t-1}, ..., y_{t-m+1}, x_t, x_{t-1}, ..., x_{t-p+1}) \tag{4.21}$$

Nous considérons h comme une période discrète sur $[1, m]$. Ayant une série d'observations $y_t, y_{t-1}, ..., y_{t-m+1}$ sur une période fixe h nous proposons d'estimer $\hat{y}(m, H)$ à des dates futures m dans un horizon H donné. Ce problème est équivalent à l'estimation du $\hat{f}$

optimal de l'équation 4.5 qui minimise :

$$Erreur = (\widehat{y} - y)^2 \tag{4.22}$$

Pour cela, en plus du modèle LSTM de référence, nous avons comparé trois autres modèles qui en découlent à savoir LSTM convolutionnel, LSTM empilé, LSTM bidirectionnel. Les critères de comparaison que nous avons choisis s'appuient sur des métriques classiques d'évaluation de leur performance. Le premier critère que nous avons employé est le $fitting$. C'est un critère d'ajustement permettant de mesurer la proximité entre les valeurs de référence et les valeurs prédites. Plus sa valeur est proche de 100 %, plus la variable est correctement prédite. Il correspond donc à un pourcentage défini par :

$$Fit(Y) = 100x(1 - \frac{|\widehat{Y} - Y|}{|Y - \bar{Y}|}). \tag{4.23}$$

Le second critère est l'erreur quadratique moyenne (MSE). C'est une moyenne arithmétique des carrés des différences entre les prévisions et les observations réelles. Plus l'erreur quadratique moyenne est faible, meilleure est la prédiction. Cette valeur est donc définie par :

$$MSE(Y) = \frac{1}{N}.\sum_{i=1}^{N}(\widehat{Y_i} - Y_i)^2. \tag{4.24}$$

La racine de cette valeur (RMSE) est également intéressante et est simplement calculée par :

$$RMSE(Y) = \sqrt{MSE(Y)}. \tag{4.25}$$

MSE et RMSE correspondent à une indication agrégée de l'erreur de prévision. Contrairement à MAE, RMSE permet de pénaliser d'une manière plus sévère les erreurs élevées. Ces deux indicateurs sont faciles à interpréter puisqu'ils sont explicités sur l'unité de la variable à expliquer malgré leur comportement instable vis à vis des modèles. L'erreur moyenne absolue (MAE), est un autre critère basé sur la moyenne arithmétique des différences entre les prévisions et les observations réelles. Dès lors que la mesure n'est pas au carré, chaque différence est gérée avec la même importance. L'objectif est bien sûr de

minimiser cette valeur, et elle est définie sur l'horizon de prévision $[t+1, t+H]$ par :

$$MAE(Y) = \frac{1}{N}.\sum_{i=1}^{N} |\widehat{Y_i} - Y_i|. \tag{4.26}$$

Enfin, le coefficient de variation (CV) est une mesure qui relate la dispersion relative des données autour de la moyenne. Le coefficient de variation se calcule comme le ratio de l'écart-type rapporté à la moyenne et s'exprime en pourcentage par la formule :

$$CV(Y) = 100.\frac{RMSE(Y)}{\bar{Y}}. \tag{4.27}$$

L'ensemble de ces métriques sont analysées d'une manière conjointe afin de chercher le modèle optimal. Les différents modèles ont été confrontés aux scénarii d'usage comme suit. Tout d'abord, nous partons d'un comportement de référence d'une chambre froide, stationnaire sans perturbations, et à température interne constante. Cette étape à pour but d'initialiser les quatre modèles. Ensuite, nous avons établi cinq séries chronologiques pour évaluer chacun de modèles. Dans ce qui suit, on note par E_i un scénario d'une série chronologique où $i = 1..5$ élaboré pour chaque cas d'utilisation i :

— E_1 (train : 127975, test : 60000) : l'hypothèse est de considérer un ensemble de mesures avec trois périodes d'effacement, uniformément réparties sur trois jours. Nous simulons ici la perturbation stochastique du système en considérant une distribution uniforme du bruit. Ainsi, $\delta T_{erasure}$ est diminué de façon aléatoire et $T_{erasure}$ est fixe ;

— E_2 (train : 223545, test : 149030) : est de considérer un ensemble de mesures avec deux périodes d'effacement par jour, uniformément réparties sur cinq jours. Dans ce cas, nous augmentons la fréquence d'occurrence du bruit comparé au cas d'usage 1 et nous augmentons le nombre total de mesures. En effet, ce cas nous permet d'étudier le biais de la fréquence d'occurrence du bruit ainsi que la quantité de données ;

— E_3 (train : 490985, test : 294590) : est de considérer un ensemble de mesures sur cinq jours avec une période d'effacement aléatoirement effectuée par jour. Cela signifie que les deux $\delta T_{erasure}$ et $T_{erasure}$ sont aléatoires.

— E_4 (train : 630920, test : 420610) : est de considérer l'union des trois hypothèses précédentes. Elle correspond à un modèle généralisé du problème de réponse à la

T_p	LSTM	ConLSTM	StackedLSTM	BidirectionalLSTM
E_1	(Mae =0.38,Fit=35.8)	(Mae=0.38, Fit=60.5)	(Mae=0.44, Fit=42.8)	(Mae =0.5, Fit=41.7)
E_2	(Mae =0.16,Fit=29.7)	(Mae=0,33, Fit=14.8)	(Mae=0.14, Fit=30.4)	(Mae =0.31, Fit=24.4)
E_3	(Mae =0,27,Fit=48.0)	(Mae=0.26, Fit=46.9)	(Mae=0.23, Fit=50.9)	(Mae =0.31, Fit=52.1)
E_4	(Mae =0.27,Fit=58.2)	(Mae=0.33, Fit=55.5)	(Mae=0.27, Fit=61.1)	(Mae =0.37, Fit=48.4)
E_5	-	-	(Mae=0.19,Fit=69.64)	-

TABLE 4.4 – Prédiction de la Température T_p pour chaque E_i avec quatre modèles dérivés du LSTM.

$CV(Compressor_{Energy})$	LSTM	ConLSTM	StackedLSTM	BidirectionalLSTM
E_1	18.9	23.5	18.9	20.3
E_2	15.7	13.7	47.1	20.08
E_3	15.1	13.8	54.42	22.6
E_4	16.2	15.9	16.1	16.1
E_5	-	28.5	-	-

TABLE 4.5 – E_i $Compressor_{Energy}$ prediction with the four derived LSTM models

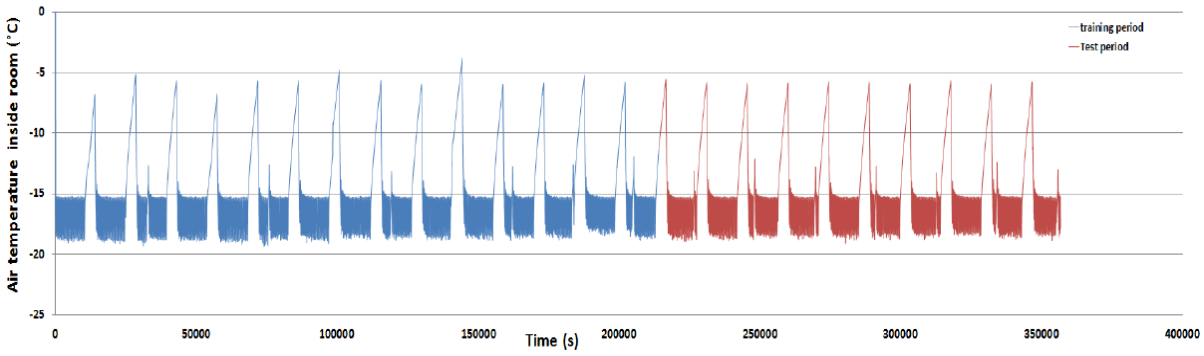


FIGURE 4.11 – Prédiction des séries temporelles par modèles LSTM sur les données de E_1

demande d'électricité, c'est-à-dire que nous avons une grande quantité de données, plus de bruit et plus d'aléa.

— E_5 (train : 214080, test : 142700) : est de considérer $\delta T_{erasure}$ et $T_{erasure}$ non aléatoires mais variant entre 1 et 3 heures en augmentant l'horizon de prédiction H.

Chaque ensemble de données E_i est respectivement divisé en 60% de données pour l'apprentissage/Validation et 40% pour le test, sauf le E_5 utilisé en tant que échantillon de test supplémentaire.

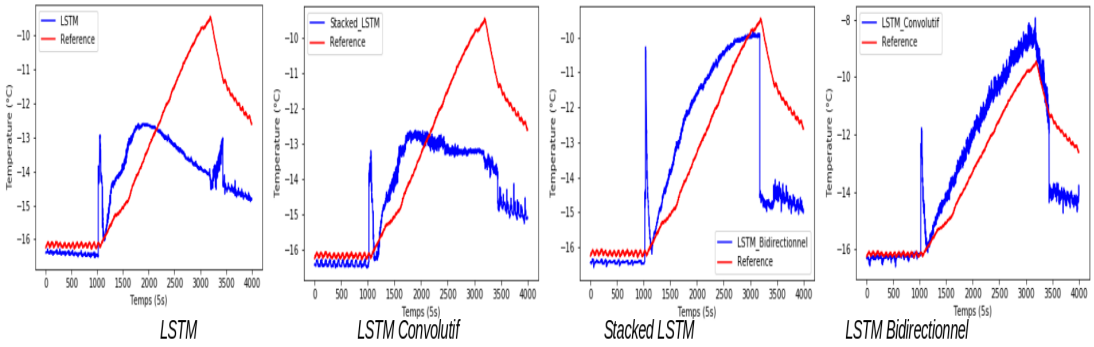


FIGURE 4.12 – Prédiction de la série temporelle $Tp_{product}$ par modèles LSTM sur les données de E_2

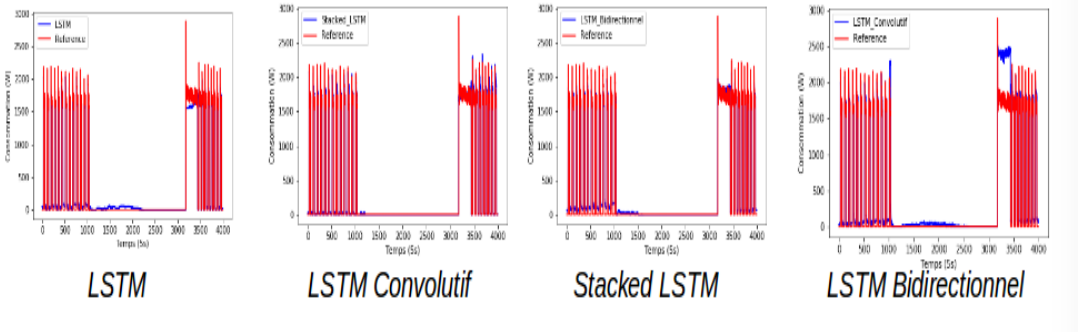


FIGURE 4.13 – Prédiction de la série temporelle $Tp_{product}$ par modèles LSTM sur les données de $Compressor_{Energy}$ sur les données de E_3

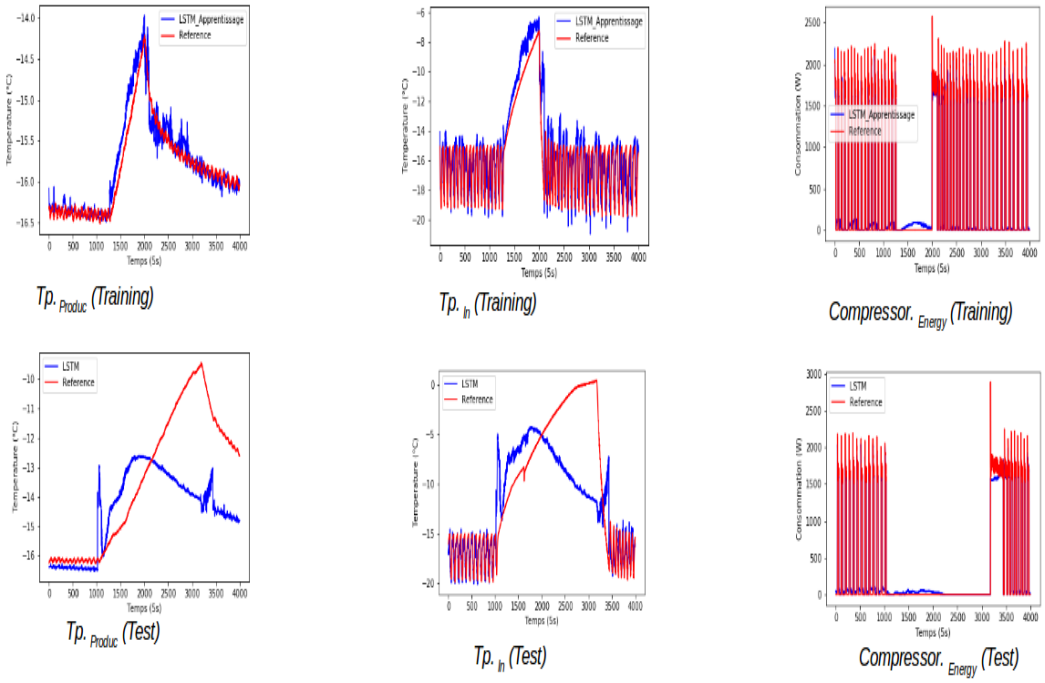


FIGURE 4.14 – Prédiction des séries temporelles par modèles LSTM sur les données de E_5

Comme le montre les figures 4.11, 4.12, 4.13, **??**, nous obtenons des résultats significatifs grâce à l'utilisation du réseau LSTM convolutif utilisant l'ensemble des données et sur toutes les variables à l'exception de Ep. Cependant, ces résultats sont encore très modestes et s'expliquent facilement par la faible quantité de données. En outre, il convient de noter que certaines caractéristiques, telle que $Tp_{produit}$, sont mal prédites par tous les modèles. Ces problèmes semblent être en partie liés aux capteurs. Une analyse plus détaillée a été effectuée dans le cadre de la thèse de Mme Akerma vérifiant les fuites de données causées par les différents capteurs. Les résultats obtenus avec l'ensemble des données E_2 sont moins performants pour prédire les caractéristiques liées à la température. En particulier, une augmentation soudaine de la température prédite a été noté dès que la coupure électrique est déclenchée, suivie d'une diminution tout aussi soudaine à la fin de la période d'effacement. En ce qui concerne la consommation d'énergie, le comportement est toujours reproduit aussi fidèlement qu'auparavant. Avec cet ensemble de données, le LSTM stacked et le bidirectionnel surpassent les autres modèles. On peut noter que ces deux derniers sont moins sensibles à la quantité de données. Avec l'ensemble des données E_3, le modèle LSTM de référence ainsi que le convolutif surpassent les autres modèles. Ils semblent être plus efficaces en présence de bruits aléatoires. Les modèles LSTM bidirectionnels et le *stacked* sont capables de prédire la dynamique des séries temporelles mais sont sensibles au bruit, en particulier pendant les périodes de démarrage de l'effacement et pendant les reprises. Le comportement *idiosyncrasique* de signaux prédits se manifeste par un signal bruit qui s'ajoute au signal prédit. Avec l'ensemble de données E_4 (rappelons que c'est l'union des trois ensembles précédents de données), nous obtenons des résultats comparables à ceux obtenus par LSTM de référence et le *stacked*. Leur précision est supérieure à celle des deux autres modèles. Il convient toutefois de noter que les quatre modèles sont entraînés sur l'union de E_1 et E_2 et testés sur l'ensemble de données E_3. Ce qui nous semble intéressant, c'est que nous nous attendions à ce que les résultats de ce cas d'utilisation soient similaires à ceux du cas d'utilisation précédent (cas d'utilisation 3). Or, on remarque que l'apprentissage de plus de bruit lié au dysfonctionnement des capteurs, permet au modèle LSTM *stacked* de mieux prédire le bruit ainsi que les données aléatoires. De plus, au vu des résultats obtenus, il est moins sensible à la quantité des données. Enfin, l'ensemble des données E_5 est utilisé comme échantillon de test pour

le LSTM *stacked* pré-entraîné. Comme la taille de l'horizon de E_5 est de vingt minutes, soit quatre fois plus que les précédents, les données des séries chronologiques sont moins nombreuses. Tant que la taille de l'horizon E_5 est quatre fois plus grande que les précédentes, les données sont plus lissées et moins bruitées. En conséquence, le modèle a pu mieux prédire avec un gain d'environ 10%. La variable $Compressor_{Energy}$ est correctement prédite dans tous les cas d'usage où l'ajustement est d'environ 90% et la MAE d'environ 0.1. Ceci est dû à son indépendance par rapport à l'état interne du système. Les résultats obtenus par ces modèles montrent leur intérêt et leur pertinence quant à la prédiction de séries temporelles multivariées. Afin d'améliorer leur performance il est nécessaire de disposer de davantage de données. Or, comme il a été décrit au début de cette section, en l'absence de bases de données de référence, les données sont très difficiles à obtenir. De plus, chaque élaboration d'un ensemble E_i nécessite au moins six semaines d'expérimentation. Pour palier ce manque de données, nous avons augmenté les données expérimentales par des données simulées. En effet, l'extension de ce travail nous a ouvert de nouvelles perspectives de recherche qui s'insèrent pleinement dans les travaux de recherche actuels sur l'apprentissage par transfert. La première perspective est l'apprentissage par transfert homogène du domaine source des données simulées $\mathcal{D}_s$ vers le domaine cible des données expérimentales $\mathcal{D}_e$. La seconde perspective vise l'apprentissage par transfert hétérogène entre le domaine source et domaine cible. La formalisation de l'apprentissage par transfert homogène permet de modéliser la prédiction de l'effacement dans les chambres froides. Alors que la formalisation de l'apprentissage par transfert hétérogène vise à modéliser la prédiction de l'effacement dans les entrepôts frigorifiques. Ces deux axes font actuellement l'objet de nos travaux recherche et qui sont par ailleurs étendus aux domaines du son, du texte et de l'image. Ces perspectives seront détaillées dans le chapitre suivant.

4.6 Prédiction à partir des données de réseaux sociaux

Les données textuelles numérisées sont devenues des données courantes pour plusieurs applications et dans différents domaines. Elles peuvent exister sous forme d'une liste de

mots-clés, de phrases, de paragraphes, ou encore sous forme plus complexes tels que les résumés. L'analyse des données textuelle est un axe à part entière qui nécessite des compétences pointue en linguistique. C'est un domaine à l'intersection de la fouille des données, du traitement automatique de la langue et de l'intelligence artificielle. Nous avons montré dans le chapitre précédent, l'intérêt de l'analyse des annotations (qui sont des séquences de phrases) pour l'enrichissement sémantique des descripteurs des images contribuant à réduire le fossé sémantique visuel. Mais le texte étant présent partout, il devient un enjeu, mais aussi un atout, pour l'enrichissement sémantique et l'aide à l'interprétation des résultats. Cependant, l'analyse des données textuelles doit rendre compte de différentes dimensions qui peuvent être dans bien des cas dépendantes. La dimension spatiale d'un mot dans une phrase contribue à la structure de celle-ci. Ce même mot devient un vecteur de sens quand on regarde sa dimension temporelle. Le sens d'un même mot peut être différent d'une phrase à une autre. Même si les mots restent génériques, leur fréquence d'usage peut dépendre plus ou moins fortement du domaine. Une autre difficulté à laquelle l'analyse textuelle devra faire face est la dimension multilingue. En effet, la traduction d'un mot d'une langue à une autre peut lui faire perdre son sens faute d'une relation bijective parfaite permettant la conservation des différentes dimensions (spatiale, temporelle, domaine). Nous abordons dans cette section, le problème de la prédiction quand les données textuelles s'ajoutent aux données numériques. Dans un contexte où la prolifération des réseaux sociaux est source d'interaction et d'échange d'information, les systèmes de recrutement assisté par ordinateur (*e-recruitment*) a explosé ces dix dernières années. Cela a pour conséquence de favoriser la multiplication des réseaux sociaux professionnels spécialisés en diffusion des offres d'emploi. Pour atteindre des objectifs stratégiques et économiques d'une part par les recruteurs et d'autre part par les demandeurs d'emploi, il devient donc nécessaire de repenser les modèles de recommandation qui aideraient les recruteurs à choisir le ou les canaux de diffusion appropriés aux offres d'emploi. En effet, disposer d'un modèle pertinent de recommandation de canaux permet à l'entreprise d'optimiser ses coûts financiers, comme elle permet au demandeur d'emploi d'optimiser ses coûts de recherche en fonction de ses compétences et profil professionnel. À moyen terme, les entreprises pourraient aussi bénéficier d'un système de recommandation de CV approprié pour chaque profil recherché par une offre d'emploi. Pour ces objectifs, plusieurs

problématiques ont été dégagées. Tout d'abord, nous avons été confrontés à l'analyse du contenu textuel des offres d'emploi. L'analyse du contenu des offres contribue d'une part à enrichir le modèle de classification des canaux de diffusion et, d'autre part, à comprendre l'évolution de leur contenu. Ces connaissances augmentées jouent un rôle précieux dans le fonctionnement du système de recommandation des canaux de diffusion afin qu'il soit pertinent face à une offre donnée. Ensuite, nous avons étudié l'évolution temporelle du contenu des canaux de diffusion à travers les consultations des utilisateurs dans le but de prédire leur attractivité. Enfin, grâce à la jonction des deux problématiques décrites ci-dessus, nous avons proposé un système de recommandation temporelle où la notion d'attractivité se conjugue avec l'évolution du contenu textuel des offres dans le temps mais aussi par bassin géographique. En effet, la dimension géographique s'est révélée un critère discriminant dans la diffusion de certaines offres et donc de certains canaux spécialisés. Ces travaux ont été élaborés dans le cadre d'un projet FUI-SONAR (tableau. 4.1), mené par le post-doctorant Sidahmed Benabderrahmane, durant deux ans et ont conduits à d'autres perspectives intéressantes de recherche, notamment en analyse du contenu dynamique et en prédiction temporelle du contenu des réseaux sociaux spécialisés. Dans la suite de cette section, nous allons nous focaliser sur deux aspects : l'analyse du contenu textuel et la prédiction des séries temporelles symboliques.

4.6.1 Word2Vec et Doc2Vec pour la représentation textuelle

Chaque offre d'emploi, en tant que document, est représentée dans notre base de données par une liste d'éléments structurés. Ils sont principalement composés par le titre de l'emploi, la description, les compétences requises, le lieu, le salaire, et des informations annexes. Chaque élément est ensuite transformé en un vecteur de termes fréquents du document en question. En outre, nous disposons d'un vocabulaire de classification des emplois, référencé par un organisme public français (code ROME [4]). Il est donc nécessaire de représenter ces données de manière idoine au modèle de traitement. Le texte est l'une des formes les plus répandues des données séquentielles. Il peut être traité soit comme une séquence de caractères, soit comme une séquence de mots, bien qu'il soit plus courant de

4. /www.pole-emploi.fr/.../les-fiches-metiers.html

travailler au niveau des mots. Les modèles d'apprentissage profonds sur des séquences que nous avons utilisés pour traiter les offres d'emploi sont capables d'exploiter du texte pour produire des connaissances augmentées contribuant à la compréhension du langage naturel. L'apprentissage profond est très largement exploité pour des applications allant de la classification de documents, l'analyse des sentiments, l'identification d'auteurs ou même la réponse à des questions dans un contexte donné [Cho17]. L'apprentissage profond pour le traitement du langage naturel est simplement un modèle de reconnaissance appliquée aux mots, phrases et paragraphes, de la même manière que la vision par ordinateur est une simple reconnaissance de formes appliquée aux pixels. Comme tous les autres réseaux neuronaux, les modèles d'apprentissage profond ne prennent pas en compte le texte brut. Ils ne fonctionnent qu'avec des tenseurs (vecteurs) numériques. La vectorisation du texte est le processus de transformation pouvant être réalisé de plusieurs façons :

— en segmentant le texte en mots, et en transformant chaque mot en un vecteur ;

— en segmentant le texte en caractères, et en transformant chaque caractère en un vecteur ;

— en extrayant des *n-grammes* de mots ou de caractères, et en transformant chaque n-gramme en un vecteur. Les *n-grammes* sont des groupes de mots ou de caractères consécutifs multiples qui se chevauchent.

Collectivement, les différentes unités dans lesquelles un texte peut être décomposé (mots, caractères ou n-grammes) sont appelés *« tokens »* (c'est le grain de décomposition du texte comme le pixel pour une image), et la décomposition du texte en tokens est appelée *« tokenisation »*. Tous les processus de vectorisation de texte consistent à appliquer un schéma de tokenisation, puis à associer des vecteurs numériques aux tokens générés. Ces vecteurs, regroupés en tokens par séquence viennent alimenter les réseaux neuronaux profonds.

Dans ce travail, nous en avons utilisé deux principales représentations : l'encodage à chaud (*« one-hot »*) et l'incorporation des tokens (utilisé exclusivement pour les mots et généralement appelé plongement de mots ou *« Word embedding »*) [Cho17][Jas16].

L'encodage à chaud est le moyen le plus courant et le plus élémentaire de transformer un mot en vecteur. Il consiste à associer un index entier unique i à chaque mot, puis à transformer cet index entier en un vecteur binaire de taille N, la taille du vocabulaire qui serait

entièrement nulle sauf pour la $i^{ème}$ entrée qui serait de 1. Le plongement de mots consiste à utiliser des vecteurs de mots, est une autre façon bien répandu et puissante permettant d'associer un vecteur à un mot. Alors que les vecteurs obtenus par un codage à chaud sont binaires, éparses (principalement constitués de zéros) et à très haute dimension voire de même dimension que le vocabulaire, les plongements de mots sont des vecteurs denses de nombres réels et à faible dimension par opposition aux vecteurs éparses. Il est courant d'associer des représentations de plongement de mots sur 256, 512 voire 1024 mots lorsqu'il s'agit de très grands vocabulaires. Cette technique est fondée sur l'hypothèse de Harris qui suppose que les mots apparaissant dans des contextes similaires ont des significations apparentées. D'autre part, les mots codés en une seule fois conduisent généralement à des vecteurs de grande dimension. Par conséquent, le plongement de mots permet de regrouper plus d'informations et d'avoir ainsi une dimension plus réduite.

La figure 4.15 donne un exemple de représentation d'un texte avec des valeurs numériques.

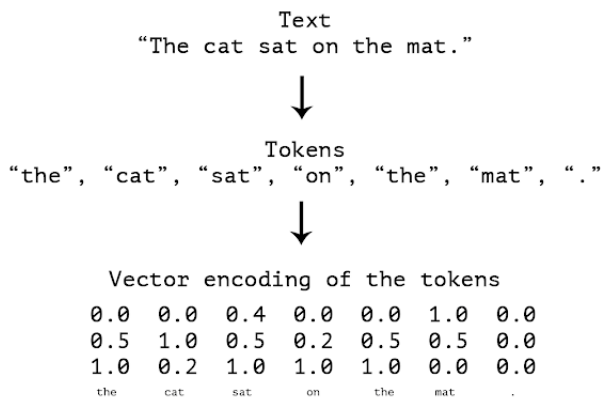


FIGURE 4.15 – Du texte aux vecteurs en passant par les tokens [Cho17].

Deux modèles se distinguent dans la littérature qui sont *Word2vec* et *doc2vec*. *Word2vec* est à la base d'un réseau de neurones à deux couches ce qui lui permet d'être rapidement entraîné et exécuté. Son objectif consiste à apprendre les représentations vectorielles des mots de telle manière que les mots ayant le même contexte soient représentés par des vecteurs assez proches. Pour ce même modèle, il existe deux architectures différentes à savoir *CBOW* et le *skip-gram*. *CBOW* permet de prédire un mot en fonction de son contexte qui peut être soit un mot unique ou un groupe de mots. Tandis que le *Skip-Gram* per-

met de prédire le contexte à partir du mot. Ces deux architectures peuvent être utilisées de façon conjointe dans le cas de corpus incomplet et où le contexte est imprécis ; plus particulièrement dans le contexte de la détection de l'incohérence dans des documents multilingues ou cross-lingues. Nous reviendrons en détail sur cette problématique de niche plus loin dans cette section. *Doc2vec* est une version modifiée du *Word2vec* pour l'adapter à la représentation des documents. *Doc2vec* peut être vu comme un changement d'échelle opéré sur les mots permettant de représenter les documents sous la forme de vecteurs de tailles fixes. Ainsi chaque document est-il représenté par un vecteurs de descripteurs de taille fixe. Cette représentation permet de prédire les mots d'un document. Rappelons que nos documents d'expérimentation sont des offres d'emploi. Ce sont des documents textuels dont le vocabulaire utilisé dans la description de l'offre est un ensemble de mots caractéristiques d'un métier donné. Par conséquent, les canaux de diffusion des offres d'emploi peuvent être analysés et classés en fonction des offres. Autrement dit, la caractérisation des offres d'emploi via leur contenu est une identité sémantique d'un canal ou d'un ensemble de canaux de diffusion. Plusieurs canaux peuvent diffuser une même catégorie de métier et donc des offres similaires. Ainsi, l'exploitation des offres suivant une représentation *Doc2Vec* permet de caractériser le métier alors que l'exploitation des mêmes documents en *Word2Vec* a pour objectif de caractériser le type de l'emploi pour un métier donné. Par exemple, pour le métier de l'informatique, on trouve plusieurs déclinaisons d'emploi tels que *administrateur web*, *développeur web*, *architecte web*, etc. Ces deux informations liées au métier de l'informatique et à la spécialité sont très complémentaires et détectables à deux échelles différentes. Elles apportent de la précision sur la spécificité des canaux de diffusion et permettent de lever un fossé sémantique entre le contenu d'une offre et la spécificité d'un canal de diffusion. Quand plusieurs canaux de diffusion peuvent être candidats potentiels à la publication d'une offre d'emploi, la question de la recommandation d'un canal attractif dédié peut se poser. Nous partons de l'hypothèse que l'attractivité d'un canal peut être mesurée par le nombre de visites comptabilisé par l'activation des liens correspondants aux offres publiées par un canal de diffusion. Les offres d'emploi sont publiées périodiquement sur un ou plusieurs canaux de diffusion à une date donnée. Une offre diffusée sur un canal dispose d'un cycle de vie limité. Pendant cette période, le nombre de liens activés (autrement dit le nombre de clics) sur les offres d'emploi dans les différents

canaux peut être facilement connu. Ainsi, le nombre de clics quotidien associé à une offre et à des canaux est disponible. Ce nombre peut être relaté sur différentes échelles temporelles : hebdomadaire, mensuelle, semestrielle ou même annuelle. Nous indiquons par T la période ou l'échelle de temps associée au nombre de clics considéré. Pour formuler cette représentation des données, en particulier le nombre de clics, nous considérons un canal JB, comme un ensemble d'offres o_j sur une période de temps donnée T :

$$JB_T = \bigcup o_j \text{ for } \ j = 1, ..., p. \tag{4.28}$$

Pour chaque canal, nous introduisons un ratio X^{JB_T} calculé comme le nombre total de clics relatifs aux offres publiées sur ce canal durant une période T :

$$X^{JB_T} = \frac{nb.click}{|JB_T|}. \tag{4.29}$$

Ici, nous considérons T comme un intervalle discret $[1, N]$. Étant donné que les clics relatifs sont des valeurs numériques, nous pouvons considérer le rapport $X_t^{JB_T}$ comme une observation temporelle des clics donnés à l'instant t. Ayant une série d'observations $X_1^{JB_T}, X_2^{JB_T}, ..., X_N^{JB_T}$ sur une période fixe T, nous proposons de prédire les clics $\hat{X}^{JB_T}(N, h)$ sur des dates futures N dans un horizon h donné. Il s'agit de prédire la réalisation future d'une variable aléatoire X^{JB_T} en utilisant les valeurs précédemment observées $X_1^{JB_T}, X_2^{JB_T}, X_N^{JB_T}$ pour chaque canal de diffusion JB, et d'ordonner par la suite les séquences temporelles prédites afin de sélectionner le canal qui maximise X^{JB_T}. Nous utiliserons ici des séries chronologiques univariées, et nous notons la variable X^{JB_T} simplement par x_t, une observation à l'instant t.

4.6.2 Prédiction de séquences symboliques

En pratique, les séries temporelles numériques sont souvent de dimension très élevée, et souffrent de difficultés de stockage, d'accès et d'interprétation. Ces difficultés ont contribué à la production de divers algorithmes de réduction de dimension. En particulier, de nombreuses méthodes de discrétisation ont été proposées dans la littérature pour co-

der les séries temporelles en séquences symboliques [CF99][CKMP02][LKLC03][MWLF05]. Parmi les algorithmes de transformation de séries temporelles numériques en séquences symboliques, on peut énumérer la transformée de Fourier, l'ACP, la transformée en ondelettes,l'Approximation par Agrégation symbolique nommé SAX [LKLC03] et la MVQ (Multi-resolution Vector Quantization) [ED04]. Toutes ces méthodes ont été évaluées d'une manière exhaustive par les travaux antérieurs du post-doctorant dans un contexte de recherche de similarités et de clustering [BQG14] [RRFD16]. En partant de ces travaux, nous avons proposé une extension de MVQ pour le rendre parallèle. Les deux méthodes SAX et PMVQ, ont été retenues pour la prédiction de nos séries symboliques. Chaque symbole prédit est une quantification des clics des utilisateurs sur les offres d'emploi. C'est donc les trajectoires comportementales des demandeurs d'emploi que nous avons suivies à travers ces nouvelles représentations symboliques. L'avantage d'une telle représentation est de pouvoir appliquer des techniques de traitement automatique de la langue sur des suites de mots.

Séquences symboliques par agrégation : la méthode SAX

La méthode SAX a pour rôle de faire correspondre une série temporelle numérique :

$$T = (X_1^{JB_T}, X_2^{JB_T}, ..., X_N^{JB_T}) \tag{4.30}$$

à une séquence de symboles codée sur un alphabet de taille réduite $\mathcal{A}$=$|\Sigma|$ [LKLC03]. La première étape de l'algorithme consiste à diviser la série temporelle de longueur N en w fenêtres (à 1 séquence est associée un mot(code)) de taille égale et calcule la valeur moyenne des données de chaque fenêtre. Ainsi, un vecteur de taille w devient la représentation réduite des données d'origine. La somme de ces moyennes est basée sur une approximation agrégée par morceaux (AAP) où le i^{th} élément est défini par :

$$C_i = \frac{w}{n} \Sigma_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} X^{JB_T}. \tag{4.31}$$

Il convient de noter qu'avant d'appliquer l'AAP, chaque série temporelle est normalisée (*i.e.* une moyenne nulle et un écart-type de 1) afin d'éviter de comparer des séries tem-

porelles présentant des décalages et des amplitudes différents. Dans la deuxième étape, chaque segment est codé par une séquence de l'alphabet $\mathcal{A}$. La conversion de la représentation AAP d'une série temporelle en SAX repose sur la production de symboles qui correspondent aux caractéristiques de la série temporelle avec une probabilité égale pour chaque symbole. Keogh *et al.* [LKLC03] ont pu montrer qu'en règle générale, les données des séries chronologiques suivent une distribution normale. Avec une telle distribution, nous pouvons facilement choisir des zones de taille égale. L'extraction des zones s'appuie principalement sur la détection des points de rupture (par exemple, les quantiles sont des points de rupture qui divisent une distribution gaussienne en un nombre arbitraire de zones de même taille). Ainsi, le nombre de points de rupture β_i correspond à la taille de l'alphabet $\mathcal{A}$ (dictionnaire de mots). L'intervalle entre deux points de rupture successifs est attribué à un symbole de l'alphabet, et chaque segment de l'AAP à l'intérieur de cet intervalle est codé par ce symbole.

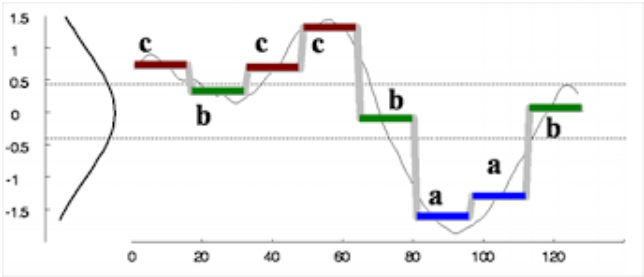


FIGURE 4.16 – Exemple d'une série chronologique encodée en une séquence SAX. À chaque position de la fenêtre, la valeur moyenne est calculée puis encodée avec un symbole.

La figure 4.16 donne un exemple de série chronologique numérique et sa séquence SAX relative. Dans cet exemple, la longueur du mot de code $w = 8$ (huit positions de la fenêtre ou fractionnées selon la dimension temporelle), et la longueur du livre de codes $a = 3$ (trois symboles de l'alphabet). La séquence SAX de cette série est $CBCCBAAB$. Il apparaît clairement qu'une telle représentation est très utile puisque l'acquisition de données et leur représentation en séries temporelles numériques peuvent engendrer des erreurs intimement liées aux capteurs ou au protocole d'acquisition. Il apparaît également évident qu'avec les séquences symboliques, nous pouvons tirer parti de la robustesse des méthodes d'exploration des données symboliques et de traitement du langage naturel, telles que la recherche de similitude, la découverte de motifs, l'exploration de motifs fréquents,

la construction de trajectoires comportementales, etc.

Séquences symboliques par PMVQ

La quantification vectorielle (VQ) est une transformée en ondelettes qui a été largement utilisée en traitement d'images, en traitement du signal et de l'encodage vidéo [Mal89]. C'est une méthode efficace de compression des données hétérogènes [Mac67]. Plus les données sont massives, plus la compression est efficace. Elle est basée sur l'extraction de la corrélation spatiale perceptive par des transformations en ondelettes. En effet, la transformation en ondelettes permet de rendre compte de l'information locale et globale à différentes résolution de la série. De plus, elle permet une représentation symbolique de la série en codant les séquences de la série d'origine par une suite de mots. Contrairement à la méthode SAX, la quantification vectorielle multirésolution utilise un ensemble de vocabulaire composé de mots. Le vocabulaire est différent pour chaque résolution. À chaque résolution, VQ est utilisé pour découvrir le vocabulaire des séquences liées aux séries initiales. Un vecteur de quantification Q, de dimension n et de taille k est une fonction de correspondance $Q : \mathbb{R}^n \rightarrow \mathcal{A}$ à partir d'un point de $\mathbb{R}^n$ vers un ensemble fini de mots $\mathcal{M} = \{m_1, ..., m_k\}$, le dictionnaire de mots (ou *codebook*) contenants k mots $m_i \in \mathbb{R}^n$.

À chaque $k - points$ composant une séquence de la série, VQ est une partition de $\mathbb{R}^n$ en k régions R_i pour tout $i = \{1, 2, .., k\}$ avec $R_i = \{x \in \mathbb{R}^n : Q(x) = m_i\}$. La fonction de correspondance Q doit fournir pour tout $k - points$:

1. la partition optimale pour un codebook $\mathcal{M}$ donné ;
2. le mot m_i optimal pour une partition R_i donnée.

Les paramètres de la méthode MVQ sont l'échelle de résolution k et la longueur d'un mot m_i. La valeur de l'échelle est calculée d'une manière intuitive par $log(n)$, n étant la dimension de la série. La longueur d'un mot m_i pour une résolution i donnée est 2^{i-1}.

Enfin, ces deux algorithmes ont été testés et leur performance prédictive a été testée en utilisant le modèle *seq2seq LSTM* comme réseau de neurones profonds. Chaque série temporelle d'un canal de diffusion des offres d'emploi est transformée en une séquence symbolique SAX et PMVQ. Différentes résolutions ont été testées. Par exemple, avec 1000

séries temporelles et 8 résolutions pour SAX et PMVQ, nous pouvons obtenir 1000x2x8 = 16000 séquences symboliques employées à l'apprentissage du modèle LSTM.

Les figures 4.17 et 4.18 illustrent la représentation spectrale de certains canaux de diffusion d'offres d'emploi avec les deux méthodes de codage lors de la production des séries symboliques.

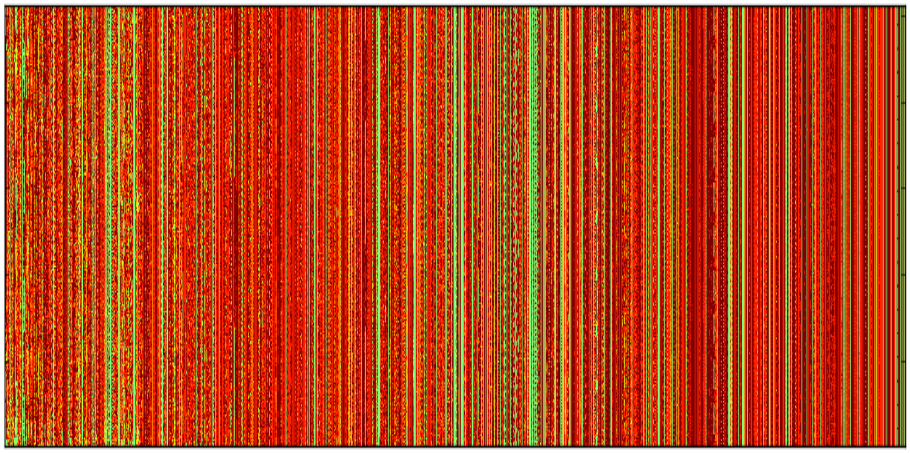


FIGURE 4.17 – Représentation spectrale de certains canaux de diffusion d'offres d'emploi avec la série symbolique SAX. Chaque ligne verticale représente une séquence symbolique d'un canal. Chaque pixel de la ligne représente la quantification des clics avec la série symbolique SAX

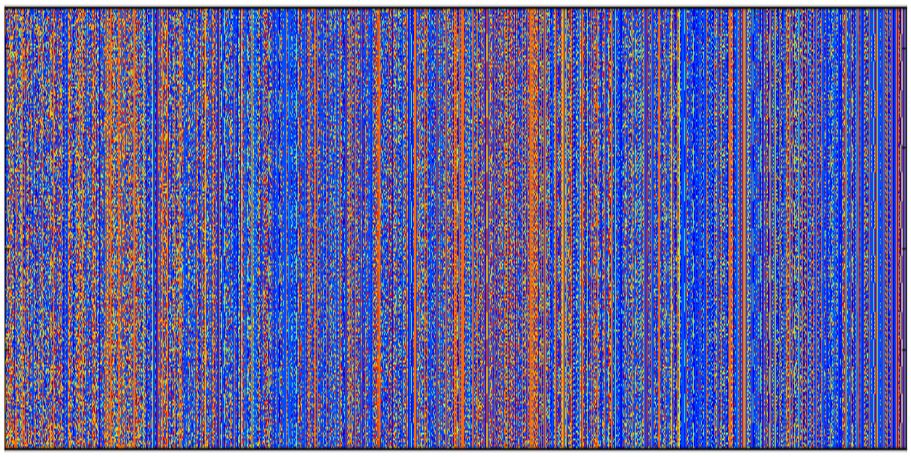


FIGURE 4.18 – Représentation spectrale de certains canaux de diffusion d'offres d'emploi avec la série symbolique PMVQ. Chaque ligne verticale représente la séquence symbolique d'un tableau d'affichage des offres d'emploi. Chaque pixel de la ligne représente la quantification des clics avec les PMVQ.

Il s'agit d'une représentation nouvelle et innovante que nous proposons pour avoir une vue d'ensemble globale sur la base de données des séries chronologiques et pour analyser

visuellement les comportements des candidats face aux offres d'emploi. Chaque ligne verticale représente une séquence symbolique d'un canal de diffusion. Chaque pixel de la ligne représente la quantification des clics avec la méthode d'encodage associée. Nous avons divisé les séquences en sous-séquences d'apprentissage et de validation afin de prédire les sous-séquences de symboles futurs comparées aux sous-séquences réelles. Avec une résolution de 8, on a obtenu des erreurs RMSE faibles. Nous avons constaté également qu'avec la méthode PMVQ, la prédiction des symboles de quantification des clics futurs est plus efficace que la méthode SAX. Une précision moyenne de $0,89$ est obtenue pour la méthode PMVQ contre $0,73$ pour la méthode SAX. La précision moyenne a été calculée pour des résolutions moindre avec SAX et PMVQ, et les résultats sont illustrés par la figure 4.19.

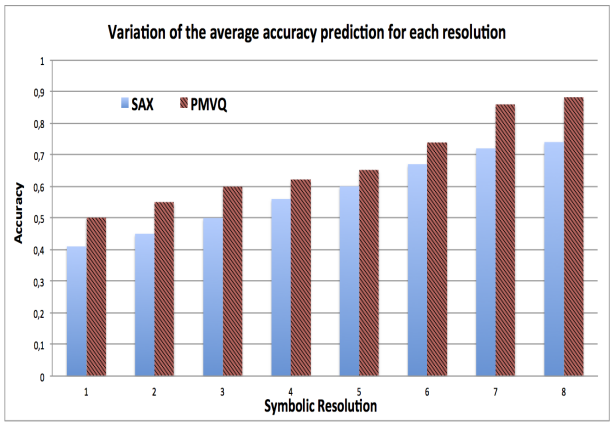


FIGURE 4.19 – Valeurs de précision moyenne de la prédiction avec des LSTM entraînés sur des séquences symboliques. Les résultats sont affichés pour chaque résolution (allant de 1 à 8) avec SAX et PMVQ.

Il ressort de cette analyse que PMVQ est plus efficace pour la prédiction de nouveaux symboles que SAX. En outre, nous pouvons observer qu'avec les réseaux neuronaux profonds les résultats de la prédiction sont meilleurs que ceux obtenus avec les N-grammes. Le détail de ces résultats a fait l'objet d'une publication dans le journal KBS [BML18].

4.7 Conclusions et premières perspectives

Dans ce chapitre, nous avons abordé la prédiction d'événements futurs en mettant les données au service de la construction du modèle. En d'autres termes, nous avons combiné les techniques d'extraction de connaissances à partir de données massives, hétérogènes, tem-

porelles, avec des techniques d'apprentissage exploitant ces connaissances pour construire un modèle de prédiction idoine. Ces modèles se construisent de manière adaptative aux connaissances cachées dans les données. Les réseaux de neurones profonds sont des méthodes qui réussissent plus ou moins bien cet exercice dès lors qu'ils disposent d'un capital de données conséquent en taille et riche en informations. Or, dans certains domaines, les données se font rares et peuvent être parfois coûteuses voire lentes à produire. Face à ce problème, des pistes de recherche sont très en vogue actuellement et consistent à combler le manque de données par des approches de transfert d'apprentissage. Cette voix ouvre de nouvelles perspectives de recherche et la résolution de certains verrous scientifiques sont alors envisagés. Par exemple, nous nous intéressons à l'usage des ontologies formelles pour exprimer le problème de transfert comme un problème d'alignement entre contextes applicatifs source et cible à différentes échelles. De plus, ce cadre d'alignement devra prendre en considération l'incertitude, l'incomplétude et l'incohérence des données, en particulier les données cibles. Par ailleurs, les données sont de plus en plus hétérogènes en type. Pour une même information, on peut disposer de données textuelles, de données images, de son, etc. La manière de fusionner ces données pour augmenter sémantiquement la précision du modèle de prédiction est un axe de recherche que nous poursuivons dans l'optique de rendre la prédiction explicable. Pour l'heure, les modèles prédictifs à base de réseaux de neurones se limitent à prédire l'événement futur sans aucune production d'un ou plusieurs schémas explicatifs des résultats. C'est vers cet aspect que nous souhaitons orienter nos perspectives scientifiques.

Chapitre 5

Conclusions et projets scientifiques

Une synthèse d'une partie de mes travaux de recherche a été présentée dans ce manuscrit. Comme expliqué dans l'introduction, mes travaux s'inscrivent dans la problématique de l'extraction de connaissances pour l'aide à la prédiction sémantique à partir de données complexes. Les données sont principalement massives et multimodales. Une hypothèse forte qui compose le fil directeur de mes travaux est de proposer une vraie rupture avec le traitement traditionnel des données. Il s'agit de construire d'une part des modèles a priori, implicites de connaissances sous-jacentes expertes et, d'autre part, d'utiliser des sources d'informations contextuelles. L'ensemble de ces données augmentées par des connaissances implicites constitue un capital d'information sémantique contribuant à augmenter la performance des modèles d'analyse prédictive et leur interprétation.

Un autre axe directeur de mes travaux de recherche est la prise en compte des différentes facettes des sources d'informations et de connaissances disponibles afin de construire des descripteurs sémantiques en tenant compte de l'**infobésité** des données et leur caractère imparfait. Par conséquent, plusieurs questions se posent relatives à :

1. l'acquisition et la construction de modèles de connaissances à partir de données complexes et d'autres sources d'informations disponibles ;
2. la construction de descripteurs sémantiques utiles à la représentation des connaissances du domaine, à la recherche sémantique et à l'interprétation de l'analyse prédictive sur les données ;

3. l'hétérogénéité des modèles de connaissances disponibles pour un même domaine applicatif et qui doivent continuer à co-exister tout en favorisant leur interopérabilité, leur accessibilité et à contribuer pleinement à l'enrichissement des connaissances transversales entre domaines ;

4. la prise en compte du facteur multimodal et multi-échelle des données visées dans le processus d'analyse prédictive ;

5. l'explicabilité des algorithmes d'analyse et d'aide à la décision qui est un problème depuis longtemps central, dans le domaine des systèmes d'expert en particulier, et s'avère cruciale dans de nombreux domaines comme par exemple en recherche d'information, en recommandation ou en prédiction.

Les différentes contributions décrites dans le manuscrit constituent des éléments de réponse à une ou plusieurs des questions ci-dessus avec pour dénominateur commun la mise au premier plan des modèles d'extraction de connaissances basés sur l'apprentissage piloté par les données et les logiques sous-jacentes de représentation et de raisonnement. Certaines de ces contributions sont plutôt d'ordre théorique avec la proposition de nouveaux formalismes ; d'autres sont plutôt d'ordre applicatif avec la proposition de nouvelles méthodes, souvent motivées par l'application et validées par des études expérimentales. Ce chapitre propose donc de rapides conclusions sur mes contributions et dresse un panorama de mes perspectives et projets de recherche à plus ou moins long terme. La motivation principale de ces perspectives reste le développement de nouvelles approches pour donner du sens aux données et de contribuer pleinement dans la réduction du fossé sémantique par l'enrichissement des connaissances, dans l'explicabilité des algorithmes de traitement et d'analyse des données complexes. Alors que l'image a été la portée principale de mes travaux passés, j'envisage très fortement d'étendre mes recherches à d'autres types de données comme, par exemple, les séquences d'images, les vidéos, le son dont la dimension dynamique, d'origine temporelle et spatiale, devra être considérée. La gestion des imperfections ou d'autres facettes de l'information, comme, par exemple, leur caractère bipolaire, restera aussi au coeur de mes actions de recherche. Ces conclusions et perspectives sont présentées en suivant une structure similaire à celle du reste du manuscrit, à savoir la réponse aux questions posées dans l'introduction et reformulées ci-dessus

5.1 Fusion Multimodale et explicabilité des algorithmes

Avec la récente résurgence des réseaux de neurones, le rapide essor des méthodes d'apprentissage profond et la prolifération de données massives non annotées, les algorithmes non supervisés ont gagné en popularité de part leur faculté à extraire de l'information depuis ces données. Les méthodes modernes d'apprentissage profond ainsi que les récentes évolutions matérielles (notamment *GPU*) permettent un apprentissage par des réseaux de neurones depuis des données quasiment brutes, c'est-à-dire sans la fastidieuse et coûteuse opération manuelle d'extraction de caractéristiques jusqu'à la prédiction de la tâche finale. Ainsi ces réseaux apprennent-ils de bout en bout à la fois une représentation des données ainsi que sa projection vers la prédiction sous-jacente à la tâche finale facilitant ainsi leur déploiement. L'apprentissage profond et donc l'apprentissage automatique de représentations n'est pas en vogue uniquement dans le domaine de la vision par ordinateur. D'autres domaines tels que le traitement des langues naturelles, la reconnaissance de la parole, les systèmes de recherche d'information multimodales et bien d'autres suivent cette tendance. Les objectifs scientifiques ont évolué de l'extraction de descripteurs et combinaison de classifieurs au développement d'architectures neuronales qui apprennent automatiquement des représentations et les exploitent efficacement pour une multitude de tâches différentes. Différents types d'architecture de réseaux profonds existent et chacune d'entre elles est dédiée à une classe de problème spécifique. L'avantage majeur de ces architectures apparaît lorsqu'on combine différent types de réseaux pour former des architectures complexes, capables de traiter différentes données, en apprendre des représentations efficaces, les combiner ou les transformer pour être efficace sur différentes tâches. C'est l'aspect le plus intéressant en apprentissage non-supervisé où de tels réseaux peuvent être entraînés en exploitant les ressources considérables de données non annotées accessibles sur le net, sans nécessité de recourir à des processus coûteux d'annotation manuelle. Les exemples typiques inclus le traitement multimodal non-supervisé de vidéos non annotées où les réseaux de neurones profonds sont capables de fusionner les informations obtenues depuis la transcription automatique de la parole et les représentations visuelles obtenues par apprentissage supervisé dans le but d'apparier correctement des segments de vidéo. Les travaux présentés dans la première partie de ce manuscrit ne constituent cependant que la première phase d'un

projet à plus long terme dont la finalité est de proposer de nouveaux formalismes permettant la fusion de grosses collections de données multimodales. Mes travaux s'orientent vers l'hybridation de modèles d'apprentissage supervisé et non supervisé par plongements neuronaux (*neural embedding*) pour la résolution du problème de fusion multimodale. Parallèlement, ces architectures présentent un défaut majeur lié au principe de boîte noire et à l'opacité du procédé d'élaboration des connexions significatives qui donne sens aux résultats obtenus en sortie. Ce procédé de calcul du sens, nous envisageons de le formaliser par une dimension sémiotique. Nous partons donc de la thèse que la perception et l'interprétation d'un artefact culturel (texte, image, film, site web, réseau social numérique, etc.) sont certes structurées par des **grammaires de production** : une articulation stratégique des signes à l'intérieur de l'artefact, qui agissent potentiellement sur le regard du récepteur comme des éléments de guidage. La construction de ces grammaires par le sujet est fonction de ses schémas de croyance et d'appartenance. Elles ne sont pas forcément universelles et résultent des représentations sociales et des schémas implicites liés au sujet et de ses automatismes cognitifs. La reconnaissance d'un objet n'implique pas forcément son interprétation. Nous proposons donc une approche complètement innovante pour la construction de ce processus d'interprétation fondée sur le plongement des algorithmes d'apprentissage dans une contextualisation sémiotique liées aux éléments perceptifs. Ce travail fait l'objet d'un projet de recherche financée par l'Université Paris 8, en collaboration avec Alexandra Saemmer, Professeure en Information et Communication. Dans ce projet, nous proposons de faire dialoguer des approches d'intelligence artificielle avec des approches sémiotiques sociales du texte et de l'image pour la production du sens. Nous proposons d'analyser tout particulièrement les couches des réseaux de neurones profonds correspondants aux descripteurs qui caractériseraient les données. Les données en entrée, les descripteurs ainsi que les labels (classes) en sortie sont représentés par des matrices. Les entrées de la matrice sont des degrés auxquels les objets représentés par les lignes, satisfont les descripteurs représentés par les colonnes. Les degrés de satisfaction sont supposés former une échelle bornée sur l'intervalle [0,1] équipée de certains opérateurs d'agrégation et conforme à la structure d'un treillis résiduel complet. Les degrés de satisfaction sont tout simplement les poids d'activation des descripteurs pour un objet. Partant de ces hypothèses, le processus d'interprétation, vu comme un problème d'abduction, utilise l'analyse

formelle de concepts pour trouver les descripteurs optimaux caractérisant les contextes formels des données homogènes. Pour pallier le problème des données hétérogènes, nous explorons également l'analyse relationnelle de concepts (ARC) qui est une extension de l'AFC où les concepts sont décrits en faisant référence à d'autres concepts via des descripteurs formés par abstraction des liens inter-objets. L'AFC est une technique pour extraire des concepts entre deux ensembles ordonnés interdépendants. Elle est souvent utilisée dans le cadre de données binaires où le lien ne peut être que 0 ou 1 entre les observations et les descripteurs. Nous proposons de l'étendre dans un contexte plus générique où le lien peut être interprété comme une probabilité de dépendance ou bien une mesure de corrélation, ou bien une mesure de relation floue. D'autre part, nous envisageons également son extension pour la prise en compte de concepts temporels. Ce travail théorique sera proposé dans un projet européen en cours de montage et qui sera piloté par l'Université Paris 8.

5.2 Image et recherche visuelle sémantique

Dans la recherche d'images basée sur le contenu, le problème clé est de savoir comment mesurer efficacement la similarité entre images. Comme les objets visuels ou les scènes peuvent subir divers changements ou transformations, il est impossible de comparer directement les images au niveau des pixels. Comme nous l'avons exposé dans le chapitre 3, nous proposons des modèles enrichis à base de motifs (*patterns*) sémantiques agrégeant différentes dimensions de représentation du sens telles que la dimension spatiale, géographique, temporelle, émotionnelle, etc. Chacune de ces dimensions représente un point de vue sur lequel une requête utilisateur peut être exprimée. Compte tenu de la contradiction entre une base de données d'images à grande échelle et la nécessité d'obtenir une réponse efficace à la requête, il est nécessaire de *packager* les caractéristiques visuelles pour fabriquer une signature visuelle sémantique multidimensionnelle facilitant l'indexation et la comparaison d'images. Pour atteindre ce but, la quantification avec des approches d'alignement est vue comme un alignement de modèles de connaissances couramment utilisé dans le domaine d'alignement d'ontologies. Parallèlement et avec l'évolution des standards de représentation des données liées aux images, je propose d'adapter le format IIIF (*International Image Interoperability Framework*) pour la représentation des patterns sémantiques

discutée au chapitre 3. En effet, IIIF désigne à la fois une communauté et un ensemble de spécifications techniques dont l'objectif est de définir un cadre d'interopérabilité pour la diffusion d'images de haute résolution sur le Web. Au delà des images statiques, j'envisage d'étendre les patterns sémantiques pour la représentation du contenu vidéo. Compte tenu de la dynamique des images composant une vidéo, j'oriente l'extension des patterns sémantiques vers des séquences de patterns sémantiques ordonnées dans le temps. L'extraction d'une séquence minimale de patterns sémantiques est vue comme un procédé de résumé sémantique d'une vidéo. À ce propos, je considère que les techniques de visualisation par l'image pourront :

— faciliter l'identification des relations de causalité dans des bases de données complexes (car multidimensionnelles et multimédia) ;

— communiquer les résultats sous une forme visuelle et synthétique.

D'où la question, peut-on représenter le son par une image ou bien par une séquence d'images ? En effet, le nouveau domaine des « MIR » (*Musique Information Retrieval*) et de l'audio sur le Web ouvre des perspectives stimulantes dans le domaine de la création musicale telles que l'identification automatique d'éléments sonores similaires, la détection parole/musique, des locuteurs, la séparation de sources, la création d'espaces collaboratifs pour la création musicale, etc. Une approche pluridisciplinaire et normalisée de la description des contenus sonores et audiovisuels permettra de croiser largement les données et les analyses, et donc d'en faciliter les accès. Dans cette optique, je m'intéresse à l'analyse audio via leur transformation en images. Les signaux sonores, tout comme les images, disposent de plusieurs échelles de représentation ou des points de vues de représentation de leur contenu. C'est dans cette optique que j'envisage l'analyse du signal à travers les images de décomposition du signal à différentes échelles dans le but de comprendre le liens entre le signal et sa perception. Ce travail fera l'objet d'une collaboration pluridisciplinaire autour d'un projet ANR soumis en 2020 (Attente du résultat final).

5.3 Apprentissage par transfert

Dans le chapitre 4, nous avons abordé la prédiction d'événements futurs en mettant les données au service de la construction du modèle. En d'autres termes, nous avons combiné les techniques d'extraction de connaissances à partir de données massives, hétérogènes, temporelles, avec des procédés d'apprentissage exploitant ces connaissances pour construire un modèle de prédiction idoine. Ces modèles se construisent d'une manière adaptative aux connaissances implicites dans les données. Les réseaux de neurones profonds sont des méthodes qui réussissent plus ou moins bien cet exercice dès lors qu'ils disposent d'un capital de données conséquent en taille et riche en informations. Or, dans certains domaines, les données se font rares et peuvent être parfois coûteuses voire lentes à produire. Face à ce problème, des pistes de recherche sont très en vogue actuellement et consistent à combler le manque de données par des approches de transfert d'apprentissage. Cette voie ouvre de nouvelles perspectives de recherche sont alors envisagés mais abordant de nouveaux verrous scientifiques à lever. Par exemple, nous nous intéressons à l'usage des ontologies formelles pour exprimer le problème de transfert comme un problème d'alignement entre contextes applicatifs source et cible à différentes échelles. De plus, ce cadre d'alignement devra prendre en considération l'incertitude, l'incomplétude et l'incohérence des données, en particulier les données cibles. C'est en effet on se pose tout d'abord la question centrale de comment quantifier la qualité des données sources et cibles. Ce sont les critères tels que la variété et la véracité que nous étudions tout particulièrement dans le contexte de l'apprentissage par transfert. Contrairement aux pratiques classiques, nous proposons d'analyser et de considérer ces deux critères sur un couple d'échantillons de données (source, cible). Formellement, le problème d'apprentissage par transfert d'un domaine $\mathcal{D}_s$ vers un domaine cible $\mathcal{D}_T$, est défini par :

— un espace $\mathcal{X}_s$ d'observations, un espace de labels $\mathcal{Y}_s$

— une tâche d'apprentissage $\mathcal{T}_s = \mathcal{Y}_s, f_s(.)$ et $\mathcal{T}_T = \mathcal{Y}_T, f_T(.)$ où $f_s(.)$ (resp. $f_T(.)$) est une fonction de prédiction apprise à partir de l'ensemble des couples x_i, y_i, pour tout $x_i \in X_s$ (resp. X_T) et $y_i \in Y_s$ (resp. Y_T)

— par une distribution de probabilité marginale $P(X)$ pour tout $X = (x_1, .., x_n \in \mathcal{X}_s$ (resp. $\mathcal{X}_s$)

— $\mathcal{D}_s = (x_1^s, y_1^s), .., (x_n^s, y_n^s)$ et $\mathcal{D}_T = (x_1^T, y_1^T), .., (x_n^T, y_n^T)$.

Compte tenu de la définition formelle du contexte de transfert, l'apprentissage par transfert est le processus visant à améliorer la fonction prédictive $f_T(.)$ en utilisant les informations relatives à $\mathcal{D}_s$ et $\mathcal{D}_T$ quand $\mathcal{D}_s \neq \mathcal{D}_T$ ou $\mathcal{T}_s \neq \mathcal{T}_T$. Deux situations sont possibles quand $\mathcal{D}_s \neq \mathcal{D}_T$: $\mathcal{X}_s \neq \mathcal{X}_T$ et/ou $P(X_s) \neq P(X_T)$. Dans le premier cas, l'apprentissage par transfert doit prendre en charge la contrainte d'hétérogénéité des données. Dans le second cas où $P(X_s) \neq P(X_T)$, la distribution marginale est complètement différente entre la source et la cible. Il se pose donc la question de l'efficacité du transfert. Nous proposons d'approfondir ce cadre formel de l'apprentissage par transfert pour le passage d'un domaine pour lequel on dispose d'un nombre d'étiquettes ou de labels associées aux données de sortie source vers un domaine cible où les labels sont très limités, voire inexistants. Nous étudions en particulier le transfert par apprentissage non supervisé vers un domaine cible renforcé par un apprentissage supervisé sur les données sources dans le but d'augmenter la sémantique au niveau des données cibles. Ce travail est inspiré des travaux de Pan [PNtS+10] [PY10] [PLXY11] qui font référence à deux types de transfert à savoir inductif et transductif. L'expérimentation de cette problématique porte sur la détection de l'incohérence dans des documents textuels. C'est grâce à une collaboration avec mon collègue Aurélien Bossard autour de son projet ANR JCJC *ASADERA* que nous avons proposé des comparaisons de modèles de transfert sur des documents multilingue avec une extension sur les documents cross-lingue.

Références

[Agn14] Vijay Agneeswaran. *Big Data Analytics Beyond Hadoop : Real-Time Applications with Storm, Spark, and More Hadoop Alternatives*. Pearson FT Press, USA, 1st edition, 2014.

[AMB+15] O. Allani, N. Mellouli, H. Baazaoui-Zghal, H. Akdag, and H. BenGhzala. A relevant visual feature selection approach for image retrieval. In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, pages 377–384, 2015.

[AMBA17] O. Allani, N. Mellouli, H. Baazaoui-Zghal, and H. Akdag. Sélection ciblée des descripteurs visuels pour la recherche d'images : une approche basée sur les règles d'association. In *Revue des Nouvelles Technologies de l'Information, vol. Extraction et Gestion des Connaissances, RNTI-E-33*, 2017.

[ATB+19] Arafat Salih Aydiner, Ekrem Tatoglu, Erkan Bayraktar, Selim Zaim, and Dursun Delen. Business analytics and firm performance : The mediating role of business process performance. *Jindal Journal of Business Research*, 96 :228–237, 2019.

[AYM+15] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K.Taha. Efficient machine learning for big data : A review. *Big Data Research*, 2(3) :87–93, 2015.

[Aze18] C.-A. Azencott. *Introduction au Machine Learning*. 2018.

[Bar04] M. Bar. Visual objects in context. *Nature Reviews Neuroscience*, 5(8) :617–29, 2004.

[BCD+14] A. K. Baughman, W. Chuang, K. R. Dixon, Z. Benz, and J. Basilico. The class imbalance problem : a systematic study. In *IEEE Transactions on Computational Intelligence and AI in Games*, volume 6, no 1, pages 55–66, 2014.

[Bel11] L. Belouaer. *Représentation de la connaissance spatiale pour la planification*. Thèse de doctorat, Université de Caen, 2011.

[BH11] H. Bannour and C. Hudelot. Toward semantic ontologies for image interpretation and annotation. In 9th *International Workshop CBMI*, pages 211–216, 2011.

[BH19] Tal Ben-Nun and Torsten Hoefler. Demystifying parallel and distributed deep learning : An in-depth concurrency analysis. *ACM Comput. Surv.*, 52(4), aug. 2019.

[Bie72] I. Biederman. Perceiving real-world scenes. *Science*, 177 :77–80, 1972.

[Bie87] I. Biederman. Recognition-by-components : A theory of human image understanding. *Psychological Review*, 94 :115–147, 1987.

[BKT00] Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Data provenance : Some basic issues. In Sanjiv Kapoor and Sanjiva Prasad, editors, *Foundations of Software Technology and Theoretical Computer Science,* 20th *Conference, FST TCS 2000 New Delhi, India, December 13-15, 2000, Proceedings*, volume 1974 of *Lecture Notes in Computer Science*, pages 87–93. Springer, 2000.

[BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American Magazine*, May 17, 2001.

[BLR96] C-L. Benhamou, E. Lespessailles, and V. Royant. Architecture osseuse et résistance mécanique du tissu osseux. *Presse Médicale*, 25(6) :249–254, 1996.

[BMC+19] H. Balti, N. Mellouli, I. Chebbi, I. Farah, and M. Lamolle. Deep semantic feature detection from multispectral satellite images. In *KDIR 2019*, 2019.

[BML18] S. Benabderrahmane, N. Mellouli, and M. Lamolle. On the predictive analysis of behavioral massive job data using embedded clustering and deep recurrent neural networks. *Knowledge-Based Systems*, 151 :95–113, 2018.

[Bor15] A. Borji. What is a salient object ? a dataset and a baseline model for salient object detection. *IEEE Transactions on Image Processing*, 24(2) :742–756, 2015.

[Bou09] G. Bouma. Normalized (pointwise) mutual information in collocation extraction. In *The Biennial GSCL Conference*, volume 156, 2009.

[BQG14] Sidahmed Benabderrahmane, Rene Quiniou, and Thomas Guyet. Evaluating distance measures and times series clustering for temporal patterns retrieval. In *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration, IRI 2014, Redwood City, CA, USA, August 13-15, 2014*, pages 434–441, 2014.

[BRL+94] C.L. Benhamou, R.Harba, E. Lespessailles, E. Jacquet, D. Toulire, and R. Jennane. Fractal organisation of trabecular bone images on calcaneus radiographs. *JBMR*, 9 :1909–1918, 1994.

[CD14] Michele Chambers and Thomas W. Dinsmore. *Modern Analytics Methodologies : Driving Business Value with Analytics.* FT Press, 1st edition, 2014.

[CdFB04] P. Carbonetto, N. de Freitas, and K. Barnard. A statistical model for general contextual object recognition. In Springer Lecture Notes in Computer Science, editor, *the European Conference on Computer Vision*, pages 350–362, 2004.

[CDP15] D. Coulomb, J-L Dupont, and A. Pichard. The role of refrigeration in the global economy, in 29th informatory note on refrigeration technologies. In *International Institute of Refrigeration (IIR)*, nov. 2015.

[CF99] Kin-pong Chan and Ada Wai-Chee Fu. Efficient time series matching by wavelets. In Masaru Kitsuregawa, Michael P. Papazoglou, and Calton Pu, editors, *Proceedings of the 15th International Conference on Data Engineering, Sydney, Australia, March 23-26, 1999*, pages 126–133. IEEE Computer Society, 1999.

[CG16] C. Calude and L. Giuseppe. The deluge of spurious correlations in big data. In *Lois des dieux, des hommes et de la nature*, pages 1–18, Nantes, France, 2016.

[CGTon] JM. Coquery, P. Gailly, and N. Tajeddine. *Neurosciences et cognition.* NOTO, 5ème édition.

[Cho17] F. Chollet. *Deep Learning with Python.* Manning Publications Company, 2017.

[CKMP02] Kaushik Chakrabarti, Eamonn J. Keogh, Sharad Mehrotra, and Michael J. Pazzani. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Trans. Database Syst.*, 27(2) :188–228, 2002.

[CL17] C. S. Calude and G. Longo. The deluge of spurious correlations in big data. *Foundations of Science*, 22 :595–612, 2017.

[Cor08] M. Cori. Des méthodes de traitement automatique aux linguistiques fondées sur les corpus. *Artificial Intelligence*, 3 :95–110, 2008.

[CPK+18] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille. Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transac- tions on pattern analysis and machine intelligence*, pages 834–848, 2018.

[CPSA17] L.C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint, 1706.05587*, 2017.

[CS99] S. Coradeschi and A. Saffiotti. Anchoring symbols to vision data by fuzzy logic. *ECSQARU, Springer*, 1638 of LNCS(1) :104–115, 1999.

[CS11] F. M. Couto and M. J. Silva. Disjunctive shared information between ontology concepts : application to gene ontology. *Journal of biomedical semantics*, 2(1), 2011.

[CUD19] K. Cemil, A. Uyar, and D. Delen. An investigation of the factors influencing cost system functionality using decision trees, support vector machines and logistic regression. *International Journal of Accounting and Information Management*, 27 :27–55, 2019.

[CV07] R. L. Cilibrasi and P. MB. Vitanyi. The google similarity distance. *IEEE Transactions on knowledge and data engineering*, 19(3), 2007.

[CYY+18] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han. When deep learning meets metric learning : Remote sensing image scene classification via learning discriminative cnns. *IEEE Trans. Geoscience and Remote Sensing*, 56(5) :2811–2821, 2018.

[dAB15] P. D. C. de Almeida and J. Bernardino. Big data open source platforms. In *Proceedings of the 2015 IEEE International Congress on Big Data*, pages 268–275, 2015.

[DBRA07] Jérôme Darmont, Omar Boussaid, Jean-Christian Ralaivao, and Kamel Aouiche. An architecture framework for complex data warehouses. *CoRR*, abs/0707.1534, 2007.

[Del99] Alberto Del Bimbo. *Visual information retrieval*. Morgan and Kaufmann, 1999.

[Del15] D. Delen. *Real-World Data Mining : Applied Business Analytics and Decision Making*. Upper Saddle River, New Jersey : Pearson Education LTD, 2015, San Francisco, CA, USA, 2015.

[DHS16] J. Dai, K. He, and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. In *The IEEE Conference on Computer Vision and Pattern Recognition*, volume 156, pages 3150–3158, 2016.

[DJM18] N. Dvornik and C. Schmid J. Mairal. Modeling visual context is key to augmenting object detection datasets. In Springer Lecture Notes in Computer Science, editor, *the European Conference on Computer Vision*, pages 375–391, 2018.

[DM14] P. B. Dongre and L. G. Malik. A review on real time data stream classification and adapting to various concept drift scenarios. In *Proceedings of the 2014 IEEE International Advance Computing Conference (IACC)*, pages 533–537, 2014.

[EAM14] S. Elavarasi, Dr. J. Akilandeswari, and K. Menaga. A survey on semantic similarity measure. *International Journal of Research in Advent Technology*, 2(3), 2014.

[ED04] M. Effros and D. Dugatkin. Multiresolution vector quantization. *IEEE TRANSACTIONS ON INFORMATION THEORY*, 50 :3130–3145, 2004.

[EFH+14] J. A. Evans, A. M. Foster, J. M. Huet, L. Reinholdt, K. Fikiin, C. Zilio, M. Houska, A. Landfeld, C. Bond, M. Scheurs, and T. W. M. Van Sambeeck. Specific energy consumption values for various refrigerated food cold stores. *Energy and Buildings*, 74 :141–151, 2014.

[EHG+10] H. J. Escalante, C. A. Hernández, J. A. Gonzalez, A. López-López, M. Montes, E. F. Morales, L. E. Sucar, L. Villaseñor, and M. Grubinger. The segmented and annotated iapr tc-12 benchmark. *Computer Vision and Image Understanding*, 114(4) :419–428, 2010.

[ET97] B. Efron and R. Tibshirani. Improvements on cross-validation : The .632+ bootstrap method. *Journal of the American Statistical Association*, 438(3) :548–560, 1997.

[Far19] M. Farouk. Measuring sentences similarity : A survey. *Indian Journal of Science and Technology*, 12(25), 2019.

[FB13] Wei Fan and Albert Bifet. Mining big data : Current status, and forecast to the future. *SIGKDD Explorations Newsletter*, 14(2) :1–5, apr 2013.

[FKZ07] K. Fundel, R. Küffner, and R. Zimmer. Relation extraction using dependency parse trees. *Bioinformatics*, 23(3) :365–371, 2007.

[FPS96] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11) :27–34, 1996.

[FR07] B. Fallery and F. Rodhain. Quatre approches pour l'analyse de données textuelles : lexicale, linguistique, cognitive, thématique. In *XVIème Conférence de l'Association Internationale de Management Stratégique AIMS*, pages 1–17, 2007.

[Fro90] H. M. Frost. Skeletal structural adaptations to mechanical usage. *The Anatomical Record Journal*, 226 :403–422, 1990.

[GB10] C. Galleguillos and S. J. Belongie. Context based object categorization : A critical survey. *Computer Vision and Image Understanding*, 114(6) :712–722, 2010.

[GDDM14] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[GH14] A. Gandomi and M. Haider. Beyond the hype : Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2) :137–144, 2014.

[GHH+14] K. Grolinger, M. Hayes, W. A. Higashino, A. LḦeureux, D. S. Allison, and M. A. M. Capretz. Challenges for map reduce in big data. In *Proceedings of 2014 IEEE World congress on Services*, pages 182–189, 2014.

[GHTC13] Katarina Grolinger, Wilson A. Higashino, Abhinav Tiwari, and Miriam A. M. Capretz. Data management in cloud environments : Nosql and newsql data stores. *Journal of Cloud Computing*, 2 :22, 2013.

[GOO+17] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv :1704.06857*, 2017.

[GR02] J. Greenberg and W.D. Robertson. Semantic web construction : an inquiry of authors' views on collaborative metadata generation. In *of the 2002 Int. Conf. on Dublin core and metadata applications : Metadata for e-communities : supporting diversity and convergence*, pages 45–52, Dublin Core Metadata Initiative, 2002.

[GWC+13] M. Ghanavati, R. K. Wong, F. Chen, Y. Wang, , and C-S. Perng. An effective integrated method for learning big imbalanced data. In *IEEE International Congress on Big Data*, pages 691–698, 2013.

[GZB+14] João Gama, Indre Zliobaite, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4) :44 :1–44 :37, 2014.

[HAB08] C. Hudelot, J. Atif, and I. Bloch. Fsro : une ontologie de relations spatiales floues pour l'interprétation d'images. In *Revue des Nouvelles Technologies de l'Information*, 2008.

[Har79] Robert M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5) :786–804, 1979.

[HHM99] C. J. Hernandez, S. J. Hazelwood, and R. B. Martin. The relationship between basic multicellular unit activation and origination in cancellous bone. *Bone*, 25(5) :585–587, 1999.

[HK98] J. Han and M. Kamber. Data mining methods for knowledge discovery. *Kluwer*, 1998.

[HK06] J. Han and M. Kamber. Data mining : Concepts and techniques. *Morgan Kaufmann*, 2006.

[HK13] M. Hofmann and R. Klinkenberg, editors. *Rapid Miner, Data Mining Use Cases and Business Analytics Applications*. Taylor & Francis Group, LLC CRC Press, 2013.

[HS98] G. Hirst and D. St-Onge. Lexical chains as representations of context for the detection and correction of malapropisms. In *Proceedings of Fellbaum*, pages 305–332, 1998.

[HSD73] Robert M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. In *Systems, Man and Cybernetics, IEEE Transactions*, volume SMC-3, pages 610–621, 1973.

[HSJ98] S. Hui, C. W. Slemenda, and C. C. Johnston. Age and bone mass as predictors of fracture in a prospective study. *Clin Invest 81*, pages 1804–1809, 1998.

[HTF01] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning : Data mining, inference and prediction. *Springer*, 2001.

[Hug68] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1) :55–63, 1968.

[HXW17] M. Hu, F. Xiao, and L. Wang. Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model. *Energy Procedia*, 105 :2759–2765, 2017.

[Jas16] Brownlee Jason. *Deep Learning With Python*. Machine Learning Mastery, 2016.

[JDM00] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition : A review. *IEEE Transactions pattern analysis and machine intelligence*, 22 :4–37, 2000.

[Jen95] R. Jennane. *Modélisation fractale de textures, application à l'analyse de l'architecture osseuse*. Thèse de doctorat, Université d'Orléans, France, 1995.

[JGL+14] H. V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, and Cyrus Shahabi. Big data and its technical challenges. *Communication of ACM*, 57(7) :86–94, 2014.

[JS02] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem : A systematic study. *Intelligent Data Analalysis*, 6(5) :429–449, 2002.

[JZS15] R Jozefowicz, W Zaremba, and I Sutskever. An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, 2015.

[KGD13] K. A. Kumar, J. Gluck, and A. Deshpande. Hone : 'scaling down' hadoop on shared-memory systems. In *Proceedings of the VLDB Endowment*, volume vol. 6, no. 12, pages 1354–1357, 2013.

[KGPP14] Camille Kurtz, Pierre Gancarski, Anne Puissant, and Nicolas Passat. Approches multi-hiérarchiques pour l'analyse d'images de télédétection. *Revue Française de Photogrammétrie et de Télédétection*, pages 19–35, 01 2014.

[KH08] Y. Kompatsiaris and P. Hobson, editors. *Semantic Multimedia and Ontologies*. Springer-Verlag, London, 2008.

[KK06] S. Kotsiantis and D. Kanellopoulos. Association rules mining : A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1) :71–82, 2006.

[KL98] J. H. Kinney and A. J. C. Ladd. The relationship between three-dimensional connectivity and the elastic properties of trabecular bone. *Journal of Bone and mineral research*, 13(5) :839–845, 1998.

[KLSS17] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sensing Lett*, 14(5) :778–782, 2017.

[KPD+11] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. Berg, and T. Berg. Baby talk : Understanding and generating simple image descriptions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1601–1608, 2011.

[KS04] Y. Ke and R. Sukthankar. Pca-sift : A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004, in Proceedings of the 2004 IEEE Computer Society Conference*, volume 2, 2004.

[KUG14] M. Khan, M. F. Uddin, and N. Gupta. Seven v's of big data understanding big data to extract value. In *Proceedings of the 2014 Zone 1 Conference of the American Society for engineering Education*, pages 1–5, 2014.

[LAJ15] Dana Lahat, Tülay Adali, and Christian Jutten. Multimodal data fusion : an overview of methods, challenges, and prospects. In *Proceedings of the IEEE*, volume 103.9, pages 1449–1477, 2015.

[Lan06] T. A. Landaue. Latent semantic analysis. *Wiley Online Library*, 2006.

[LBAM20] Katerina Lepenioti, Alexandros Bousdekis, Dimitris Apostolou, and Gregoris Mentzas. Prescriptive analytics : Literature review and research challenges. *International Journal of Information Management*, 50 :57–70, 2020.

[LC98] C. Leacock and M. Chodorow. Combining local context and wordnet similarity for word sense identification. *WordNet : An Electronic Lexical Database, MIT Press*, pages 265–283, 1998.

[LHG+98] C. M. Langton, T. J. Haire, P. S. Ganney, C. A. Dobson, and M. J. Fagan. Dynamic stochastic simulation of cancellous bone resorption. *Bone Journal*, 22(4) :375–380, 1998.

[LKLC03] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Yuan-chi Chiu. A symbolic representation of time series with implications for streaming algorithms. In Mohammed Javeed Zaki and Charu C. Aggarwal, editors, *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, DMKD 2003, San Diego, California, USA, June 13, 2003*, pages 2–11. ACM, 2003.

[LMSR17] G. Lin, A. Milan, C. Shen, and I. Reid. Refinenet : Multi-path refinement networks for high-resolution semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[LSD15] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional net- works for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[LW04] L. Ledwich and S. Williams. Reduced sift features for image retrieval and indoor localisation. In *Australian Conference on Robotics and Automation*, 2004.

[LZLM07] Y. Liu, D. Zhang, G. Lu, and W-Y. Ma. Survey of content-based imageretrieval with high-level semantics. In *Pattern Recognition*, volume 40, pages 262–282, 2007.

[LZZ+17] X. Lian, C. Zhang, H. Zhang, C-J. Hsieh, W. Zhang, , and J. Liu†. Can decentralized algorithms outperform centralized algorithms ? a case study for decentralized parallel stochastic gradient descent. In *31st Conference on Neural Information Processing Systems (NIPS 2017).*, pages 1–11, 2017.

[Mac67] J.B. Macqueen. Some methods of classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symp. on MSP*, pages 281–297, 1967.

[Mah20] A. Mahdjouba. *Impact énergétique de l'effacement dans un entrepôt frigorifique - Analyse des approches systémiques : boîte noire / boîte blanche.* Thèse de doctorat de génie des procédés, INRAE, Anthony, France, 2020.

[Mal89] S.G. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell*, 11 :674–693, 1989.

[Mar10] David Marr. *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information.* The MIT Press, 2010.

[Mel19a] *A Big Remote Sensing Data Analysis Using Deep Learning Framework*, July 2019.

[Mel19b] *Deep Semantic Feature Detection from Multispectral Satellite Images*, September 2019.

[Mil14] Thomas W. Miller. *Modeling Techniques in Predictive Analytics : Business Problems and Solutions with R.* PH Professional Business, 2014 by pearson education, inc. upper saddle river, new jersey 07458 edition, August 2014.

[MM97] R. W. McCalden and J. A. McGeough. Age-related changes in the compressive strength of cancellous bone : The relative importance of changes in density and trabecular architecture. *J. Bone Joint Surg. 79*, 191 :421–427, 1997.

[MNW+19] Azlinah Mohamed, Maryam Khanian Najafabadi, Yap Bee Wah, Ezzatul Akmal Kamaru Zaman, and Ruhaila Maskat. The state of the art and taxonomy of big data analytics : view from new big data framework. *Artificial Intelligence Review*, Feb 2019.

[MWLF05] Vasileios Megalooikonomou, Qiang Wang, Guo Li, and Christos Faloutsos. A multiresolution symbolic representation of time series. In *ICDE*, pages 668–679, 2005.

[Nav09] R. Navigli. ACM computing surveys. *Artificial Intelligence*, 41 :1–69, 2009.

[NB14] R. Narasimhan and T. Bhuvaneshwari. Big data - a brief study. *International Journal of Scientific & Engineering Research*, 5(9) :350–353, 2014.

[NHP13] T. Nguyen, J. Han, and P-C. Park. Satellite image classification using convolutional learning. In *AIP conference Proceedings*, pages 2237–2240, 2013.

[NMMB98] C. Nastar, M. Mitschke, C. Meilhac, and N. Boujemaa. Reduced sift features for image retrieval and indoor localisation. In *Proceedings of the sixth ACM international conference on Multimedia*, pages 339–344, 1998.

[NS12] R. Navigli and P. Simone. Babelnet : The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193 :217–250, 2012.

[NVK+15] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R.Wald, and E. Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1) :87–93, 2015.

[OT07] Aude Oliva and Antonio Torralba. The role of context in object recognition. *TRENDS in Cognitive Sciences*, 11(12) :520–527, 2007.

[PACB10] F. Peyrin, D. Attali, C. Chappard, and CL. Benhamou. Local plate-rod descriptors of 3d trabecular bone micro-ct images from medial axis topologic analysis. *Med. Physics*, 3 :4364–4376, 2010.

[Par12] C. Parker. Unexpected challenges in large scale machine learning. In *Proceedings of the* 1st *International Workshop on Big Data, Streams and Heterogeneous Source Mining Algorithms, Systems, Programming Models and Applications*, pages 1–6, 2012.

[Pat19] A. Patil. Distributed programming frameworks in cloud platforms. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(6) :611–619, 2019.

[PCD15] P.O. Pinheiro, R. Collobert, and P. Dollár. Learning to segment object candidates. *In Advances in Neural Information Processing Systems*, pages 1990–1998, 2015.

[PCT+15] B. Perret, J. Cousty, O. Tankyevych, H. Talbot, and N. Passat. Directed connected operators : Asymmetric hierarchies for image filtering and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6) :1162–1176, 2015.

[PHB+01] F. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarre, and A. Bonami. nthorder fractional brownian motion and fractional gaussian noises. In *IEEE Transactions on Signal Processing*, volume 45, pages 1049–1059, 2001.

[PHMD18] D. J. Power, C. Heavin, J. McDermott, and M. Daly. Defining business analytics : an empirical approach. *Journal of Business Analytics*, 1(1) :40–53, 2018.

[PLH+98] L. Pothuaud, E. Lespessailles, R. Harba, R. Jennane, V. Royant, E. Eynard, and C-L. Benhamou. Fractal analysis of trabecular bone texture on radiographs : discriminant value in postmenopausal osteoporosis. In *Steoporosis International*, volume 8, pages 618–625, 1998.

[PLXY11] Weike Pan, Nathan Nan Liu, Evan Wei Xiang, and Qiang Yang. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the* $22^n d$ *International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 2318–2323. IJCAI/AAAI, 2011.

[PNtS+10] Sinno Jialin Pan, Xiaochuan Ni, Jian tao Sun, Qiang Yang, and Zheng Chen. Cross-domain sentiment classification via spectral feature alignement. In *WWW*, pages 751–760, 01 2010.

[PY10] Sinno Jialin Pan and Qiang Yang. A survey on transfert learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10) :1345–1359, 2010.

[QWD+16] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng. A survey of machine learning for big data processing. In *EURASIP Journal on Advances in Signal Processing*, volume 67, pages 1–16, 2016.

[Rec89] R. R. Recker. Low bone mass may not be the only cause of skeletal fragility in osteoporosis. *Proc. Soc. Exp. Biol. Med.*, 191 :272–274, 1989.

[Res95] Ph. Resnik. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.

[RJY15] J. Ren, X. Jiang, and J. Yuan. A chi-squared- transformed subspace of lbp histogram for visual recognition. *IEEE Transactions on Image Processing*, 24(6) :1893–1904, 2015.

[RM16] Anne Ricordeau and Nédra Mellouli. Exploring 3d-structure analysis tools for a simulated bone remodelling process. *CMBBE : Imaging & Visualization*, 4(3-4) :253–263, 2016".

[Ros79] Azriel Rosenfeld. Digital topology. *American Mathematical Monthly*, 86, 10 1979.

[RRFD16] X. Renard, M. Rifqi, G. Fricout, and M. Detyniecki. EAST representation : fast discovery of discriminant temporal patterns from time series. In *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*, Riva Del Garda, Italy, Sep 2016.

[RVHH05] R. Ruimerman, B. Van Rietbergen, P. Hilbers, and R. Huiskes. The effects of trabecular-bone loading variables on the surface signaling potential for bone remodeling and adaption. *Annals of Biomed. Eng.*, 33(1) :71–78, 2005.

[SBCS01] S. Sevestre-Ghalila, A. Benazza-Benyahia, H. Cherif, and W. Souid. Texture analysis for osteoporosis detection with morphological tools. In *Medical Imaging. SPIE Conf.*, volume 4322, pages 1534–1541, 2001.

[SBR+04] S. Sevestre-Ghalila, A. Benazza-Benyahia, A. Ricordeau, N. Mellouli, and C-L. Benhamou C. Chappard. Texture image analysis for osteoporosis detection with morphological tools. In *Proceedings of the MICCAI'04, St-Malo, France*, pages 26–30, 2004.

[Sch10] E. Schmidt. Every 2 days we create as much information as we did up to 2003, Aug. 2010.

[SDB+91] V. Shen, D. W. Dempster, R. Birchman, R. Xu, and R. Lindsay. Loss of cancellous bone mass and connectivity in ovariectomized rats can be restored by combined treatment with prathyroid hormone and estradiol. *Journal Clin Invest.*, pages 2479–2487, 1991.

[Ser03] J. Serra. Connexions et segmentation d'image. *Traitement du Signal*, 20(3) :243–254, 2003.

[Shm10] Galit Shmuel. To explain or to predict ? *Statistical Science*, 25 :289–310, 2010.

[SLJ+15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, pages 1–9, 2015.

[SNIU19] G. Swarnendu, D. Nibaran, D. Ishita, and M. Ujjwal. Understanding deep learning techniques for image segmentationt. *arXiv preprint arXiv :1907.06119v1*, 2019.

[Soi03] P. Soille. Morphological image analysis : principles and applications. In *Springer*, 2003.

[SR15] D. Singh and C. K. Reddy. A survey on platforms for big data analytics. *Journal of big data*, 2(1) :1–20, 2015.

[SSS+15] B. Saha, H. Shah, S. Seth, G. Vijayaraghavan M. Murthy, and C. Curino. A unifying framework for modeling and building data processing applications. In *2015 ACM SIGMOD International Conference on Management of Data (SIGMOD'15)*, pages 1357–1369, 2015.

[Suk14] S. R. Sukumar. Machine learning in the big data era : Are we there yet ? In *in Proceedings of the* 20th *ACM SIGKDD Conference on Knowledge Discovery and Data Mining : Workshop on Data Science for Social Good*, 2014.

[SYH+10] PK. Saha, Y.Xu, H.Duan, A.Heiner, and G.Liang. Volumetric topological analysis : A novel approach for trabecular bone classification on the continuum between plates and rods. *IEEE Trans. Med. Imaging*, 29(11) :1821–1838, 2010.

[SZ14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv :1409.1556*, 2014.

[SZ15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR 2015*, 2015.

[Tab10] Z. Tabor. Anisotropic resolution biases estimation of fabric from 3d gray-level images. *Journal of Medical Engineering & Physics*, 32 :39–48, 2010.

[TG80] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1) :97–136, 1980.

[TKC05] I. W. Tsang, J. T. Kwok, and P-M. Cheung. Core vector machines : Fast svm training on very large data sets. *Journal of Machine Learning Research*, 6 :363–392, 2005.

[TLD18] Lynda Tamine-Lechani and Mariam Daoud. Evaluation in contextual information retrieval : Foundations and recent advances within the challenges of context dynamicity and data privacy. *ACM Computing Surveys - CSUR*, 51(4) :1–36, September 2018.

[Tor03] Antonio Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2) :169–191, 2003.

[Tsy04] Alexey Tsymbal. The problem of concept drift : Definitions and related work. Technical report, Computer Science Department, Trinity College Dublin, 2004.

[Tve77] A Tversky. Features of similarity. *Psycological Review*, pages 327–352, 1977.

[Vap95] C. Vapnick. Support-vector network. *Machine Learning*, 20(3) :273–297, 1995.

[Vin11] P. Vinukonda. *A Study of the Scale-Invariant Feature Transform on a Parallel Pipeline*. PhD thesis, Louisiana State University, 2011.

[VJS+14] D. Vannella, D. Jurgens, D. Scarfini, D. Toscani, and R. Navigli. Validating and extending semantic knowledge bases using video games with a purpose. *ACL*, 1 :1294–1304, 2014.

[VMD+13] V.K. Vavilapalli, A.C. Murthy, C. Douglas, S. Agarwal, M. Konar R. Evans, T. Graves, J. Lowe, H. Shah, and S. Seth. Apache hadoop yarn : Yet another resource negotiator. In *Proceedings of the 4th annual Symposium on Cloud Computing*, 2013.

[WCP+14] J. Wang, D. Crawl, S. Purawat, M. Nguyen, and I. Altintas. Big data provenance : Challenges, state of the art and opportunities. In *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*, pages 2509–2516, 2014.

[WFHP16a] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining : Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series, 2016.

[WFHP16b] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining : Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[Whi74] WJ. Whitehouse. The quantitative morphology of anisotropic trabecular bone. *Journal of Microscopy*, 101 :1. :53–68, 1974.

[WP94] Z. Wu and M. Palmer. Verb semantics and lexical selection. In *32nd annual Meeting of the Association for Computational Linguistics*, pages 27–30, 1994.

[WZD14] X. Wu, X. Zhu, and WuW. Ding. Data mining with big data. *IEEE transactions on Knowledge and data engineering*, 26(1) :97–107, 2014.

[XWYC14] X. Xue, S. Wang, C. Yan, and B. Cui. A fast chiller power demand response control strategy for buildings connected to smart grid. *Applied Energy*, 137 :77–87, 2014.

[YRMP18] WENLU YANG, Maria Rifqi, Christophe Marsala, and Andrea Pinna. Physiological-Based Emotion Detection and Recognition in a Video Game Context. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 194–201, Rio, Brazil, July 2018.

[YTF73] S. Yokoi, I. Toriwaki, and T. Fukumura. Topological properties in digitized binary pictures. *Syst. Comput. Contr.*, 4(6) :32–39, 1973.

[ZCF+10] M. Zaharia, M. Chowdhury, M.J. Franklin, S. Shenker, and I. Stoica. Spark : cluster computing with working sets. In *The 2nd USENIX conference on Hot topics in cloud computing*, 2010.

[Zhe15] Y. Zheng. Methodologies for cross-domain data fusion : an overview. *IEEE Transactions on Big Data*, 1(1) :16–34, 2015.

[ZTAM20] J. Zarka, L. Thiry, T. Angles, and S. Mallat. Deep network classification by scattering and homotopy dictionary learning. In *ICLR 2020*, 2020.