

Laboratoire d'informatique de Paris Nord

LIPN

Rapport d'activité 2017-2022
et projet 2023-2028

Équipe A3 :

Apprentissage Artificiel
et
Applications

cnrs

UNIVERSITÉ
SORBONNE PARIS NORD

Partie I

Le LIPN

page 5

The LIPN at a glance

Environment
Scientific activities
Doctoral studies

7

Partie II

Equipe A3

page 13

Informations générales sur l'équipe A3

Membres de l'équipe A3
Profil d'activités liées à la recherche
Les thématiques scientifiques et leurs enjeux
Prise en compte des recommandations du précédent rapport

15

Introduction du portfolio

29

Auto-évaluation de l'équipe A3

Domaine 2. Attractivité
Domaine 3. Production scientifique
Domaine 4. Inscription des activités de recherche dans la société

31

Trajectoire de l'équipe A3

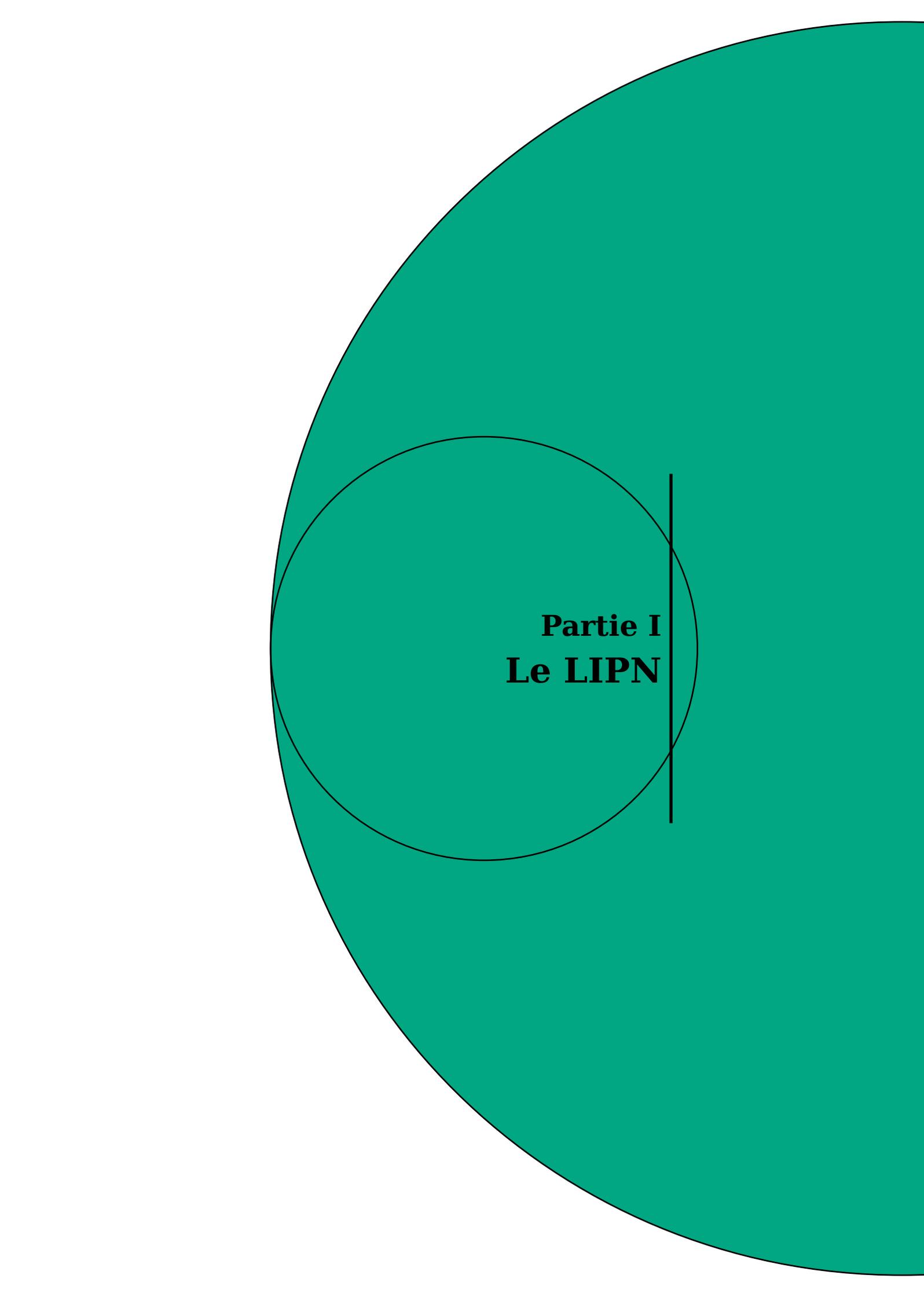
Analyse AFOM/SWOT
Organisation et ambitions de l'équipe
Perspectives et projet de recherche

53

Partie III

Acronymes

page 63

The image features a teal background. A white circle is positioned in the center-left area. A vertical white line is located to the right of the circle, extending from the top to the bottom of the circle's vertical span. The text 'Partie I' and 'Le LIPN' is centered within the white circle.

Partie I
Le LIPN

Chapitre 1

The LIPN at a glance

The LIPN (Laboratoire d'Informatique de Paris-Nord) was created in 1985, has been associated with CNRS since January 1992, before becoming a CNRS UMR in January 2001. The LIPN is the computer science laboratory of University Sorbonne Paris Nord (USPN) which is part of the Alliance Sorbonne-Paris-Cité ASPC.

The laboratory comprises 5 teams with more than 70 permanent staff (full-time researchers, professors and associate professors or research engineers), 9 technical and administrative staff, as well as PhD students and postdocs. It has overall more than 150 members most of the time. After having experienced impressive growth during the last two five-year contracts, the number of teacher-researchers at LIPN has decreased, weakening some of its areas of expertise.

The scientific activities display many collaborations, be they national, international, or through research projects, in particular with the industry. LIPN members are extremely involved on the national scene, in national boards (at CNRS, CNU — national universities council, HCERES, etc.), and in institutional cooperation projects (ANR). On the international scene, members are widely present in editorial boards, programme committees, expert pools of different countries, international schools, and conduct collaborative research with colleagues worldwide.

1.1 Scientific activities

The laboratory is structured into 5 teams and a transverse axis, focussing on the following main areas.

- **A³ (Apprentissage Artificiel et Applications): Machine Learning and Applications**

The A³ team tackles machine learning problems and covers a wide spectrum of issues, ranging from supervised and unsupervised learning to reinforcement learning. Its research is fed, coordinated and evaluated thanks to various applications in the field of pattern recognition and data mining. Research in A³ focuses on the following main topics: algebraic and logical models of learning, collaborative and transfer learning, cluster analysis, dimensionality reduction, link prediction in social networks, recommender systems, and learning structures from complex data.

- **AOC (Algorithmes et Optimisation Combinatoire): Algorithms and Combinatorial Optimisation**

The AOC team develops research in optimisation on graphs, mathematical programming, parallel and distributed computing. Optimisation on graphs is conducted with a particular focus on complexity, polyedral theory and approximation. The team has expertise in design and analysis of mathematical programming approaches and algorithms, and develops both exact and heuristic approaches for solving linear and non-linear problems. Many distributed environments issues are also considered: distributed middleware and architectures, distributed programming and distributed algorithms.

- **CALIN (Combinatoire, ALgorithmique et INTERactions): Combinatorics, ALgorithmics, and INTERactions**

The CALIN team brings together researchers with skills in a variety of aspects of combinatorics (analytic, bijective, geometric, and algebraic). They are interested in the complexity of algorithms and in the fine-grained analysis of data structures. They also study problems of physics with a distinct combinatorial flavour, or apply methods from physics to combinatorial problems. The team is organised into two intersecting subgroups that reflect these interests: one of them focuses on the analysis of algorithms and combinatorial structures, and the other is devoted to the interactions between combinatorics and physics. Geometry, both as a tool and a subject of study is transversal to these two subgroups.

- **LoVe (Logique et Vérification : Logic and Verification)**

The LoVe team tackles different aspects of computational and software models, from theory to applications. It is organised into two subgroups. The logic, programming and complexity group studies the relationship between logic and programming languages, especially for what concerns denotational semantics of programs and proofs, under the Curry-Howard correspondence. It has a special expertise in the quantitative aspects of programming, including the links with computational complexity. It also hosts experts in formal proofs. The verification group develops model checking approaches for concurrent systems, in particular in a parametric, distributed and timed setting. The research spans from fundamental aspects to application domains, including an open verification platform software development.

- **RCLN (Représentation des Connaissances et Langage Naturel): Knowledge Representation and Natural Language**

The RCLN team carries out innovative work for the analysis and exploration of text corpora as well as for the acquisition of knowledge from texts or knowledge graphs. The team is particularly active in problems that are little studied in language, such as the treatment of under-endowed or specialised languages, and is working on approaches to deal with the complexity of problems in texts or knowledge graphs, such as syntactic analysis, information extraction, or graph completion using methods from deep learning, combinatorial optimisation, data mining or inductive logic programming.

- **The transverse axis Complexités: Complexities**

The *Complexités* research axis spreads across the whole laboratory and brings together all those people whose work is related, directly or indirectly, with computational complexity. It is primarily composed of members of the LoVe, CALIN and AOC teams, but its themes potentially intersect also with the other two teams of LIPN. The activities of the axis are centered around a weekly seminar.

1.2 Environment

The LIPN has many research cooperations with its environment: other laboratories at University Sorbonne Paris Nord, but also in the Alliance Sorbonne-Paris-Cité (ASPC), with industry, etc.

- **MathSTIC research federation, and University Sorbonne Paris laboratories**

Research collaborations with other laboratories of University Sorbonne Paris Nord have been favoured over the years, in particular with LAGA (UMR 7539, mathematics laboratory), L2TI (UR3043, signal and image processing laboratory).

This led to the creation of a new research federation in Mathematics and Information Technology MathSTIC, which became a CNRS Research Federation – FR 3734 – in January 2016), gathering members of LAGA, LIPN, and L2TI laboratories, in order to enhance cross-fertilisation in four of their major areas of expertise:

- Optimisation and learning applied to digital contents;
- High-performance computing, distributed systems;
- Mathematical physics, statistical physics, combinatorics;
- Categories: between calculation and topology.

- **Regional setting: DIM RFSI and EFL LabEx**

All LIPN teams strongly collaborate with other groups in the Île-de-France region. Furthermore, LIPN has been heading the regional initiative in computer science, DIM RFSI, which fostered regional collaborative projects between different laboratories and companies in computer science or application domains. It also allowed for setting up a large regional cluster infrastructure for HPC and deep learning.

The team RCLN at LIPN is a major actor in the EFL LabEx (Empirical Foundations of Linguistics). Its multidisciplinary nature is a key to evolution of research topics at the meeting point of linguistics and computer science.

- **International cooperations**

Members of the laboratory have many collaborations worldwide, with exchanges of researchers. Moreover, approximately 15 foreign researchers are invited for a period of two to four weeks at our laboratory each year to carry collaborative research. Co-operation projects involve laboratories in Denmark, Netherlands, Norway, Poland, Italy,

1.3 Doctoral studies

Tunisia, Marocco, Canada, United States, Argentina, Brazil, Chile, Mexico, Uruguay, Singapore, Taiwan, Vietnam, Australia.

• Industrial collaborations

Most research areas of the LIPN also participate in collaborations with the industry, mainly through research projects (ANR, FUI, industrial collaborations) and CIFRE PhDs. The LIPN is involved in a joint laboratory (LabCom) on satellite image processing. Furthermore some researchers of the LIPN have also created a start-up company, or benefit from a few months industrial experience. Moreover, the LIPN is involved in different boards of 2 business clusters (pôles de compétitivité) which favour technology transfer and research projects with companies at a regional level.

1.3 Doctoral studies

The doctoral studies are organised within the doctoral school ED146 *Galileo : Science, Technology, Health*, which addresses three major scientific areas, LIPN being part of the first group:

1. mathematics, computer science, signal processing;
2. physics, materials, engineering sciences;
3. health sciences, medicine, human biology, chemistry, ethology.

The doctoral school organises mandatory courses, and allocates scholarships to doctoral students. Every year, approximately five of them are allocated to the LIPN (out of 25 for the whole doctoral school).

• PhDs and Habilitations

Over the 2017–2022 period, 62 PhD defences took place. Half of the doctoral students were financed by a scholarship from the doctoral school, and the others through research projects, industry, or international support.

During the same period, 11 habilitations (Habilitations à diriger les Recherches) were delivered.

• Math & Computer Science Graduate School

The M&CS Graduate school is an international 5-year high-level curriculum founded for 8 years. It is endorsed by the Galileo Doctoral School and three local masters: Mathematics, Computer Science, Engineering and Innovation in Images and Networks. The LIPN and the two other laboratories of MathSTIC research federation actively participate in this graduate school. The interactions between Mathematics and Computer Science are highlighted in the graduate school to promote interdisciplinarity in teaching and research.

• Masters courses

A masters programme in computer science is delivered at University Sorbonne Paris Nord. It currently features two specialities:

- EID² (Exploration Informatique des Données et Décisionnel): Machine Learning and Data Science and its international curriculum Data Science;
- PLS (Programmation et Logiciels Sûrs): Programming and Secure Software.

Both specialities lead to either industrial or research careers, depending on the master thesis.

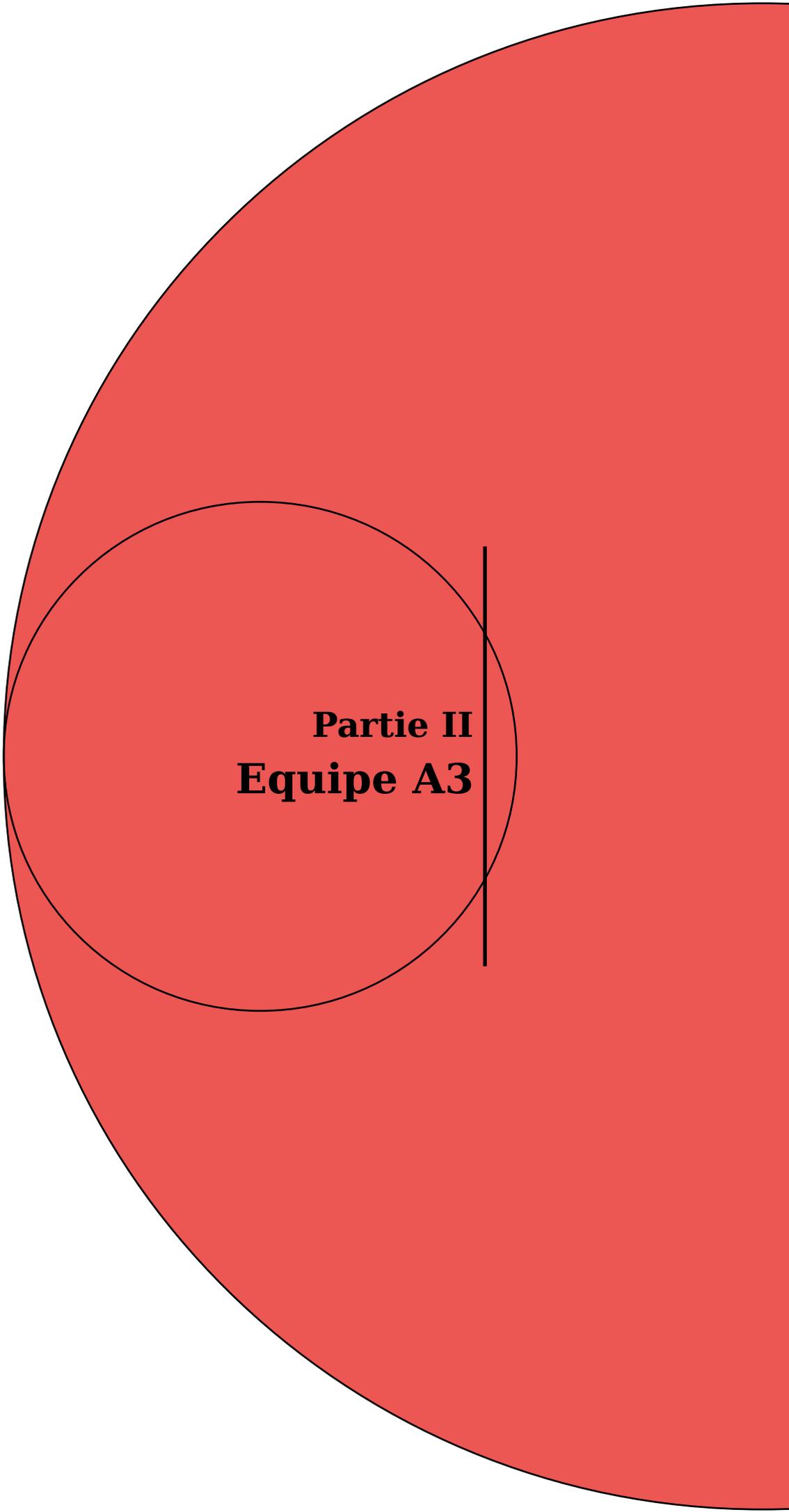
LIPN researchers also deliver courses in other masters programmes of universities in the parisian region.

Furthermore, they are involved in international cooperations for setting up masters programmes in foreign countries, e.g. Madagascar, Vietnam.

- **Summer schools and conferences**

LIPN supports the organisation and participation in thematic schools, by encouraging doctoral students attendance (which is recognised as mandatory courses), talks by members of the laboratory, as well as financial support to the organisation of such events.

In particular, Column Generation School 2018, Mathématiques expérimentales: méthodes et pratiques (MathExp) 2018, Kaleidoscope research school 2019, Deep Learning & Data science thematic school every two years, were co-organised by LIPN members. Some of these thematic schools take place in University Sorbonne Paris Nord, while others are elsewhere in France or abroad.



Partie II
Equipe A3

Overview

Machine learning is a subfield of artificial intelligence that involves the design and development of algorithms and models, allowing computers to learn from observations and prior knowledge, without being explicitly programmed, to improve their performance in the execution of tasks (predictions, decision-making, etc.). Machine learning borrows techniques and theoretical backgrounds from several fields, including mathematics (Linear algebra, calculus, probability theory, geometry, optimization, information theory, etc.), statistics providing the framework for evaluating the accuracy and reliability of machine learning models, computer science, and cognitive psychology.

The latest advances in machine learning, particularly those initiated by transformers and more recently ChatGPT, are profoundly changing all areas of society. They are transforming relationships with work, education, research, health, entertainment, control systems, people, and access to personal data, among others.

In the context of classical machine learning and metalearning, our group covers a broad range of topics, from symbolic and statistical Learning to graph mining and reinforcement learning. This allows us to be quite reactive to cope with new challenges raised by emerging applications of Machine Learning. It also makes it possible to study innovative combinations of learning methods for tackling complex problems.

The structure of the group has evolved over the years, and now consists of three research axes :

- *Learning from Data and Learners*, which gathers research that deals with collaboration and transfer learning;
- *Relational Learning and Graphs*, motivated by learning from models expressed as logical, explicit and explainable programs and from data structured in graphs;
- *Meta-learning and Structures Learning*, which generalizes the learning problem by treating the learning process itself. It also allows simultaneous modeling of data and clustering (deep unsupervised learning).

In each of these axes, theoretical issues are addressed, and algorithms and software are developed, in the context of collaborative projects, including both academic and industrial partners. The team publishes in main Machine Learning international conferences (IJCNN, ICML, ECML, AAAI, IJCAI to name a few) as well as main international journals (Artificial Intelligence, Machine Learning, Pattern Recognition, ...).

The strong links of the group with industrial partners is in particular testified by the high number of PhD in the group that take place in an industrial context.

The research topics of the team for the next five years aim at investigating promising areas where the team may emerge as a leader :

- Using emerging mathematical formalisms: algebraic topology, optimal transport
- Learning to learn including bias learning, decision boundaries, weak convergence and selection of admissible functions.
- Interpretability and symbolic machine learning
- Mining augmented and heterogeneous graphs
- New challenges for simultaneous modeling of data and clustering (deep unsupervised learning) using multi-varied temporal data, large-scale learning, images and audio or video documents.

The group actively participates in national projects, and has several international collaborations. The researchers of the team are involved in teaching at the undergraduate and graduate levels, in USPN and in other institutions. In particular, the group has created a highly visible second year master specialization on Data Mining, Analytics, and Knowledge Discovery (*Exploration Informatique des Données et Décisionnel* (EID²)) of the USPN master in computer science.

Chapitre 1

Informations générales sur l'équipe A3

Sommaire

| | |
|---|-----------|
| 1.1 Membres de l'équipe A3 | 16 |
| 1.1.1 Membres permanents | 16 |
| 1.1.2 Membres non-permanents au 01/01/2023 | 17 |
| 1.1.3 Autres personnels temporaires du 01/01/2017 au 31/12/2022 | 17 |
| 1.1.4 Évolution de l'équipe (permanents) | 18 |
| 1.2 Les thématiques scientifiques et leurs enjeux | 18 |
| 1.2.1 Axe ADA : Apprentissage à partir de Données et d'Apprenants | 18 |
| 1.2.2 Axe ARG : Apprentissage Relationnel et Graphes | 21 |
| 1.2.3 Axe MAR : Métaapprentissage et Apprentissage de Structures | 23 |
| 1.3 Profil d'activités liées à la recherche | 26 |
| 1.4 Prise en compte des recommandations du précédent rapport | 26 |

Nom de l'équipe : Apprentissage Artificiel et Applications

Responsable de l'équipe : Aomar OSMANI

1.1 Membres de l'équipe A3

1.1 Membres de l'équipe A3

Tableau synthétique de l'évolution des membres de l'équipe (nombre de membres au 1er janvier de chaque année) :

| Corps | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | (2023) |
|-------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|
| DR | 0 | 0 | 0 | 0 | 0 | 0 | (0) |
| CR | 0 | 0 | 0 | 0 | 0 | 0 | (0) |
| PU | 3 | 2 | 2 | 2 | 2 | 2 | (2) |
| MCF | 18 | 16 | 16 | 15 | 14 | 12 | (12) |
| <i>Total permanents</i> | <i>21</i> | <i>18</i> | <i>18</i> | <i>17</i> | <i>16</i> | <i>14</i> | <i>(14)</i> |
| Post-docs et ATER | 4 | 5 | 3 | 2 | 1 | 1 | (1) |
| Ingénieurs (CDD) | 0 | 0 | 0 | 0 | 0 | 1 | (0) |
| Doctorants | 11 | 14 | 15 | 14 | 11 | 14 | (9) |
| <i>Total</i> | <i>36</i> | <i>37</i> | <i>36</i> | <i>33</i> | <i>28</i> | <i>30</i> | <i>(24)</i> |

Informations
générales sur
l'équipe A3

1.1.1 Membres permanents

(membres ou, *en italique*, anciens membres, et, **en gras**, responsable de l'équipe au 01/01/2023)

| | <i>Nom</i> | <i>Prénom</i> | <i>Situation</i> | <i>Institution</i> |
|---|-------------------|-----------------|------------------|--------------------|
|  | Azzag | Hanane | HDR | IUTV |
|  | Bennani | Younès | PU | IG |
|  | Bouthinon | Dominique | MCF | IUTV |
|  | Cabanes | Guénaël | MCF | IG |
|  | Champesme | Marc | MCF | IG |
|  | <i>Chevaleyre</i> | <i>Yann</i> | PU | IG |
|  | Gérard | Pierre | MCF | IUTV |
|  | <i>Grozavu</i> | <i>Nistor</i> | MCF | IG |
|  | Guérif | Sébastien | MCF | IG |
|  | Kanawati | Rashed | MCF | IUTV |
|  | <i>Lebbah</i> | <i>Mustapha</i> | MCF | IUTV |
|  | Matei | Basarab | HDR | IG |
|  | Osmani | Aomar | HDR | IUTV |
|  | Rouveirol | Céline | PU | IG |

1.1 Membres de l'équipe A3



Santini Guillaume MCF IUTV



Soldano Henry MCF IG

1.1.2 Membres non-permanents au 01/01/2023

| | Nom | Prénom | Situation | Financement |
|--|---------------|------------|-----------|----------------------|
| | Akouba | Meliha | doctorant | CD USPN |
| | Amekeo | Kodjo M. | doctorant | CIFRE Groupe BPCE |
| | Andriantsiory | Dina F. | doctorant | CD USPN |
| | Barbara | Abir | doctorant | DeepKnowledge |
| | Ben Salem | Nosseiba | doctorant | EUR M&CS |
| | Faye | Bilal | doctorant | CD USPN |
| | Khoufeche | Reda M. | doctorant | CIFRE DIM AI4DF |
| | Kazi Aoual | Malik | doctorant | CIFRE Nukkai |
| | Khalafaoui | Yasser | doctorant | CIFRE Alteca |
| | Leclercq | Etienne | doctorant | CIFRE Alterway-Smile |
| | Madane | Abdellah | doctorant | CIFRE SAFRAN |
| | Modaresi | Seyed M.R. | doctorant | Ambassade de France |
| | Zangato | Théo | doctorant | Ministère des Armées |

Informations
générales sur
l'équipe A3

1.1.3 Autres personnels temporaires du 01/01/2017 au 31/12/2022

| Nom | Prénom | Situation | Début | Fin | Financement |
|-------------|-----------------|-----------|------------|------------|------------------------|
| Attaoui | Mohammed Oualid | doctorant | 01/09/2018 | 30/06/2021 | Campus France |
| Beck | Gaël | doctorant | 01/03/2016 | 14/10/2019 | CIFREKameleoon |
| Ben Bouazza | Fatima-Ezzahraa | doctorant | 01/09/2017 | 26/11/2020 | Bourse Maroc |
| Benlamine | Kaoutar | doctorant | 01/09/2017 | 3/11/2020 | CD USPN |
| Benyettou | Assia | doctorant | 15/10/2014 | 02/11/2017 | Bourse Algérie |
| Brahmi | Ibtissam | Postdoc | 01/09/2016 | 30/11/2019 | Bourse CNRST |
| Chahdi | Hatim | doctorant | 01/09/2014 | 04/07/2017 | ANR Coclico |
| Chelly | Zaineb | Postdoc | 01/03/2016 | 28/02/2018 | MARIE SKLODOWSKA-CURIE |
| Chevallier | Marc | doctorant | 15/07/2019 | 23/11/2022 | CIFRE Synaltic |
| Coutant | Anthony | Postdoc | 15/01/2016 | 30/06/2018 | ChistEra Adalab |

1.2 Les thématiques scientifiques et leurs enjeux

| | | | | | |
|-----------|------------|------------------|------------|------------|----------------------|
| Dilmi | Djallel | Postdoc | 01/02/2022 | 02/06/2023 | Amies |
| El Hamri | Mourad | doctorant | 01/09/2019 | 15/12/2022 | CD USPN |
| El Hamri | Mourad | ATER | 01/09/2022 | 30/08/2023 | USPN |
| Falih | Issam | doctorant | 01/10/2015 | 08/03/2018 | CD USPN |
| Forest | Florent | doctorant | 01/12/2017 | 22/03/2021 | CIFRE SAFRAN |
| Foucade | Yohan | doctorant | 01/03/2019 | 10/06/2022 | Contrat Facebook |
| Goffinet | Etienne | doctorant | 01/12/2018 | 15/12/2021 | CIFRE Renault |
| Hamidi | Massinissa | doctorant | 01/10/2017 | 20/12/2022 | CD USPN |
| Khoufeche | Reda M. | Ing.d'étude CNRS | 01/01/2022 | 31/12/2022 | France Relance ACOMI |
| Léger | Hippolyte | doctorant | 01/09/2015 | 14/02/2019 | CD USPN |
| Rastin | Parisa | doctorant | 01/09/2015 | 26/06/2018 | CIFRE Mindlytix |
| Rifi | Mouna | doctorant | 01/02/2016 | 03/05/2019 | CIFRE EDF |
| Sarazin | Tugdual | doctorant | 04/02/2013 | 26/06/2018 | CIFRE Smokewatchers |
| Veillon | Anne-Lise | doctorant | 01/10/2014 | 08/03/2018 | CD USPN |
| Veillon | Anne-Lise | Postdoc | 01/02/2018 | 30/06/2018 | ChistEra Adalab |
| Viard | Tiphaine | Postdoc | 01/10/2019 | 31/08/2020 | FUI-PCU |
| Zaiou | Ahmed | doctorant | 10/10/2020 | 03/12/2022 | CIFRE EDF |
| Zevio | Stella | doctorante | 15/09/2017 | 23/02/2021 | FUI-PCU |
| Zouinina | Sarah | doctorant | 23/03/2017 | 06/10/2020 | Bourse Maroc |

Informations
générales sur
l'équipe A3

1.1.4 Évolution de l'équipe (permanents)

Les effectifs de l'équipe A³ ont subi une forte baisse depuis 2017 avec le départ de plusieurs personnes actives. Il n'y a eu aucune arrivée dans l'équipe durant cette période. Un recrutement MCF est prévu pour la rentrée 2023/2024. Un fait remarquable concernant la composition de l'équipe est qu'elle continue d'accueillir de nombreux doctorants, en particulier avec des financements CIFRE, malgré la baisse de l'effectif des permanents.

Départs

- 2017** Y. Chevaleyre, PU, IG (mutation PU Univ. Paris-Dauphine)
- 2021** H. Soldano, MCF, IG, (retraite)
- 2022** N. Grozavu, MCF, IG, (PU, Université de Cergy)
- 2022** M. Lebbah, MCF, IUTV, (PU, Université Versailles)

Arrivées

Aucune arrivée

1.2 Les thématiques scientifiques et leurs enjeux

En 2017, l'équipe A³ était structurée en quatre axes : un axe apprentissage pour l'aide à la décision, un axe apprentissage non supervisé pour le transfert et la collaboration, un axe apprentissage dans les graphes et un axe apprentissage de modèles topologiques à partir de données massives.

Le départ de plusieurs membres actifs de l'équipe a conduit à la fermeture ou à la réorientation des axes sur lesquels ils étaient actifs. Le retour d'un membre actif d'une délégation pour création d'entreprise a conduit à la création d'un nouvel axe. En 2023, l'organisation de l'équipe a trouvé un nouvel équilibre et ses activités sont structurées en trois axes : l'axe ADA (Apprentissage à partir de Données et d'Apprenants), l'axe ARG (Apprentissage Relationnel et Graphes) et l'axe MAS (Méta-apprentissage et Apprentissage de Structures).

1.2.1 Axe ADA : Apprentissage à partir de Données et d'Apprenants

Le thème central de l'axe de recherche ADA est l'apprentissage de représentations à partir de données et d'apprenants, pour la collaboration et le transfert. Il se focalise sur trois paradigmes d'apprentissage multi-modèles fondés sur trois formalismes mathématiques : apprentissage collaboratif/fédératif, apprentissage par transfert et apprentissage à partir de données multimodales à travers des formalismes émergents (pour l'apprentissage) tels que le formalisme quantique, le transport optimal, et la topologie algébrique. Les contributions de l'axe ADA abordent à la fois des recherches à caractère fondamental et des recherches plus appliquées, le plus souvent

1.2 Les thématiques scientifiques et leurs enjeux

soutenues par des projets collaboratifs académiques et industriels dans différents domaines tels que : santé, marketing digital et recommandation, diagnostic des systèmes complexes, qualité et anonymisation de données, analyse des réseaux sociaux, etc.

Dans le cas de la collaboration, nous nous intéressons à l'apprentissage non supervisé collaboratif (ou *collaborative clustering*). Il s'agit d'un paradigme d'apprentissage automatique permettant de préserver notamment la confidentialité des données en utilisant d'autres résultats de classification non supervisée, sans avoir recours aux données de ces dernières. Ayant une collection de bases de données distribuées sur plusieurs sites différents, le problème consiste à partitionner chacune de ces bases en considérant les données locales et les résultats des classifications distantes des autres bases collaboratrices, sans partage de données entre les différents sites. Le principe fondamental du clustering collaboratif est d'appliquer des algorithmes de clustering localement sur les différents sites, puis de faire collaborer les sites en partageant les résultats obtenus lors de la phase locale [RI-6]. Il existe deux grandes familles de clustering collaboratif : l'approche horizontale pour la collaboration des bases de données qui décrivent les mêmes individus avec des variables différentes ; l'approche collaborative dite verticale pour la collaboration de plusieurs bases de données décrites par les mêmes variables représentant des populations différentes. Parmi les cadres théoriques qui formalisent nos études, on peut citer le transport optimal (TO) qui définit un ensemble d'outils géométriques pour comparer les mesures de probabilité prises sur un espace métrique. Ces outils ont trouvé de multiples applications dans des domaines tels que l'économie, l'imagerie et la mécanique des fluides pour modéliser les transferts, les morphings et les déplacements. La géométrie du TO a des propriétés qui la rendent particulièrement adaptée pour résoudre des problèmes d'apprentissage artificiel, pour lesquels la capacité de comparer des mesures de probabilité empirique en grande dimension est souvent cruciale. Cette direction de recherche vise à établir un pont entre les problèmes de TO et le fondement de la science des données. Il cherche à développer de nouveaux outils méthodologiques pour l'analyse et la représentation de données accessibles sous la forme de distributions empiriques ainsi qu'à revisiter plusieurs cadres classiques de l'apprentissage artificiel (apprentissage profond, factorisation matricielle, ...) sous le point de vue du TO. Pour le contexte de l'apprentissage par transfert, les membres de l'axe ADA cherchent à utiliser un jeu de tâches pour influencer l'apprentissage et améliorer les performances sur une autre tâche. Cependant, ce paradigme d'apprentissage peut en réalité gêner les performances si les tâches (sources et cibles) sont trop dissemblables. Un défi pour l'apprentissage par transfert est donc de développer des approches qui détectent et évitent le transfert négatif des connaissances utilisant très peu d'informations sur la tâche cible. Un cas particulier de ce type d'apprentissage est l'adaptation de domaine [LI-1]. C'est une situation où les tâches sources et cibles sont identiques mais dans des domaines différents. Bien que l'informatique quantique et l'apprentissage artificiel puissent sembler a priori très différents, ils ont déjà pu se rencontrer et interagir de multiples façons par le passé. L'apprentissage quantique est le domaine qui est né de l'intersection et des rencontres entre l'informatique quantique et l'apprentissage automatique. Il s'agit d'une discipline en émergence présentant de nombreuses perspectives pertinentes pour la science. Parallèlement aux autres thèmes de l'axe ADA, nous tentons de modifier les algorithmes traditionnels d'apprentissage automatique pour les reformuler en termes quantiques. Aussi, nous avons proposé des algorithmes pour les ordinateurs quantiques pour résoudre efficacement certains problèmes qui ne peuvent pas être traités par des ordinateurs conventionnels.

Informations
générales sur
l'équipe A3

• **Théorie du transport optimal pour le clustering collaboratif, le co-clustering et l'apprentissage par transfert**

Clustering collaboratif : Le travail de recherche développé concerne le développement d'approches de clustering multi-modèles à partir de données distribuées. Nous avons travaillé sur deux volets principaux en apprentissage automatique multi-modèles. Le premier concerne le clustering multi-vues où nous visons à apprendre un modèle optimal global à partir des modèles locaux des différentes vues. Le second volet porte sur le clustering collaboratif où l'idée principale est de trouver un modèle d'échange de connaissances entre les différents collaborateurs afin d'améliorer leurs propres performances. Pour le clustering multi-vues, nous avons proposé deux approches basées sur la théorie du transport optimal [CI-61]. La première approche (CPA : Consensus Projection Approach) consiste à trouver un modèle de consensus à partir des modèles locaux, en projetant les distributions des différentes vues sur un espace global. Il s'agit ici d'un apprentissage coopératif. La seconde approche (CNR : Consensus with New Representation) vise à apprendre une nouvelle représentation consensuelle à partir de la représentation

1.2 Les thématiques scientifiques et leurs enjeux

locale des distributions. Dans le cas du clustering collaboratif, nous avons introduit une nouvelle approche basée sur la théorie du transport optimal (Co-OT : Collaborative clustering through Optimal Transport) qui vise à améliorer le mécanisme de la collaboration et la manière de transporter les informations entre les collaborateurs avec un coût minimum [CI-75]. Pour ce faire, nous avons proposé une fonction objective pour la collaboration basée sur la distance de Wasserstein. Nous avons aussi proposé une solution pour choisir les meilleurs collaborateurs par comparaison de la distribution locale des prototypes et l'analyse de la diversité entre les collaborateurs. Pour une autre approche dans ce cadre, nous avons proposé un nouveau modèle de collaboration guidé par la sélection des caractéristiques, où l'idée principale est de choisir les caractéristiques qui donnent la meilleure représentation pour chaque collaborateur et garantissant la meilleure communication entre eux, tout en préservant la confidentialité des données de chaque collaborateur [CI-74]. Cette dernière approche collaborative est aussi développée dans le cadre de la théorie du transport optimal.

Informations générales sur l'équipe A3

Co-clustering : Nous avons proposé une nouvelle méthode de co-clustering, une approche d'apprentissage non supervisé qui vise à découvrir des groupes homogènes d'instances de données et de caractéristiques en les regroupant simultanément [CI-14]. La méthode proposée utilise le transport optimal régularisé par l'entropie entre les mesures empiriques définies sur les instances de données et les caractéristiques afin d'obtenir une fonction de densité de probabilité conjointe représentée par la matrice de couplage optimale. Cette matrice est ensuite factorisée pour obtenir les partitions de lignes et de colonnes induites en utilisant l'approche des représentations multi-échelles. Pour justifier notre méthode sur le plan théorique, nous montrons comment la solution du transport optimal régularisé peut être vue du point de vue de l'inférence variationnelle, ce qui motive son utilisation pour le co-clustering. L'algorithme dérivé de la méthode proposée et sa version à base de noyaux utilisant la notion de distance de Gromov-Wasserstein sont rapides, précis et peuvent déterminer automatiquement le nombre de clusters en ligne et en colonne. Ces caractéristiques sont démontrées de façon éclatante par des évaluations expérimentales approfondies.

Apprentissage par transfert et adaptation de domaine avec garanties théoriques :

Nous avons proposé une approche consacrée au transport optimal hiérarchique pour l'adaptation de domaine [CI-107]. Elle apporte quelques réponses à la question de l'apprentissage des structures cibles cachées en utilisant le transport optimal, et leur incorporation ensuite dans le processus du transport pour résoudre le problème d'adaptation de domaine. Il s'agit d'une formulation structurelle du transport optimal qui exploite, au-delà des informations géométriques capturées par la métrique de base, des informations structurelles plus riches dans les domaines source et cible. L'information structurelle dans le domaine source étiqueté est formée instinctivement en regroupant les échantillons dans des structures en fonction de leurs étiquettes de classe. L'exploitation des structures cachées dans le domaine cible non étiqueté est réduite au problème d'apprentissage des mesures de probabilité à travers le barycentre de *Wasserstein*, que nous prouvons théoriquement être équivalent au *clustering spectral*. Pour l'approche proposée, nous avons développé un nouveau cadre théorique d'adaptation de domaine à travers le transport optimal hiérarchique [CI-14]. Ce paradigme fournit des garanties théoriques sous la forme de bornes de généralisation en permettant de considérer l'organisation structurelle implicite des échantillons dans les deux domaines en classes ou *clusters*. De plus, nous avons fourni une nouvelle mesure de divergence entre les domaines source et cible, appelée la distance de *Wasserstein* Hiérarchique, qui indique, sous des hypothèses modérées, quelles structures doivent être alignées pour mener à une adaptation réussie.

• **Apprentissage quantique à partir de données séquentielles**

Dans ce thème, notre objectif est de fournir des solutions moins complexes en utilisant des algorithmes d'apprentissage automatique quantiques pour traiter des problèmes, notamment dans le domaine des études Probabilistes de Sécurité (*PSA : Probabilistic Safety Assessment*) avec un temps raisonnable. Notre travail a été guidé par les besoins et la complexité du problème traité. Nous avons abordé les deux aspects du problème des *PSA*, statique et dynamique. Pour le cas statique, nous nous sommes intéressés à trouver toutes les combinaisons d'événements de base du système qui peuvent générer des accidents graves, nous avons proposé d'obtenir ces combinaisons d'événements de base par un algorithme quantique, en utilisant des graphes orientés au lieu

de chercher toutes les solutions d'un problème *SAT*. Notre proposition est un algorithme quantique qui utilise un nombre linéaire de *qubits* et grâce à un filtre classique, nous pouvons trouver toutes les combinaisons d'événements de base qui peuvent générer ces accidents [CI-96]. Dans le cas dynamique, nous nous sommes intéressés à trouver toutes les séquences accidentelles d'un système. Notre principal intérêt est le traitement de ces séquences. Dans le cas classique, afin de trouver toutes ces séquences, nous utilisons le graphe d'état du système et nous cherchons tous les chemins entre l'état courant et tous les états critiques. Comme ce problème est NP-complet, nous avons proposé une solution quantique pour trouver tous ces chemins. Nous avons proposé deux algorithmes quantiques, tous les deux basés sur la philosophie des marches quantiques. Le premier algorithme permet de trouver tous les chemins entre un sommet source et plusieurs sommets destination dans un graphe orienté acyclique. Cet algorithme utilise N *qubits* et M portes afin de trouver tous les chemins. Le second est une version hybride du premier, capable de traiter de grands graphes même avec un nombre réduit de *qubits*. Une autre contribution propose une approche quantique de l'algorithme *Dynamic Time Warping (DTW)* afin de calculer la similarité entre ces séquences, ainsi qu'une version capable de trouver la meilleure correspondance entre les séquences en utilisant des sous-séquences dont la longueur varie dynamiquement. Nous avons proposé également une stratégie d'apprentissage pour les Modèles de *Markov Cachés Quantiques (QHMM)* afin de générer des scénarios accidentels à partir de n'importe quel état initial du système et également pour gérer le système en temps réel [CI-96]. Nous avons proposé enfin une version améliorée de *k-means* quantique [CI-73].

Informations
générales sur
l'équipe A3

1.2.2 Axe ARG : Apprentissage Relationnel et Graphes

Cet axe rassemble des recherches dans le domaine de l'apprentissage symbolique supervisé (apprentissage de règles dans des langages logiques, notamment en Datalog) ou non supervisé (fouille de données, et notamment fouille de graphes). C'est un thème « historique » de l'équipe A³ qui s'est retrouvé vivifié par l'intérêt récent dans la communauté d'IA pour l'explication de modèles d'apprentissage.

• Fouille de graphes attribués

Cet axe rassemble nos travaux portant sur le développement d'algorithmes d'extraction de connaissances dans les graphes. Ils s'inscrivent dans la suite de nos travaux de formalisation théorique et de développement d'algorithmes efficaces pour l'analyse de réseaux attribués et de graphes biologiques, et dans le développement de l'approche de fouille de motifs *fermés* dans les graphes *attribués*. Cette approche s'appuie sur le développement d'opérateurs d'*abstraction de graphe*. L'extension d'un motif d'attribut (classiquement un sous-ensemble d'items) est constituée de l'ensemble des sommets qui sont des occurrences de ce motif. En appliquant un opérateur d'*abstraction* on réduit cet ensemble à son cœur, c'est-à-dire la partie dont le sous-graphe induit vérifie une certaine propriété topologique. Un motif fermé abstrait est alors le plus spécifique parmi ceux partageant la même extension abstraite.

Nous inspirant des travaux de J. Kleinberg sur l'importance relative des sommets selon leurs rôles nous avons d'abord introduit la définition d'un nouveau cœur de graphe (*Hub-Autorité*), spécifique des graphes orientés, dans lequel un sommet est retenu en raison de son degré sortant suffisamment grand (il a alors un rôle de *hub*) ou en raison d'un degré entrant suffisamment grand (il a alors un rôle d'*autorité*) dans le sous-graphe induit. Nous avons alors étendu notre algorithme d'énumération de motifs fermés abstraits dans ce cadre. Lorsque les sommets peuvent ainsi avoir plusieurs rôles dans un sous-graphe, il est intéressant de considérer des bi-motifs, constitué de deux composantes chacune associée à un rôle dans le sous-graphe. Nous avons ainsi défini de nouveaux sous-graphes cœurs, défini les bi-motifs fermés abstraits associés, et étendu notre algorithme d'énumération à ces bi-motifs. Ceux-ci peuvent être définis et utilisés pour analyser des graphes orientés ou des graphes bi-modaux, graphes bi-partis liant deux types de nœuds comme par exemple auteurs et articles dans des données bibliographiques, ou encore pour faire apparaître des structures cœur/périphérie dans des graphes non-orientés [CO-7, CO-11].

Nous nous sommes également intéressés à la sélection d'un sous-ensemble de ces motifs fermés abstraits d'abord en considérant les motifs individuellement. La *modularité locale* d'un sous-graphe, issue des travaux sur la détection de communautés dans les graphes, peut-être utilisée pour sélectionner des motifs intéressants au sens où ils ont plus d'arêtes internes qu'attendu. Une telle contrainte, non anti-monotone, avait auparavant été introduite dans l'énumération de motifs via un estimateur anti-monotone optimiste. Pendant la visite de M. Atzmueller dans le laboratoire

1.2 Les thématiques scientifiques et leurs enjeux

nous avons appliqué cette idée à notre algorithme d'énumération de motifs fermés abstraits, évitant ainsi une coûteuse étape de sélection des motifs de modularité locale suffisante postérieure à l'énumération [TU-1, CI-54].

Cependant, si notre approche vise à minimiser le nombre de motifs énumérés en ne considérant que les motifs fermés supportés par un cœur de graphe, et en sélectionnant des motifs individuellement intéressants pendant l'énumération, il reste des situations où leur nombre est très important. C'est notamment le cas avec les bi-motifs dont la contrepartie au gain d'expressivité est une explosion combinatoire du nombre de motifs. C'est une question omniprésente en fouille de données : peut-on sélectionner un nombre suffisamment restreint de motifs pour qu'ils soient présentés aux utilisateurs et experts du problème abordé ? Nous nous sommes intéressés à la sélection de sous-ensembles de motifs respectant collectivement une contrainte de diversité et avons développé un algorithme glouton de sélection d'un sous-ensemble de motifs divers [CI-4].

Nous avons appliqué et adapté ces méthodes à l'exploration de plusieurs graphes de terrain. En collaboration avec le sociologue E. Lazega nous avons exploré les relations de conseils entre pairs dans un réseau professionnel d'avocats sous l'angle des hubs et des autorités [CI-4] et avec S. Bary nous avons étudié un jeu de données épistémologiques. La densité du graphe construit par S. Bary nous a conduit en particulier à proposer une méthode de partitionnement des arêtes, permettant notamment de faciliter l'exploration et la visualisation de graphes denses [CI-27]. La fouille de bi-motifs a été développée lors de l'étude de ces deux domaines, pour rendre compte de la diversité des structures de conseils en ce qui concerne les avocats (ceux qui demandent un avis n'ont pas le même profil que ceux à qui on le demande) et pour examiner les liens entre campagnes et participants dans l'étude épistémologique de campagnes d'exploration des fonds sous-marins.

Dans le cadre d'un projet international CHIST-ERA, nous avons, par l'analyse de réseaux de co-régulateurs de l'expression des gènes de la levure, identifié des motifs dont les gènes formant l'ensemble support sont des candidats pertinents à l'expérimentation biologique pour l'évaluation de leur implication dans la transition diauxic [RI-11]. En collaboration avec l'équipe RCLN du LIPN, dans le cadre du projet FUI-PCU, nous avons exploré les réseaux de collaborations scientifiques pour identifier des réseaux d'experts et d'expertises [CO-4], [CO-8]. Dans ce contexte nous avons été conduits à développer une approche permettant d'imposer aux deux parties d'un bi-motif de partager une description commune sur une partie du langage de description [CI-78].

Pour finir, nous avons profité du post-doc de Tiphaine Viard pour généraliser notre approche de fouille de motifs aux flux de liens. Ceux-ci offrent une représentation de l'évolution temporelle des interactions entre individus là où les graphes ne peuvent que capturer les interactions à un instant donné. Nous avons développé pour cela un cadre d'énumération, un algorithme et une implémentation de la fouille de motifs dans les flux de liens attribués [CI-112].

• Programmation Logique Inductive

La Programmation Logique Inductive (PLI dans la suite) est un domaine de l'apprentissage symbolique qui s'intéresse à l'apprentissage, le plus souvent supervisé, de modèles exprimés dans des restrictions de la logique des prédicats, comme Datalog. Les travaux réalisés autour de la Programmation Logique Inductive se sont poursuivis et développés pendant cette période, notamment en collaboration avec *Nukkai* et dans le cadre de la délégation de H. Soldano (2019-2021), de la demi-délégation de C. Rouveïrol (2019-2020).

Exemples ambigus Cette recherche fait suite au cadre *d'apprentissage relationnel à partir d'exemples ambigus* que nous avons conçu, où chaque exemple est un ensemble de descriptions $\{x_1, \dots, x_n\}$, une seule description, non déterminée, représentant la réalité. Nous appelons cette description *l'exemple caché*. Nous avons montré que pour tenir compte de l'ambiguïté des exemples il fallait remplacer une classique relation de couverture entre une hypothèse h et un exemple e , positif ou négatif par deux relations dites de *compatibilité* : l'une pour les exemples positifs, h compatible⁺ avec e si h couvre au moins une description de e , l'autre pour les exemples négatifs, h compatible⁻ avec e si h ne couvre pas au moins une description de e .

Suivant ce cadre, nous avons implémenté un système nommé *Lear*, où les descriptions d'un exemple sont les modèles d'une théorie clauseale composée de clauses générales, les hypothèses des clauses de Horn et la relation *couvre* la thêta-subsumption. *Lear* cherche une hypothèse compatible⁺ avec les exemples positifs et compatible⁻ avec les exemples négatifs.

Nous avons poursuivi cette recherche en associant à chaque description d'un exemple ambigu la probabilité d'être l'exemple caché. Ainsi nous pouvons définir la probabilité qu'une hypothèse

h couvre l'exemple caché dans e : c'est la probabilité que l'exemple caché soit dans les possibilités de e couvertes par h . Les relations de compatibilité ne tenant pas compte des probabilités, nous utilisons un critère de justesse pour guider la recherche d'une solution : la justesse d'une hypothèse h est l'espérance de la probabilité que h classe correctement un exemple e tiré au hasard, c'est-à-dire couvre l'exemple caché dans e si e est un exemple positif, et ne le couvre pas si e est un exemple négatif. Le but de l'apprentissage est de trouver une hypothèse de justesse maximale.

Nous avons conçu un nouveau système, nommé *nukLear*¹, qui adopte la représentation des hypothèses et exemples de *Lear*, mais associe des probabilités aux possibilités et utilise le critère de justesse pour chercher une hypothèse solution.

ILP et explications Nous avons étudié, également en collaboration avec *Nukkai*, l'utilisation de systèmes d'apprentissage supervisé de PLI dans le cadre d'une boucle d'interaction avec des experts de Bridge, dans le but de produire des explications associées à une prise de décision. Dans un premier temps, nous l'avons étudié une méthodologie pour résoudre un problème relativement simple de Bridge (décision d'ouverture dans le cas de mains « limites ») [CI-32], puis à un problème plus complexe d'enchères [CO-19] où les classes des exemples ont été attribuées par un processus boîte noire utilisant notamment des simulations. Dans les deux cas, des modèles à la fois explicables et précis ont été appris, notamment grâce à une interaction avec des experts de Bridge et au fait que les modèles produits par PLI soient explicites.

D'autres travaux en cours dans le cadre de la thèse CIFRE de M. Kazi Aoual abordent le problème de l'explication d'une décision dans un cadre *MDP* : étant donné un état initial, une séquence d'actions est effectuée et la récompense associée à la séquence d'actions est perçue dans l'état final. Dans le cadre du jeu de Bridge, les actions sont des cartes jouées, l'état initial est une donne ; un joueur (déclarant) joue contre un adversaire (défenseur) dont on connaît *a posteriori* les cartes, le but est d'expliquer quelles décisions (cartes jouées) ont mené à un gain de la partie, et pourquoi. Nous avons choisi de représenter les trajectoires (séquences d'actions) comme des interprétations, décrivant à la fois les actions effectuées, mais également les propriétés des états atteints par ces trajectoires, certaines de ces propriétés perdurant sur des intervalles de temps le long de la trajectoire. Étant donné un groupe T^+ de trajectoires optimales et un groupe T^- de trajectoires non optimales, nous avons conçu un algorithme permettant de construire des explications *pour ce groupe*, comme les sous-ensembles minimaux et corrects de la généralisation maximale spécifique de T^+ . Une implémentation est proposée permettant de résoudre les problèmes de complexité de ces étapes pour des langages relationnels (Datalog).

Les jeux sont un terrain d'applications privilégié pour identifier et expliquer des trajectoires optimales dans un cadre *MDP*, mais ces questions peuvent être transférées à d'autres problèmes réels. Par exemple, nous sommes également intéressés à des applications d'enseignement numérique. Pour ce travail, un exercice d'algorithmique est posé à un étudiant, qui va le résoudre progressivement en appliquant des actions prédéfinies, et où le problème est d'identifier à des fins de recommandations et d'explication pourquoi une séquence d'actions est optimale pour résoudre le problème. Ce travail, très exploratoire, a été initié dans le cadre d'un BQR financé par l'Université Sorbonne Paris Nord, qui a permis de tester une première représentation des traces de résolution d'exercices et d'évaluer les principales difficultés pour apprendre une politique définissant quelle action doit être appliquée dans un état – une solution partielle à l'exercice – donné par des techniques d'apprentissage par renforcement classique.

Enfin, dans le cadre d'une collaboration avec l'équipe RCLN, nous sommes intéressés à l'application de l'apprentissage relationnel à l'apprentissage de règles discriminantes dans le cadre de données graphes [CI-83]. Ces travaux ont permis d'évaluer les avantages et limites de chacune des méthodes en présence (arbre de décision relationnel et fouille de règles relationnelles « à la AMIE » sur le problème donné et d'initier une collaboration sur l'apprentissage de règles en présence de très peu d'exemples étiquetés et d'une théorie du domaine de grande taille représentée sous forme de graphe de connaissances.

1.2.3 Axe MAR : Métaapprentissage et Apprentissage de Structures

Cet axe regroupe les activités en métaapprentissage et en modélisation simultanée des données et du clustering (apprentissage non supervisé profond) à large échelle appliqués aux données temporelles multivariées, aux images et aux documents audios et vidéos.

1. *nukLear* est développé par *Nukkai*, dans le cadre d'une collaboration avec le LIPN.

1.2 Les thématiques scientifiques et leurs enjeux

Le métaapprentissage comprend l'ensemble des processus qui permettent d'apprendre à apprendre. Il inclut, en particulier, l'analyse des algorithmes d'apprentissage pour sélectionner le plus approprié en optimisant l'efficacité et l'efficacité, la conciliation, la composition ou le renforcement de modèles de solutions issus de perspectives différentes, le transfert partiel ou total et le contrôle du changement d'espaces de représentations, le choix des fonctions d'évaluation et la prise en compte de toutes les formes de connaissances a priori et leur relativité, en plus des données, pour améliorer l'apprentissage courant et futur en minimisant l'effort, le temps et les ressources.

Mieux apprendre ne s'évalue pas uniquement en fonction des critères d'efficacité sur les données du problème spécifique, mais aussi sur l'efficacité de la solution globale incluant, par exemple, l'évaluation des critères tels que le respect de l'éthique, le respect de la loi, le respect de la vie privée, la qualité de l'explication ou de l'interprétation, l'effort en ressources consommées (facteur écologique, etc.) ou le nombre d'expériences nécessaires pour apprendre.

Un modèle intégré unique, basé souvent sur des modèles de fonctions non linéaires complexes optimisées par des biais de choix des hypothèses minimisant des mesures générales ne suffisent pas pour répondre aux autres exigences, qui s'évaluent sur d'autres critères souvent impossibles à exprimer dans ces modèles. Ceci nécessite le développement de nouveaux paradigmes d'apprentissage multimodèles chacun responsable de certifier un sous-ensemble de fonctions d'évaluations du système global. Notre définition du métaapprentissage c'est aussi la proposition de nouvelles architectures de conception capables d'auditer, de qualifier et plus généralement d'évaluer les modèles d'apprentissage selon des normes préétablies par de larges communautés, dépassant celles des experts en intelligence artificielle.

Cet axe regroupe les recherches qui tentent de proposer des solutions et modèles génériques pour ces problèmes à travers l'exploration de divers paradigmes complémentaires pour améliorer les performances avec moins d'exemples en tenant compte de la structure explicite ou non, en recherchant des régularités par la minimisation du risque structurel notamment, en modélisant le contexte et les divers biais, en explicitant la nature des données (car les données n'ont pas toutes le même impact, ni les mêmes caractéristiques -vérité, certitude, imprécision, etc.-) [RI-18] et en modélisant les interactions entre les concepts à apprendre et concepts appris selon divers points de vue. C'est aussi tenir compte de l'évolution des théories apprises et de la métamodélisation des hyperparamètres. L'objectif ultime est de construire un cadre général et générique pour le processus d'apprentissage lui-même. Nous détaillons ci-dessous les principales contributions.

Intégration de biais de contraintes structurelles Durant cette période exploratoire de l'impact du métaapprentissage (MML) et de la structuration du risque structurel sur la qualité de l'apprentissage, nous avons choisi de travailler essentiellement sur les applications issues de l'internet des objets [CI-81]. Ce choix, très ciblé, répond à plusieurs problématiques fondamentales du MML dont la présence de plusieurs points de vue sur les phénomènes d'intérêt (relativité), la coexistence de plusieurs théories cohérentes, mais d'apparence incompatibles (apprentissage fédéré, distribué, etc.), la multiplicité des modalités peu étudiées ensemble (son, image et les divers types de capteurs), existence de modèles équationnels qui régissent une partie des phénomènes observés, le mode de disponibilité des données (souvent par fenêtres temporelles), leur hétérogénéité et leurs distributions, etc. qui s'ajoute aux problématiques habituelles liées au choix de meilleures hypothèses qui représentent au mieux les exemples d'apprentissage. Cette famille d'applications impose de nouveaux paradigmes d'apprentissage automatique qui doivent inclure la modélisation à la fois des données, des structures des espaces d'hypothèses, des fonctions d'évaluation, de l'adaptation, du renforcement, du transfert, de l'explicabilité, de l'interprétabilité des résultats et des décisions, de l'incrémentalité de l'apprentissage, de la quantification des méthodes utilisées pour apprendre en plus de la quantification liée aux données, du respect de diverses normes et contraintes, etc. Nous avons proposé plusieurs approches basées sur les biais de contraintes structurelles et sur la structuration du processus d'apprentissage. La première catégorie de travaux porte sur la prise en compte de différentes formes de structures du domaine et leur intégration dans le processus d'apprentissage [CI-52, CI-86, CI-80, 1]. Alors que la deuxième utilise les connaissances et la sémantique de l'espace des catégories pour organiser le processus d'apprentissage [CI-81, CI-97]. L'objectif est la fois la réduction du nombre d'exemples nécessaires pour apprendre et la recherche de modèle de prise en compte de ce type de contraintes dans une modélisation générique du MML.

Abstraction du contexte et relativité Les données réelles qu'on manipule sont essentiellement issues de processus de transformation via des capteurs en utilisant des fonctions de

transferts. Des biais intrinsèques entachent les données par ce simple fait [CI-81]. De plus, l'emplacement des sources de données apportent des biais de point de vue dont la connaissance améliore sensiblement la qualité de l'apprentissage. Nous avons travaillé sur les mécanismes de conciliation des apprenants par l'abstraction de contexte et également, quand la connaissance du domaine le permet, la modélisation explicite ou implicite des données relatives afin de distinguer les données qui dépendent du contexte local des données générales utilisables dans un contexte de centralisation de la théorie de l'apprentissage [CI-98, CI-105].

Quantification des biais Quantifier les biais consiste à prendre en compte tout le pipeline d'apprentissage, depuis la modélisation de l'environnement pour définir la typologie puis la topologie des capteurs des données, à l'acquisition de l'information via les fonctions de transfert, à la prise en compte des biais dans les processus de transformation et de transmission des données et au choix des fonctions d'évaluation. Dans nos derniers travaux, en plus de ceux plus théoriques, nous avons exploré l'impact de ces biais dans diverses applications [TU-3] sur le biais des modèles d'évaluation ou l'article [TU-3] sur l'étude de la segmentation et de l'évaluation comme problèmes eux-mêmes à apprendre. L'outil développé est disponible ici : (accéder au Code). ou la redéfinition de la segmentation comme un problème de décomposition nécessitant de quantifier la perte d'information issue de la segmentation, l'apprentissage après segmentation et le lien qui existe entre la solution du problème d'apprentissage sur le modèle segmenté et le problème initial [CI-101]. L'outil développé est accessible ici (accéder au Code) et ici pour l'évaluation de la décomposition (Code) [CI-99]. Dans le domaine de l'analyse du son, l'étude du problème d'évaluation d'événements sonores selon l'angle du métaapprentissage, nous a permis de présenter une méthode multimodale simple et visuelle pour l'évaluation des événements sonores [CI-100] (accéder au Code) qui permet d'analyser les biais liés au choix des modalités et les résultats obtenus).

Informations
générales sur
l'équipe A3

Multi-Coclustering topologique et visualisation En apprentissage automatique, les modèles de mélanges topologiques et leurs variantes stochastiques sont très présents en statistique; plus généralement, l'hypothèse markovienne est souvent invoquée à chaque fois que des probabilités sont utilisées pour modéliser l'état d'un processus dynamique. Ces modèles de Markov sont souvent combinés avec d'autres algorithmes de classification. Nous avons étudié la possibilité d'utiliser conjointement les modèles de mélange et/ou les modèles topologiques pour traiter des séries temporelles multivariées (e.g., fonctionnelles). Cela consiste à estimer automatiquement et simultanément le modèle, la structure cachée ainsi que le nombre de classes, supposé lui aussi inconnu. Les mélanges de lois permettent de modéliser des distributions de probabilités d'une manière à la fois flexible (une grande diversité de distributions peuvent être modélisées sous cette forme) et parcimonieuse (la distribution ainsi modélisée est décrite par un nombre réduit de paramètres) [CI-87]. Parmi les résultats obtenus, nous citons le développement d'un modèle de Multi-Coclustering non paramétrique. Ces travaux ont été valorisés dans le cadre d'applications concrètes dans le contexte d'un projet *CIFRE* avec *Renault*.

Pour améliorer les résultats du clustering, nous avons également abordé d'autres travaux liés aux changements de représentations tel que le travail sur le regroupement de sous-espaces (subspace clustering), nous définissons un double système de pondération pour détecter les attributs et les sous-espaces d'attributs "influenceurs" dans un clustering. Nous avons étudié deux types de pondération, la pondération globale dans laquelle un seul vecteur de pondération est attribué pour chaque attribut et sous-espace, et la pondération locale où un vecteur de pondération différent est attribué pour chaque attribut et sous-espace d'un cluster [CI-93].

Dans la même direction, nous nous sommes également intéressés, au clustering multiobjectifs en utilisant des méthodes évolutionnaires. Contrairement aux méthodes classiques qui détectent les clusters en se basant sur une seule mesure de similarité, les approches évolutionnaires élargissent l'espace de recherche et permettent de détecter les clusters via des fonctions objectifs et un algorithme génétique [CI-93]

Apprentissage profond et Clustering Les succès récents des architectures profondes pour le clustering (unsupervised deep learning) nous a conduit à développer un modèle qui apprend conjointement la représentation des données et la représentation des référents des cartes topologiques (SOM : Self-Organizing Map). Notre modèle est composé d'un autoencodeur et une couche SOM optimisée lors d'une procédure d'apprentissage conjointe (DESOM : Deep Embedded Self-Organizing Map) [CI-57, CI-35]. les résultats obtenus dans le cadre de la *CIFRE* avec *Safran* sur les séries temporelles nous ont confronté dans l'intérêt de coupler l'apprentissage profond aux méthodes d'apprentissage non supervisé afin d'encoder/vectoriser des données

brutes non formatées (par exemple les séries temporelles, les images, etc.). Les données disponibles pour le clustering sont issues de sources hétérogènes, ces problèmes d'encodage multimodales constituent une dimension intéressante de recherche qui reste encore à explorer.

De plus, la considération de la contrainte de scalabilité des modèles nous a conduit à la proposition d'un nouvel algorithme utilisant un système de hachage (LSH : Locality Sensitive Hashing) tout en bénéficiant du récent paradigme MapReduce appliqué aux systèmes distribués. Nous avons proposé également la variante de l'algorithme de clustering pour les données catégorielles en utilisant le codage binaire et la distance de Hamming [CI-45], aussi nous avons introduit une amélioration du Clusterwise en combinant l'algorithme de clustering et un modèle de régression [RI-13]. L'amélioration de sa complexité en temps, en appliquant le clustering avant une tâche de régression PLS, a permis d'assurer le passage à l'échelle dans le domaine de la sélection de variables. Nos résultats sont mis en œuvre et testés dans le cadre d'applications réelles et ont conduit au projet opensource Clustering4Ever et une collaboration CIFRE avec Kameleoon.

Références complémentaires

- [1] M. [Hamidi](#) and A. [Osmani](#). Description of Structural Biases and Associated Data in Sensor-Rich Environments. *CoRR*, abs/2104.04885, 2023.

1.3 Profil d'activités liées à la recherche

Activités (répartir 100 points sur ces 7 items)

| | |
|---|----|
| Administration et animation de la recherche : pilotage de la recherche (VP, direction d'institut, DAS, par exemple), participation à des instances d'évaluation (CNU, CoNRS, CSS, Hcéres, par exemple), responsabilité de dispositifs Idex ou Isite, direction de projets (ANR, Horizon Europe, ERC, CPER, PIA, France 2030, par exemple), responsabilités éditoriales dans des revues ou collections nationales et internationales. | 10 |
| Aide aux politiques publiques et expertise technique : pouvoirs publics aux niveaux européen, national et régional, entreprises, instances internationales comme FAO, OMS, etc. | 5 |
| Contribution à l'adossement d'enseignements innovants à la recherche : EUR, SFRI, etc. | 5 |
| Dissémination de la recherche : partage de connaissances avec le grand public, médiation scientifique, interface sciences et société | 5 |
| Recherche et encadrement de la recherche. | 50 |
| Valorisation, transfert, innovation. | 20 |
| Autres activités. (à préciser en une ligne maximum). | 5 |

1.4 Prise en compte des recommandations du précédent rapport

Lors de la précédente évaluation avait jugé que : *"le contexte, favorable à l'IA et à l'apprentissage automatique, devrait permettre à l'équipe de gagner en visibilité et attractivité"... "Les productions sont de très bon, d'excellent niveau, pourraient être renforcées par la considération de conférences plus centrales de l'apprentissage automatique, (ECML/PKDD) ou (NIPS). L'équipe doit se poser la question des conférences qu'elle souhaite cibler... pour renforcer son attractivité et sa visibilité internationale."*

Les dernières années le contexte est devenu encore plus favorable à l'IA et à l'apprentissage automatique, au point que les postes universitaires deviennent largement moins compétitifs que ceux proposés par les grands groupes internationaux. L'équipe a perdu plusieurs de ses membres depuis 2017 et n'a eu aucun recrutement. Malgré ce contexte, la production de l'équipe reste soutenue et plusieurs papiers sont acceptés dans des conférences centrales de l'apprentissage automatique comme ECML/PKDD, PAKDD, NIPS, IJCAI ou AAAI.

Le comité avait également indiqué que *"les thématiques mises en avant par l'équipe sont parfois peu "lisibles" et gagneraient à être simplifiées.* De plus, le comité a indiqué que l'équipe

1.4 Prise en compte des recommandations du précédent rapport

jouit d'une *Excellente production scientifique, avec des publications dans des conférences et des journaux de tout premier plan. Elle a une visibilité et une attractivité fortes au niveau national et participe de façon importante à l'animation de la communauté apprentissage en France. L'attractivité et la présence de l'équipe à l'international peuvent toutefois être renforcées.*

Un effort a été fait pour structurer les activités de l'équipe en passant de quatre axes à trois axes avec un affichage beaucoup plus clair et une identité renforcée. Malgré l'amélioration des collaborations à l'international avec, notamment, plusieurs (8) séjours dans des laboratoires étrangers par des membres de l'équipe, la participation ou l'organisation à plusieurs conférences et workshops internationaux, l'équipe doit fournir des efforts supplémentaires pour améliorer sa présence et sa visibilité à l'international.

Le comité a indiqué que *la vie de l'équipe mérite d'être renforcée avec des activités plus régulières permettant de mieux partager l'information et d'impliquer tous ses membres. Il est important de faire en sorte que tous les membres participent à la production et aux collaborations de l'équipe. L'équipe devra aussi être attentive dans ses recrutements futurs et veiller à un bon équilibre F/H... le ratio rang A/rang B, extrêmement défavorable*, et les responsabilités administratives importantes assurées par plusieurs membres seniors représentent toutefois des facteurs de risque pour le futur. Il est important que l'équipe A3 puisse s'appuyer sur plus de membres de rang A et se "décharge" de certaines responsabilités administratives actuelles.

Plusieurs membres de l'équipe ont été et sont encore très impliqués dans des responsabilités administratives (VP USPN, directrice adjointe de l'IG en charge de l'école d'ingénieurs, directeurs de département d'IUT, chargé de mission auprès de la présidence). Ce facteur et l'absence de recrutement de membres de rang A constituent un facteur de risque pour l'équipe.

L'équilibre homme/femme s'est amélioré du fait du départ de plusieurs membres et grâce au soin pris à recruter des doctorantes (sur la période, il y a eu 11 doctorantes sur un total de 23 doctorant.e.s). L'équipe compte une professeure et un professeur et une maîtresse de conférences contre neuf maîtres de conférences.

**Informations
générales sur
l'équipe A3**

Chapitre 2

Introduction du portfolio

Le portfolio qui suit met en exergue une partie de l'expertise de l'équipe à travers cinq réalisations :

Transport optimal Des avancées récentes en adaptation du domaine sont dues à la théorie du transport optimal. Nous présentons nos résultats publiés en 2022 dans *Machine Learning*.

Données relatives et apprentissage de représentation Le métaapprentissage s'intéresse à la structuration et à l'apprentissage de tous les biais permettant d'apprendre à apprendre. Il s'agit ici de décrire une série de résultats [CI-105, CI-97, CI-99] qui présentent deux biais importants et leurs apports pour l'apprentissage : la relativité des données et leur impact sur le succès de l'apprentissage et l'utilisation d'invariants pour la structuration de l'espace de représentation des concepts à apprendre.

Un livre sur la théorie de l'adaptation du domaine écrit en collaboration avec un enseignant chercheur de l'équipe, ce livre présente les résultats les plus récents en apprentissage par transfert.

La startup HelpIA Les collaborations industrielles de membres de l'équipe ont conduit au développement, en 2020, d'un projet de création d'une startup à succès mise à l'honneur dans ce portfolio.

Apprentissage de modèles structurés et projet CHIST-ERA Adalab Des membres de l'équipe ont participé à ce projet européen Dont l'objectif était d'accélérer les découvertes scientifiques grâce à la conception et au développement de robots scientifiques autonomes capables de mettre à jour un modèle, de réaliser et de choisir des expériences plus rapidement et de manière fiable [RI-11]. Les bons résultats obtenus, dans le cadre de ce projet, sont détaillés dans ce portfolio.

Chapitre 3

Auto-évaluation de l'équipe A3

Sommaire

| | |
|---|-----------|
| 3.1 Domaine 2. Attractivité | 32 |
| 3.1.1 Référence 1. L'équipe est attractive par son rayonnement scientifique et s'insère dans l'espace européen de la recherche. | 32 |
| 3.1.2 Référence 2. L'équipe est attractive par la qualité de sa politique d'accompagnement des personnels. | 33 |
| 3.1.3 Référence 3. L'équipe est attractive par la reconnaissance de ses succès à des appels à projets compétitifs. | 34 |
| 3.2 Domaine 3. Production scientifique | 34 |
| 3.2.1 Référence 1. La production scientifique de l'équipe satisfait à des critères de qualité. | 35 |
| 3.2.2 Référence 2. La production scientifique de l'équipe est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels. | 35 |
| 3.2.3 Publications | 36 |
| 3.3 Domaine 4. Inscription des activités de recherche dans la société | 51 |
| 3.3.1 Référence 1. L'équipe se distingue par la qualité et la quantité de ses interactions avec le monde non-académique. | 51 |

3.1 Domaine 2. Attractivité

| | |
|---|----|
| 3.3.2 Référence 2. L'équipe développe des produits à destination du monde culturel, économique et social. | 51 |
|---|----|

3.1 Domaine 2. Attractivité

3.1.1 Référence 1. L'équipe est attractive par son rayonnement scientifique et s'insère dans l'espace européen de la recherche.

L'équipe est active dans sa communauté scientifique, à la fois en termes de publications de haut niveau, de participation aux comités de programme de conférences de premier plan ou d'organisation ou d'invitations à des conférences et workshops internationaux. La nouvelle structuration de l'équipe commence à donner ses fruits. A la fois au niveau théorique (transport optimal, métaapprentissage ...) et au niveau des applications de haut niveau.

Invitations dans des institutions académiques ou des congrès internationaux

L'équipe a été invitée au Colloque international de l'UM6SS sur le thème *Artificial Intelligence for health care*, à la journée FHU MOSAIC' sur le thème *Multiscale Optimised Strategy for Artificial intelligence-based Imaging biomarkers in digestive Cancer*, à la conférence CIFSD'18-Tanger sur *Limits of Artificial Intelligence*, par le centre de conférences CURI-USMBA de Fès sur *Behind Deep Learning*. Elle a également participé au colloque *sciences et sens de l'intelligence artificielle* avec deux conférences invitées l'une sur la *Chirurgie augmentée par l'Intelligence Artificielle* et l'autre sur *l'Intelligence Artificielle : entre confiance et défiance* qui ont eu lieu au campus Condorcet.

Durant cette période, les membres de l'équipe ont effectué huit séjours dans des laboratoires étrangers. Il y a eu deux séjours en Italie, à Caserta, à l'Université Campania Luigi Vanvitelli et à Naples; un séjour à l'université du Texas; deux séjours au Maroc, à Tanger (ENSAT) et Fès (centre CURI); un séjour en Vietnam (USTH); un séjour au Sénégal et deux séjours aux États-Unis (université de Californie et université du Texas).

Auto-évaluation de l'équipe A3

Conférences où des membres de l'équipe sont membres du PC

Chaque année, les membres de l'équipe participent aux comités de programmes de plusieurs conférences et journaux internationaux de premier plan. Parmi ces conférences, on trouve régulièrement : American Conference on Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), ICONIP (International Conference on Neural Information Processing), International Conference on Machine Learning (ICML), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), IJCNN (International Joint Conference on Neural Networks), Inductive Logic Programming conference (ILP), ISCV (International Conference on Intelligent Systems and Computer Vision).

Plusieurs membres de l'équipe participent aux comités de programmes et l'organisation de la recherche au niveau national. Ils participent aux comités de programmes des conférences tels que : Conférence Francophone d'Apprentissage (CAP), Conférence Internationale Francophone Extraction et Gestion des Connaissances (EGC), Rencontres annuelles de la Société Francophone de Classification (SFC), Apprentissage Artificiel et Fouille de Données (AAFD), French Regional Conference on Complex Systems (FRCC), APIA (Applications Pratiques de l'Intelligence Artificielle) sur la plateforme de l'AFIA (Association Française de l'Intelligence Artificielle).

Organisation de conférences Une année sur deux plusieurs l'équipe organise le workshop international ALML (International Workshop on Advances in Learning from/-

with Multiple Learners). Des membres de l'équipe ont participé à l'organisation de la 23ème conférence IV 2019 (23th International Conference Information Visualisation), de la conférence ICHS (International Conference 24 hours Holographic Surgery), de la conférence internationale ACONTA (International Conference on Augmented Complex Networks), du symposium international CS2 (IEEE International Symposium on Cloud and Services Computing), de la conférence EGC (Extraction et de Gestion des Connaissances), de la conférence CIFSD (Conférence Internationale Francophone sur la Science des données), de deux workshop LSISA (Workshop on Large-scale Industrial Time Series Analysis) en marge de la conférence ICDM et du workshop international IWMAN (International workshop on Mining Attributed Networks).

Elle organise des sessions adossées à des conférences internationales comme la session *Applying Complex Network Analysis for Cybersecurity* à la conférence internationale EICC à deux reprises, la session *Data Science Approaches or Modeling, Analyzing and Mining Networks on Networks* à la conférence DSAA, la session *Applying Complex Network Analysis to Industrial Risk Studies* à la conférence Networks et la session *Representation learning for Transfer and Collaborative approaches* associée au Symposium IEEE SSCI (Series on Computational Intelligence).

L'équipe organise également régulièrement l'école thématique "Deep Learning & Data Science".

Responsabilités éditoriales dans des revues et des collections de haut niveau

L'équipe a eu des activités de responsabilité éditoriales durant la période. Younès Bannani est *Associate Editor* du journal international Knowledge and Information Systems (Kais) édité par les éditions Springer et est *Deputy/Managing Editor* du journal Marocain *Pure and Applied Analysis*. Hanane Azzag et Mustapha Lebbah ont participé à l'édition des actes de la conférence internationale IV en 2019 et à ceux de la conférence EGC en 2018.

Participation à des instances de pilotage de la recherche et d'expertise scientifique

Plusieurs membres de l'équipe ont participé à l'expertise de projets ANT, ANRT. Younès Bannani était notamment nommé expert sur des dossiers Masters pour l'HCERES et auprès du Pôle de Compétitivité Cap Digital. Rushed Kanawati était également nommé expert d'un projet pour la Czech Science Foundation en 2020. Mustapha Lebbah était nommé expert scientifique pour le ministère de l'enseignement, de la recherche et l'innovation et membre externe du conseil scientifique du centre de développement des technologies avancées en Algérie (CDTA). Hanane Azzag est membre du bureau de l'association EGC et membre de son comité de pilotage depuis 2019. Céline Rouveïrol est membre du conseil d'administration de l'AFIA depuis 2020 et responsable du *collège Apprentissage Automatique*.

Auto-évaluation de l'équipe A3

3.1.2 Référence 2. L'équipe est attractive par la qualité de sa politique d'accompagnement des personnels.

Le laboratoire a mis en place des dispositifs d'accompagnement des doctorants dont profite totalement l'équipe. L'équipe bénéficie des allocations de recherche de l'école doctorale réparties en favorisant les meilleurs dossiers et une équité entre les équipes. Par ailleurs, les membres de l'équipe disposent d'un bon réseau de partenaires industriels qui compte des entreprises comme SAFRAN, Renault, Meta, EDF, Orange, BPCE et Nukkai leur permettant de nombreux financements. Sur la période concernée par l'évaluation, le nombre de doctorants dans l'équipe est passé à 33 dont 14 thèses CIFRE (dont 18 thèses soutenues), en augmentation substantielle par rapport à la période précédente (23 doctorants dont 10 thèses CIFRE). Cette augmentation du nombre s'est également accompagnée de l'amélioration de la qualité du recrutement des doctorants. Tous les membres HDR de l'équipe participent également à des co-encadrements de

3.2 Domaine 3. Production scientifique

thèses en collaboration avec des membres des autres équipes (Calin, AOC, RCLN), avec les autres laboratoires de la fédération MathStic (L2TI et LAGA) et avec d'autres universités (UPEC, UVSQ, Cergy).

La durée moyenne des thèses est d'environ 41 mois contre 45 mois lors de la précédente période d'évaluation.

L'équipe a accueilli également six post-doctorants entre 2017 et 2022. Deux dans le cadre du projet Européen ChistEra Adalab, une dans le cadre de la bourse Marie SKLODOWSKA-CURIE, une dans le cadre du projet FUI-PCU et un dans le cadre de la startup HelphIA.

L'équipe n'a accueilli qu'un ingénieur d'étude CNRS durant un an, financé dans le cadre du projet France relance ACOMI. Ce manque d'ingénieurs constitue une des faiblesses de l'équipe. Plusieurs développements de logiciels sont faits, plusieurs plateformes d'expérimentations sont développées sans arriver à leurs termes par manque de compétences techniques et permanentes dans l'équipe pour assurer le maintien, le suivi et la diffusion appropriée.

L'équipe compte également plusieurs membres associés qui sont au nombre de douze en janvier 2023.

Les membres de l'équipe se font inviter dans plusieurs laboratoires dans le monde. Il y a eu huit séjours durant cette période : en Italie (l'Université Campania "Luigi Vanvitelli", Caserta et à l'université de Naples), aux USA (université du Texas et université de Californie), au Vietnam (USTH), au Sénégal (AIMS) et au Maroc (Tanger et au centre CURI-USMBA de Fès). Sept chercheurs ont été invités dans l'équipe entre 2017 et 2022. Ils participent à l'animation de la vie scientifique et collaborent à divers projets.

3.1.3 Référence 3. L'équipe est attractive par la reconnaissance de ses succès à des appels à projets compétitifs.

Auto-évaluation de l'équipe A3

Les membres de l'équipe participent à plusieurs projets nationaux. Ces projets représentent la source principale de financement des doctorants et post-doctorants.

Durant cette période, l'équipe a eu deux contrats campus France, un PROFAS et une bourse d'excellence de l'ambassade de France en Iran qui a permis le co-encadrement d'un doctorant. Elle a participé à deux projets Européen, un projet CHIST-ERA (Adalab) et un ERASMUS K107 pour un montant total de 243k€, un Labcom ANR IRISER (363 k€). Les membres de l'équipe ont participé à un contrat avec un laboratoire public du ministère de la défense pour un montant de 150k€ et un projet AMIES (99k€), un projet France Relance ACOMI (122k€), et un projet DIM4IDF (122k€), Un projet IUH, dont équipe est partenaire, est accepté en mai 2023. La source principale de financements des activités de l'équipe vient des projets de partenariats avec l'environnement socio-économique. Les membres de l'équipe ont participé à 19 contrats durant la période de référence pour un montant total de 2,13 millions d'euros. Il s'agit des contrats CIFRE MINDLYTIX 6k€, EDF-2020 (45k€), ALTER WAY (15k€), NUKKAI(20k€), SYNALTIC (28k€), RENAULT CIFRE (45k€), RENAULT PostDoc (56k€), BPCE (79k€), SAFRAN (57k€), REGIENOV (45k€), EDF (45k€), FaceBook (120k€), Deep Knowledge (120k€), SAFRAN (69k€) et Novedia (30k€). Elle a été sollicité pour une prestations de service pour un montant par REGIENOV pour un montant de 126k€ puis pour un montant de 122k€. Elle a, également, bénéficié d'un projet de maturation d'un montant de 470 k€. Le montant total des financements obtenus par l'équipe durant cette période d'évaluation avoisine les 3 millions d'euros.

3.2 Domaine 3. Production scientifique

Le tableau suivant résume la production scientifique de l'équipe durant la période.

| | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | Total |
|-----------------------------|------|------|------|------|------|------|------|-------|
| Livres | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| Chapitres de livre | 2 | 0 | 2 | 0 | 1 | 2 | 0 | 7 |
| Revue internationale | 3 | 3 | 9 | 1 | 8 | 4 | 1 | 29 |
| Autres revues | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 5 |
| Conférences internationales | 20 | 25 | 17 | 13 | 19 | 17 | 1 | 112 |
| Autres communications | 3 | 4 | 7 | 8 | 5 | 4 | 1 | 32 |
| Total | 29 | 32 | 38 | 22 | 33 | 29 | 3 | 186 |

3.2.1 Référence 1. La production scientifique de l'équipe satisfait à des critères de qualité.

La production scientifique est importante et de très grande qualité. Elle repose sur des fondements théoriques et méthodologiques solides et s'appuie sur des applications réelles nombreuses et de qualité. L'équipe travaille sur diverses thématiques allant de la programmation logique inductive, aux modèles théoriques d'apprentissage collaboratif et par transfert et aux mécanismes qui permettent d'apprendre pour apprendre. Ces thématiques, tout en utilisant les dernières avancées en apprentissage automatique, restent originales et constituent des orientations différenciantes au niveau régional et national.

Plusieurs membres de l'équipe sont impliqués dans des responsabilités administratives et d'enseignement chronophages qui limitent considérablement leurs implications en recherche durant de nombreuses années. Durant la période les membres de l'équipe ont pu publier plus d'une centaine d'articles dans des conférences internationales dont un article à IJCAI, deux articles AAAI, 2 articles ECML/PKDD, 2 articles PAKDD et une trentaine d'articles dans les revues internationales dont un article dans machine learning journal, un article dans transactions of the American Mathematical Society, un article dans les Proceedings National Academy of Sciences (USA) et un article dans Neurocomputing.

L'accent est moins mis sur les revues.

Les réalisations logicielles sont nombreuses, mais souvent mal mises en valeur. L'équipe gagnerait à mettre ses productions logicielles et ses démonstrateurs dans une plateforme visible et accessible. L'absence de personnel dédié dans l'équipe explique en partie cette situation.

Les diverses participations aux projets nationaux démontrent la vigueur et l'activité intense des membres de l'équipe. Cependant, il y a encore peu de participations aux projets internationaux.

Auto-évaluation de l'équipe A3

3.2.2 Référence 2. La production scientifique de l'équipe est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels.

Depuis la dernière évaluation, plusieurs membres sont partie en mutation ou en promotion dans d'autres universités fragilisant l'équipe et son équilibre. Dans ce contexte, l'arrivée de nouveaux collègues dynamiserait le potentiel de l'équipe. Il n'y a pas eu de recrutement sur la période, mais, un MCF est en cours de recrutement. Il est à noter que le recrutement de profils répondant à nos besoins est devenu très difficile à cause de la concurrence avec le secteur privé qui offre de meilleures conditions de travail et de meilleures opportunités.

L'accueil de doctorant est très actif et s'améliore en qualité et en production scientifique.

3.2 Domaine 3. Production scientifique

L'équivalent temps plein travaillé (ETP) de l'équipe est passé de 8,33 en 2017 à 6,83 en 2022. La moyenne annuelle sur la période est de 7,80. La production annuelle s'établit de ce fait à $26/46,03=0,56$ revues par personne et par an et $111/46,03=2,4$ conférences internationales par personne et par an. La production totale, incluant chapitres de livres et autres publications s'établit à $168/46,03=3,65$ par personne et par an. Cette situation cache cependant, des disparités importantes entre membres de l'équipe. Il est à noter que les doctorants participent à la production scientifique.

Les restructurations internes de l'équipe n'ont pas permis l'établissement de plans d'accompagnement des enseignants chercheurs qui souhaitent reprendre leur activité de recherche. Il s'agit d'un challenge à relever dans les prochaines années. Cet accompagnement peut se faire, en définissant des projets fédérateurs intégrant les compétences des collègues souhaitant reprendre l'activité de recherche à la fois dans le montage des projets et l'encadrement de doctorants.

3.2.3 Publications

Les références

- comportant « * » sont acceptées et à paraître.
- indiquées sur fond gris sont les productions phare de chaque catégorie.
- en gras constituent les 20% les meilleures de la production scientifique de chaque catégorie.

Signification des soulignements :

- *simple* : enseignant-chercheur ou chercheur titulaire,
- *double* : doctorant,
- *pointillé* : post-doc, ingénieur ou autre situation.

Articles dans des revues avec comité de lecture et chapitres de livres (ACL)

Articles dans des revues internationales avec comité de lecture sélectif

- 2023** [RI-29] M. O. Attaoui, N. Dif, H. Azzag, M. Lebbah et K. Nabil. Regions of interest selection in histopathological images using subspace and multi-objective stream clustering. *The Visual Computer*, 39 :1683-1701, 2023.
- 2022** [RI-28] K. Benlamine, Y. Bennani, B. Matej, N. Grozavu et I. Falih. Collaborative Learning to Improve the Non-uniqueness of NMF. *International Journal of Computational Intelligence and Applications*, 21(1) :2250001 :1-24, 2022.
- [RI-27] N. Dif, M. O. Attaoui, Z. Elberrichi, M. Lebbah et H. Azzag. Transfer learning from synthetic labels for histopathological images classification. *Applied Intelligence*, 52(1) :358-377, 2022.
- [**RI-26**] M. El Hamri, Y. Bennani et I. Falih. Hierarchical Optimal Transport for Unsupervised Domain Adaptation. *Machine Learning*, 111(11) :4159-4182, 2022.
- [**RI-25**] E. Goffinet, M. Lebbah, H. Azzag, G. Loïc et A. Coutant. Functional non-parametric latent block model : A multivariate time series clustering approach for autonomous driving validation. *Computational Statistics & Data Analysis*, 176 :107565 (23 pages), 2022.
- 2021** [RI-24] P. Alizadeh, A. Osmani, M. Essaid, A. Chibani et Y. Amirat. Reinforcement Learning for Interactive QoS-Aware Services Composition. *IEEE Systems Journal*, 15(1) :1098-1108, 2021.
- [RI-23] P. Alizadeh, E. Traversi et A. Osmani. Deterministic policies based on maximum regrets in MDPs with imprecise rewards. *AI Communications*, 34(4) :229-244, 2021.
- [RI-22] M. O. Attaoui, H. Azzag, M. Lebbah et N. Keskes. Subspace data stream clustering with global and local weighting models. *Neural Computing and Applications*, 33(8) :3691-3712, 2021.
- [RI-21] F. E. Ben Bouazza, Y. Bennani, G. Cabanes et A. Touzani. Unsupervised collaborative learning based on Optimal Transport theory. *Journal of Intelligent Systems*, 30(1) :698-719, 2021.

3.2 Domaine 3. Production scientifique

- [RI-20] F. Forest, M. Lebbah, H. Azzag et J. Lacaille. Deep embedded self-organizing maps for joint representation learning and topology-preserving clustering. *Neural Computing and Applications*, 33(24) :17439–17469, 2021.
- [RI-19] M. Hamidi et A. Osmani. Domain models for data sources integration in HAR. *Neuro-computing*, 444 :244–259, 2021.
- [RI-18] M. Hamidi et A. Osmani. Human Activity Recognition : A Dynamic Inductive Bias Selection Perspective. *Sensors*, 21(21) :7278 (42 pages), 2021.
- [RI-17] S. Zouinina, Y. Bennani, N. Rogovschi et A. Lyhyaoui. Data Anonymization through Collaborative Multi-view Microaggregation. *Journal of Intelligent Systems*, 30(1) :327–345, 2021.
- 2020** [RI-16] Z. C. Dagdia, C. Zarges, G. Beck et M. Lebbah. A Scalable and Effective Rough Set Theory based Approach for Big Data Pre-processing. *Knowledge and Information Systems*, 62 :3321–3386, 2020.
- 2019** [RI-15] E. Agora, J. Aantezana, C. Cabrelli et B. Matei. Existence of quasicrystals and universal stable sampling and interpolation in LCA groups. *Transactions of the American Mathematical Society*, 372(7) :4647–4674, 2019.
- [RI-14] M. Atzmueller, H. Soldano, G. Santini et D. Bouthinon. MinerLSD : efficient mining of local patterns on attributed networks. *Applied Network Science*, 4 :Article 43 (33 pages), 2019.
- [RI-13] G. Beck, T. Duong, M. Lebbah, H. Azzag et C. Cérin. A distributed approximate nearest neighbors algorithm for efficient large scale mean shift clustering. *Journal of Parallel and Distributed Computing*, 134 :128–139, 2019.
- [RI-12] B. Chien, A. Mokraoui et B. Matei. Contrast enhancement and details preservation of tone mapped high dynamic range images. *Journal of Visual Communication and Image Representation*, 58 :589–599, 2019.
- [RI-11] A. Coutant, K. Roper, D. Trejo-Banos, D. Bouthinon, M. Carpenter, J. Grzebyta, G. Santini, H. Soldano, M. Elati, J. Ramon, C. Rouveirol, L. N. Soldatova et R. D. King. Closed-loop cycles of experiment design, execution, and learning accelerate systems biology model development in yeast. *The Proceedings of the National Academy of Sciences (PNAS)*, 116(36) :18142–181, 2019.
- [RI-10] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largerone et A. Sala. Feature-rich networks : going beyond complex network topologies. *Applied Network Science*, 4 :Article 4 (13 pages), 2019.
- [RI-9] P. Rastin, G. Cabanes, B. Matei, Y. Bennani et J.-M. Marty. A new sparse representation learning of complex data : Application to dynamic clustering of web navigation. *Pattern Recognition*, 91 :291–307, 2019.
- [RI-8] P. Rastin, B. Matei, G. Cabanes, N. Grozavu et Y. Bennani. Impact of Learners' Quality and Diversity in Collaborative Clustering. *Journal of Artificial Intelligence and Soft Computing Research*, 9(2) :149–165, 2019.
- [RI-7] H. Soldano, G. Santini, D. Bouthinon, S. Bary et E. Lazega. Bi-pattern mining of attributed networks. *Applied Network Science*, 4 :Article 37 (26 pages), 2019.
- 2018** [RI-6] A. Cornuéjols, C. Wemmert, P. Gançarski et Y. Bennani. Collaborative clustering : Why, when, what and how. *Information Fusion*, 39 :81–95, 2018.
- [RI-5] H. Léger, D. Bouthinon, M. Lebbah et H. Azzag. An Instance Based Model for Scalable θ -Subsumption. *International Journal on Artificial Intelligence Tools*, 26(7) :1860011 (13 pages), 2018.
- [RI-4] L. Picchi, G. Cabanes, C. Ricci-Bonot et M. Cristina. Quantitative Matching of Clutch Size in Reciprocating Hermaphroditic Worms. *Current Biology*, 28(20) :3254–3259 E3, 2018.
- 2017** [RI-3] Y. Chevaleyre, U. Endriss et N. Maudet. Distributed Fair Allocation of Indivisible Goods. *Artificial Intelligence*, 242 :1–22, 2017.
- [RI-2] M. Ghesmoune, H. Azzag, S. Benbernou, M. Lebbah, T. Duong et M. Ouziri. Big Data : From Collection to Visualization. *Machine Learning*, 106 :837–862, 2017.
- [RI-1] J. Sublime, B. Matei, G. Cabanes, N. Grozavu, Y. Bennani et A. Cornuéjols. Entropy based probabilistic collaborative clustering. *Pattern Recognition*, 72 :144 – 157, 2017.

3.2 Domaine 3. Production scientifique

Articles dans des revues nationales ou internationales

- 2022** [RE-5] M. Chevallier, N. Rogovschi, F. Boufarès et N. Grozavu. Detecting near duplicate dataset with machine learning. *International Journal of Computer Information Systems and Industrial Management Applications*, 14(32) :12, 2022.
- [RE-4] A. F. Dos Reis, Y. Medjahdi, B. Sens, J. Sublime, G. Brante et F. Bader. Low Complexity LSTM-NN-Based Receiver for Vehicular Communications in the Presence of High-Power Amplifier Distortions. *IEEE Access*, 10 :121985–122000, 2022.
- 2019** [RE-3] S. Oujdi, H. Belbachir et F. Boufarès. C4.5 Decision Tree Algorithm for Spatial Data, Alternatives and Performances. *Journal of Computing and Information Technology*, 27(3) :29–43, 2019.
- [RE-2] J. Sublime, G. Cabanes et B. Matei. Study on the Influence of Diversity and Quality in Entropy Based Collaborative Clustering. *Entropy*, 21(10) :1–21, 2019.
- 2017** [RE-1] A. Benyettou, Y. Bannani, A. Benyettou, A. Bendahmane et G. Cabanes. Semi-Supervised Multi-Label Classification Through Topological Active Learning. *International Journal on Communications Antenna and Propagation (IRECAP)*, 7(3) :222–232, 2017.

Chapitres de livre

- 2022** [CL-7] N. Grozavu, Y. Bannani, B. Matei et K. Benlamine. Multi-view clustering based on non-negative matrix factorization, Pedrycz, éditeur, *Recent Advancements in Multi-View Data Analytics*, volume 106 de *Studies in Big Data*, pages 177–200. Springer, 2022.
- [CL-6] P.-A. Murena, J. Sublime et B. Matei. Rethinking Collaborative Clustering : A Practical and Theoretical Study Within the Realm of Multi-view Clustering, *Recent Advancements in Multi-View Data Analytics*, volume 106 de *Studies in Big Data*, pages 97–130. Witold Pedrycz, springer édition, 2022.
- 2021** [CL-5] M. Hamidi et A. Osmani. Combining Domain Knowledge and Deep Learning to Improve HAR Models, Z. Chen, M. Wu et X. Li, éditeurs, *Generalization With Deep Learning : For Improvement On Sensing Capability*, pages 55–77. World Scientific, 2021.
- 2019** [CL-4] S. Benbernou, M. Bentounsi, C. Cérin, M. Ouziri, P. Garteiser, L. Abidi, M. Lebbah, H. Azzag et M. Smadja. A Big Data Platform for Enhancing Life Imaging Activities, J. Darmont et S. Loudcher, éditeurs, *Utilizing Big Data Paradigms for Business Intelligence*, pages 39–71. IGI Global, 2019.
- [CL-3] A. Osmani et M. Hamidi. Bayesian optimization of neural architectures for human activity recognition, N. Kawaguchi, N. Nishio, D. Roggen, S. Inoue, S. Pirttikangas et K. Van Laerhoven, éditeurs, *Human Activity Sensing*, pages 171–195. Springer, 2019.
- 2017** [CL-2] M. Pujari et R. Kanawati. Link prediction in Large-scale multiplex networks, S. Cordier, N. Debarsy, C. Ertur, F. Nemo, D. Nourrit-Lucas, G. Poisson et C. Vrain, éditeurs, *Understanding Interactions in complex systems*, Chapitre 7, pages 107–136. Cambridge Scholar Publishing, 2017.
- [CL-1] H. Soldano, G. Santini et D. Bouthinon. Formal Concept Analysis of Attributed Networks, R. Missaoui, S. Obiedkov et S. Kuznetsov, éditeurs, *Formal Concept Analysis of Social Networks*, pages 143–170. Lecture Notes in Social Network. Springer, 2017.

Tutoriaux dans des conférences internationales

- 2023** [TU-3] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. Multi-Modal Evaluation Approach for Medical Image Segmentation. *arXiv*, 2023.
- 2017** [TU-2] N. Grozavu et R. Kanawati. Topological & Graph-based Clustering : Recent Algorithmic Advances : A tutorial. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, Alaska, USA, May 14-19, 2017, pages 120–120, 2017.
- [TU-1] R. Kanawati et M. Atzmueller. Mining Attributed Networks : A Tutorial. *The 4th IEEE International conference on Data Science and Advanced Analytics (DSAA 2017)*, Tokyo, Japon, October 17-21, 2017, 2017.

Communications avec actes (ACT)

Communications dans des conférences internationales avec comité de lecture

- 2023** [CI-112] T. Viard, H. Soldano et G. Santini. Exploring and mining attributed sequences of interactions. *The 11th International Conference on Complex Networks and their Applications (COMPLEX NETWORKS XI), Palermo, Italy, November 8-10, 2022*, volume 1078 de *Studies in Computational Intelligence*, pages 537–549, 2023.
- 2022** [CI-111] P. Alizadeh, A. Osmani et S. Taleb. Improving Imitation Learning by Merging Experts Trajectories. *The 31st ACM International Conference on Information & Knowledge Management (CIKM'22), Atlanta, GA, USA, October 17-21, 2022*, pages 3796–3800, 2022.
- [CI-110] M. Chevallier, N. Grozavu, F. Boufarès et N. Rogovschi. Detecting Near Duplicate Dataset. *The 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022), Seattle, WA, United States, December 14-16, 2022*, volume 417 de *Lecture Notes in Networks and Systems*, pages 394–403, 2022.
- [CI-109] M. Chevallier, N. Grozavu, N. Rogovschi, F. Boufarès et C. Clairmont. Trade Between Population Size and Mutation Rate for GAAM (Genetic Algorithm with Aggressive Mutation) for Feature Selection. *The 18th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022), Hersonissos, Crete, Greece, June 17-20, 2022*, volume 646 de *IFIP Advances in Information and Communication Technology*, pages 432–444, 2022.
- [CI-108] M. El Hamri, Y. Bennani et I. Falih. Incremental Unsupervised Domain Adaptation Through Optimal Transport. *The 2022 International Joint Conference on Neural Networks (IJCNN 2022) Padua, Italy, July 18-23, 2022*, pages 1–8, 2022.
- [CI-107] M. El Hamri, Y. Bennani et I. Falih. When Domain Adaptation Meets Semi-supervised Learning Through Optimal Transport. *The 18th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022), Hersonissos, Crete, Greece, June 17-20, 2022*, volume 646 de *IFIP Advances in Information and Communication Technology*, pages 58–69, 2022.
- [CI-106] W. Fogos da Rocha, H. Azzag, M. Lebbah et A. Mokraoui. Super-Resolution GAN Improving YOLO's Performance Benchmark. *The 21st IEEE International Conference on Machine Learning and Applications (ICMLA 2022), Nassau, The Bahamas, December 12-14, 2022*, pages 1–8, 2022.
- [CI-105] M. Hamidi et A. Osmani. Context Abstraction to Improve Decentralized Machine Learning in Structured Sensing Environments. *The 2022 Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2022), Grenoble, France, September 19-23, 2022*, volume 13715 de *Lecture Notes in Artificial Intelligence*, pages 647–663, 2022.
- [CI-104] M. Hamidi, A. Osmani, L. Rasaq, G. Dogan et N. Alotaibi. Multiview Representation Learning for Human Activity Recognition. *The 9th IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA 2022), Chemnitz, Germany, June 15-17, 2022*, pages 1–5, 2022.
- [CI-103] Y. Khalafaoui, N. Grozavu, B. Matei et L.-W. Goix. Multi-modal Multi-view Clustering based on Non-negative Matrix Factorization. *The 2022 IEEE Symposium Series on Computational Intelligence (SSCI 2022), Singapore, Singapore, December 4-7, 2022*, pages 1386–1391, 2022.
- [CI-102] J.-V. Loddo et R. Kanawati. Mariotel : A Web-Based Virtual Remote Computer Science Lab. *The 20th International Conference on Information Technology Based Higher Education and Training (ITHET'22), Antalya, Turkey, November 7-9, 2022*, pages 1–6, 2022.
- [CI-101] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. Evaluation of Early Diagnosis of COVID-19 Algorithms. *The 34th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2022), Macau, China, October 31-November, 2022*, pages 1–8, 2022.
- [CI-100] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. Multimodal Evaluation Method for Sound Event Detection. *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022), Singapore, 23-27 May 2022*, pages 1026–1030, 2022.

3.2 Domaine 3. Production scientifique

- [CI-99] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. Uniform Evaluation of Properties in Activity Recognition. *The 26th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2022), Chengdu, China, May 16-19, 2022*, volume 13281 de *Lecture Notes in Computer Science*, pages 83–95, 2022.
- [CI-98] A. Osmani et M. Hamidi. Reduction of the Position Bias via Multi-level Learning for Activity Recognition. *The 26th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2022), Chengdu, China, May 16-19, 2022*, volume 13281 de *Lecture Notes in Computer Science*, pages 289–302, 2022.
- [CI-97] A. Osmani, M. Hamidi et P. Alizadeh. Clustering Approach to Solve Hierarchical Classification Problem Complexity. *The 36th AAAI Conference on Artificial Intelligence (AAAI 2022), Virtual Conference, February 22-March 1, 2022*, volume 36(7) de *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7904–7912, 2022.
- [CI-96] A. Zaiou, Y. Bennani, M. Hibti et B. Matei. Quantum Approach for Vertex Separator Problem in Directed Graphs. *The 18th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022), Hersonissos, Crete, Greece, June 17-20, 2022*, volume 646 de *IFIP Advances in Information and Communication Technology*, pages 495–506, 2022.
- [CI-95] A. Zaiou, Y. Bennani, B. Matei et M. Hibti. A quantum learning approach based on Hidden Markov Models for failure scenarios generation. *The 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML 2022), Hangzhou, China, March 25-27, 2022*, pages 62–67, 2022.
- 2021 [CI-94] D. F. Andriantsiory, J. Ben Geloun et M. Lebbah. Multi-Slice Clustering for 3-order Tensor Data. *The 20th IEEE International Conference on Machine Learning and Applications (ICMLA 2021), Pasadena, CA, USA, December 13-16, 2021*, pages 173–178, 2021.
- [CI-93] M. O. Attaoui, H. Azzag, N. Keskes et M. Lebbah. A New Subspace Multi-Objective Approach for the Clustering and Selection of Regions of Interests in Histopathological Images. *The 2021 IEEE Congress on Evolutionary Computation (CEC 2021), Kraków, Poland, June 28 - July 1, 2021*, pages 556–563, 2021.
- [CI-92] G. Beck, M. Lebbah, H. Azzag et T. Duong. A New Nearest Neighbor Median Shift Clustering for Binary Data. *The 30th International Conference on Artificial Neural Networks (ICANN 2021), Bratislava, Slovakia, September 14-17, 2021*, volume 12895 de *Lecture Notes in Computer Science*, pages 101–112, 2021.
- [CI-91] M. Chevallier, F. Boufarès, N. Grozavu et N. Rogovschi. Near duplicate Column identification : a machine learning approach. *The 2021 IEEE Symposium Series on Computational Intelligence (SSCI 2021), Orlando, FL, USA, December 5-7, 2021*, pages 1–7, 2021.
- [CI-90] M. El Hamri, Y. Bennani et I. Falih. Inductive Semi-supervised Learning Through Optimal Transport. *The 28th International Conference on Neural Information Processing (ICONIP 2021), Sanur, Bali, Indonesia, December 8-12, 2021*, volume 1516 de *Communications in Computer and Information Science*, pages 668–675, 2021.
- [CI-89] M. El Hamri, Y. Bennani et I. Falih. Label Propagation Through Optimal Transport. *The 2021 International Joint Conference on Neural Networks (IJCNN 2021), Shenzhen, China, July 18-22, 2021*, pages 1–8, 2021.
- [CI-88] Y. Foucade, Y. Bennani et Z. Aabbou. Collaborative Random Forests Learning. *The 2021 IEEE 23rd International Conference on High Performance Computing & Communications; 7th International Conference on Data Science & Systems; 19th International Conference on Smart City; 7th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys 2021), Hainan, Hainan, China, December 20-22, 2021*, pages 1147–1154, 2021.
- [CI-87] E. Goffinet, M. Lebbah, H. Azzag, L. Giraldi et A. Coutant. Multivariate Time Series Multi-Cocustering. application to advanced driving assistance system validation. *The 9th European Symposium on Artificial Neural Networks (ESANN 2021), Bruges, Belgium, October 6-8, 2021*, pages 463–468, 2021.
- [CI-86] M. Hamidi et A. Osmani. Data generation process modeling for activity recognition. *The 2020 Joint European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2020), Ghent, Belgium, Septemebr 14-18, 2021*, volume 12460 de *Lecture Notes in Artificial Intelligence*, pages 374–390. Springer, 2021.

- [CI-85] M. Kaddouh, G. Cabanes et Y. Bennani. Incorporating Neighborhood Information During NMF Learning. *The 28th International Conference on Neural Information Processing (ICONIP 2021), Sanur, Bali, Indonesia, December 8-12, 2021*, volume 1516 de *Communications in Computer and Information Science*, pages 591–598, 2021.
- [CI-84] P. Le Jeune, M. Lebbah, A. Mokraoui et H. Azzag. Experience feedback using Representation Learning for Few-Shot Object Detection on Aerial Images. *20th IEEE International Conference on Machine Learning and Applications (ICMLA 2021), Pasadena, CA, USA, December 13-16, 2021*, pages 662–667, 2021.
- [CI-83] T. Mecharnia, N. Pernelle, C. Rouveirol, F. Hamdi et L. Chibout. Mining Contextual Rules to Predict Asbestos in Buildings. *The 26th International Conference on Conceptual Structures (ICCS 2021), Berlin, Germany, September 20-22, 2021*, volume 12879 de *Lecture Notes in Artificial Intelligence*, pages 170–184, 2021.
- [CI-82] A. Osmani, P. Alizadeh et C. Rodrigues. Equational Model Guided by Real-time Sensor Data to Monitor Industrial Robots. *17th IEEE International Conference on Automation Science and Engineering, CASE 2021, Lyon, France, August 23-27, 2021*, pages 1848–1854, 2021.
- [CI-81] A. Osmani et M. Hamidi. Hierarchical Learning of Dependent Concepts for Human Activity Recognition. *The 25th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2021), Virtual Event, May 11-14, 2021*, volume 12713 de *Lecture Notes in Artificial Intelligence*, pages 79–92, 2021.
- [CI-80] A. Osmani, M. Hamidi et S. Bouhouche. Augmented Experiment in Material Engineering Using Machine Learning. *The 35th AAAI Conference on Artificial Intelligence (AAAI 2021), Virtual Event, February 2-9, 2021*, volume 35 de 10, pages 9251–9258, 2021.
- [CI-79] G. Santini, H. Soldano et S. Zevio. Experiments on F-Restricted Bi-pattern Mining. *The 10th International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS X), Madrid, Spain, November 30-December 2, 2021*, volume 1015 de *Studies in Computational Intelligence*, pages 539–550, 2021.
- [CI-78] G. Santini, H. Soldano et S. Zevio. Restricted Bi-pattern Mining of Attributed Networks. *The 26th International Conference on Conceptual Structures (ICCS 2021), Virtual Event, September 20-22, 2021*, volume 12879 de *Lecture Notes in Artificial Intelligence*, pages 200–207, 2021.
- [CI-77] A. Zaiou, Y. Bennani, B. Matei et M. Hibti. Balanced K-means using Quantum annealing. *The 2021 IEEE Symposium Series on Computational Intelligence (SSCI 2021), Orlando, FL, USA, December 5-7, 2021*, pages 1–7, 2021.
- [CI-76] A. Zaiou, B. Matei, Y. Bennani et M. Hibti. Convex Non-negative Matrix Factorization Through Quantum Annealing. *The 23rd IEEE International Conference on High Performance Computing & Communications; 7th International Conference on Data Science & Systems; 19th International Conference on Smart City; 7th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys 2021), Haikou, Hainan, China, December 20-22, 2021*, pages 1253–1258, 2021.
- 2020 [CI-75] F. E. Ben Bouazza, Y. Bennani, G. Cabanes et A. Touzani. Collaborative Clustering Through Optimal Transport. *29th International Conference on Artificial Neural Networks (ICANN 2020), Bratislava, Slovakia, September 15-18, 2020*, volume 12397 de *Lecture Notes in Computer Science*, pages 873–885, 2020.
- [CI-74] F. E. Ben Bouazza, Y. Bennani, A. Touzani et G. Cabanes. Subspace Guided Collaborative Clustering Based on Optimal Transport. *The 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020), Hyderabad, India, December 15-18, 2020*, volume 1383 de *Advances in Intelligent Systems and Computing*, pages 113–124. Springer, 2020.
- [CI-73] K. Benlamine, Y. Bennani, N. Grozavu et B. Matei. Quantum Collaborative K-means. *The 2020 International Joint Conference on Neural Networks (IJCNN 2020), Glasgow, United Kingdom, July 19-24, 2020*, pages 1–7, 2020.
- [CI-72] K. Benlamine, Y. Bennani, B. Matei et N. Grozavu. Quantum Semi Non-negative Matrix Factorization. *The 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020), Hyderabad, India, December 15–18, 2020*, volume 1383 de *Advances in Intelligent Systems and Computing*, pages 135–144, 2020.

3.2 Domaine 3. Production scientifique

- [CI-71] M. Chevallier, N. Rogovschi, N. Grozavu et F. Boufarès. Seeding Initial Population, in Genetic Algorithm for Features Selection. *The 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020)*, Online, India, December 15–18, 2020, volume 1383 de *Advances in Intelligent Systems and Computing*, pages 572–582, 2020.
- [CI-70] G. Dupont, E. Kalinicheva, J. Sublime, F. Rossant et M. Pâques. Unsupervised Change Detection Using Joint Autoencoders for Age-Related Macular Degeneration Progression. *The 29th International Conference on Artificial Neural Networks (ICANN 2020)*, Bratislava, Slovakia, September 15-18, 2020, volume 12397 de *Lecture Notes in Computer Science*, pages 813–824, 2020.
- [CI-69] F. Forest, Q. Cochard, C. Noyer, A. Cadut, M. Joncour, J. Lacaille, M. Lebbah et H. Azzag. Large-scale Vibration Monitoring of Aircraft Engines from Operational Data using Self-organized Models. *The 14th Annual Conference of the Prognostics and Health Management Society (PHM22)*, Nashville, Tennessee, November 1-4, 2022, pages 1–6, 2020.
- [CI-68] F. Forest, A. Mourer, M. Lebbah, H. Azzag et J. Lacaille. An Invariance-guided Stability Criterion for Time Series Clustering Validation. *The 25th International Conference on Pattern Recognition (ICPR 2020)*, Virtual Event, Milan, Italy, January 10-15, 2021, pages 9296–9303, 2020.
- [CI-67] Y. Foucade et Y. Bennani. Unsupervised Learning from Data and Learners. *The 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020)*, Hyderabad, India, December 15-18, 2020, volume 1383 de *Advances in Intelligent Systems and Computing*, pages 496–507, 2020.
- [CI-66] E. Goffinet, M. Lebbah, H. Azzag et L. Giraldi. Autonomous Driving Validation with Model-Based Dictionary Clustering. *The 2020 European Conference Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2020)*, Ghent, Belgium, September 14-18, 2020, volume 12460 de *Lecture Notes in Computer Science*, pages 323–338, 2020.
- [CI-65] M. Hamidi, A. Osmani et P. Alizadeh. A Multi-View Architecture for the SHL Challenge. *The 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20)*, Virtual Event, Mexico, September 12-17, 2020, pages 317–322, 2020.
- [CI-64] J. Zamora et J. Sublime. A New Information Theory Based Clustering Fusion Method for Multi-view Representations of Text Documents. *The 12th International Conference on Social Computing and Social Media (SCSM 2020) held as Part of the 22nd HCI International Conference (HCI 2020)*, Copenhagen, Denmark, July 19-24, 2020, volume 12194 de *Lecture Notes in Computer Science*, pages 156–167, 2020.
- [CI-63] S. Zouinina, Y. Bennani, N. Rogovschi et A. Lyhyaoui. A Two-Levels Data Anonymization Approach. *The 16th IFIP WG 12.5 International Conference on The Artificial Intelligence Applications and Innovations (AIAI 2020)*, Neos Marmaras, Greece, June 5-7, 2020, *Proceedings, Part I*, volume 583 de *IFIP Advances in Information and Communication Technology*, pages 85–95, 2020.
- 2019** [CI-62] M. O. Attaoui, M. Lebbah, N. Keskes, H. Azzag et M. Ghesmoune. Soft Subspace Growing Neural Gas for Data Stream Clustering. *The 28th International Conference on Artificial Neural Networks (ICANN 2019)*, Munich, Germany, September 17-19, 2019, pages 569–580, 2019.
- [CI-61] F. E. Ben Bouazza, Y. Bennani, M. El, G. Cabanes, B. Matei et A. Touzani. Multi-view Clustering through Optimal Transport. *The 26th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Society (ICONIP 2019) Sydney, Australia, December 12-15 2019*, volume 15 de *Australian Journal of Intelligent Information Processing Systems*, pages 1–9, 2019.
- [CI-60] K. Benlamine, Y. Bennani, A. Zaiou, M. Hibti, B. Matei et N. Grozavu. Distance Estimation for Quantum Prototypes Based Clustering. *The 26th International Conference on Neural Information Processing (ICONIP 2019)*, Sydney, NSW, Australia, December 12-15, 2019, volume 11955 de *Lecture Notes in Computer Science*, pages 561–572, 2019.
- [CI-59] K. Benlamine, N. Grozavu, Y. Bennani et B. Matei. Collaborative Non-negative Matrix Factorization. *The 28th International Conference on Artificial Neural Networks (ICANN 2019)*, Munich, Germany, September 17-19, 2019, volume 11730 de *Lecture Notes in Computer Science*, pages 655–666, 2019.

- [CI-58] I. Brahmi, G. Cabanes, Y. Bennani et A. Touzani. Deepness Analysis of Learned Factors in Multilayer NMF. *The 26th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Society (ICONIP 2019) Sydney, Australia, December 12-15 2019*, volume 16 de *Australian Journal of Intelligent Information Processing Systems*, pages 24–31, 2019.
- [CI-57] F. Forest, M. Lebbah, H. Azzag et J. Lacaille. Deep Embedded SOM : joint representation learning and self-organization. *The 27th European Symposium on Artificial Neural Networks (ESANN 2019), Bruges, Belgium, April 24-26, 2019*, pages 1–6, 2019.
- [CI-56] E. Kalinicheva, J. Sublime et M. Trocan. Change Detection in Satellite Images Using Reconstruction Errors of Joint Autoencoders. *The 28th International Conference on Artificial Neural Networks (ICANN 2019), Munich, Germany, September 17-19, 2019*, volume 11729 de *Lecture Notes in Computer Science*, pages 637–648, 2019.
- [CI-55] E. Kalinicheva, J. Sublime et M. Trocan. Object-Based Change Detection in Satellite Images Combined with Neural Network Autoencoder Feature Extraction. *The 9th International Conference on Image Processing Theory, Tools and Applications (IPTA 2019), Istanbul, Turkey, November 6-9, 2019*, pages 1–6, 2019.
- [CI-54] R. Kanawati et M. Atzmueller. Modeling and Mining Feature-Rich Networks. *The 2019 World Wide Web Conference (WWW 2019), San Francisco, CA, USA, May 13-17, 2019*, pages 1306–1307, 2019.
- [CI-53] H. Kordestani, R. Mojarad, A. Chibani, A. Osmani, Y. Amirat, K. Barkaoui et W. Zahran. Hapicare : A Healthcare Monitoring System with Self-Adaptive Coaching using Probabilistic Reasoning. *The 16th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA 2019), Abu Dhabi, UAE, November 3-7, 2019*, pages 1–8, 2019.
- [CI-52] A. Osmani, M. Hamidi et S. Bouhouche. Monitoring of a Dynamic System Based on Autoencoders. *The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), Macao, China, August 10-16, 2019*, pages 1836–1843, 2019.
- [CI-51] P. Rastin, Y. Bennani et R. Verde. Automatic detection of the support points in relational clustering. *The 2019 International Joint Conference on Neural Networks (IJCNN 2019), Budapest, Hungary, July 14-19, 2019*, pages 1–8, 2019.
- [CI-50] P. Rastin, G. Cabanes, R. Verde, Y. Bennani et T. Couronne. Generative Histogram-Based Model Using Unsupervised Learning. *The 26th International Conference on Neural Information Processing (ICONIP 2019), Sydney, NSW, Australia, December 12-15, 2019*, volume 11955 de *Lecture Notes in Computer Science*, pages 634–646. Springer, 2019.
- [CI-49] N. Rogovschi, S. Zouinina, B. Matei, I. Falih, N. Grozavu et S. Ozawa. t-Distributed Stochastic Neighbor Embedding Spectral Clustering using higher order approximations. *The 26th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Society (ICONIP 2019) Sydney, Australia, December 12-15 2019*, volume 17 de *Australian Journal of Intelligent Information Processing Systems*, pages 78–86, 2019.
- [CI-48] H. Soldano, G. Santini et D. Bouthinon. Attributed Graph Pattern Set Selection Under a Distance Constraint. *The 8th International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS VIII), Lisbon, Portugal, December 10-12, 2019*, volume 882 de *Studies in Computational Intelligence*, pages 228–241, 2019.
- [CI-47] J. Sublime. Incremental Collaborative Clustering Using Information Theory and Information Compression. *The 5th International Conference on Fuzzy Systems and Data Mining (FSDM 2019), Kitakyushu City, Japan, October 18-21, 2019*, volume 320 de *Frontiers in Artificial Intelligence and Applications*, pages 457–464, 2019.
- [CI-46] S. Zouinina, Y. Bennani, M. Ben-Fares, N. Rogovschi et A. Lyhyaoui. Preserving Utility during Attribute-oriented Data Anonymization Process. *26th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Society (ICONIP 2019) Sydney, Australia, December 12-15 2019*, volume 16 de *Australian Journal of Intelligent Information Processing Systems*, pages 25–35, 2019.
- 2018 [CI-45] G. Beck, H. Azzag, S. Bougeard, N. Niang et M. Lebbah. A New Micro-Batch Approach for Partial Least Square Clusterwise Regression. *The 3rd INNS Conference on Big Data and Deep Learning (INNS BDDL 2018), Sanur, Bali, Indonesia, April 17-19, 2018*, volume 144 de *Procedia Computer Science*, pages 239–250, 2018.

3.2 Domaine 3. Production scientifique

- [CI-44] K. Benlamine, N. Grozavu, Y. Bennani, N. Rogovschi, K. Haddadou et A. Amamou. Domain Name Recommendation based on Neural Network. *The 2018 INNS Conference on Big Data and Deep Learning 2018, Sanur, Bali, Indonesia, 17-19 April 2018*, volume 144 de *Procedia Computer Science*, pages 60–70, 2018.
- [CI-43] S. Boulbazine, G. Cabanes, B. Matei et Y. Bennani. Online Semi-supervised Growing Neural Gas for Multi-label Data Classification. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018), Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.
- [CI-42] I. Brahmi, G. Cabanes, Y. Bennani et B. Matei. Learning Useful Representations Through Stacked Self-Organizing Maps. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018), Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.
- [CI-41] C. Cérin, T. Menouer et M. Lebbah. Accelerating the Computation of Multi-Objectives Scheduling Solutions for Cloud Computing. *The 8th IEEE International Symposium on Cloud and Service Computing (SC2 2018), Paris, France, November 18-21, 2018*, pages 49–56, 2018.
- [CI-40] Z. C. Dagdia, C. Zarges, G. Beck, H. Azzag et M. Lebbah. A Distributed Rough Set Theory Algorithm based on Locality Sensitive Hashing for an Efficient Big Data Pre-processing. *The IEEE International Conference on Big Data (Big Data 2018), Seattle, WA, USA, December 10-13, 2018*, pages 2597–2606, 2018.
- [CI-39] N.-Q. Doan, H. Azzag et M. Lebbah. Hierarchical Laplacian Score for unsupervised feature selection. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018), Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–7, 2018.
- [**CI-38**] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. Community detection in Attributed Network. *The 2018 Word Wide Web Conference (WWW 2018), Lyon, France, April 23-27, 2018*, pages 1299–1306, 2018.
- [CI-37] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. Topological Multi-View Clustering for Collaborative Filtering. *The 3rd INNS Conference on Big Data and Deep Learning 2018 (INNS BDDL 2018), Sanur, Bali, Indonesia, 17-19 April 2018*, volume 144 de *Procedia Computer Science*, pages 306–312, 2018.
- [CI-36] I. Falih, N. Grozavu, R. Kanawati, Y. Bennani et B. Matei. Collaborative Multi-View Attributed Networks Mining. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018), Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.
- [CI-35] F. Forest, J. Lacaille, M. Lebbah et H. Azzag. A Generic and Scalable Pipeline for Large-Scale Analytics of Continuous Aircraft Engine Data. *The 2018 IEEE International Conference on Big Data (Big Data 2018), Seattle, WA, USA, December 10-13, 2018*, pages 1918–1924, 2018.
- [CI-34] M. Ghesmoune, M. Lebbah, H. Azzag, S. Benbernou, M. Ouziri et T. Duong. A Complete Data Science Work-flow For Insurance Field. *The 2018 IEEE International Conference on Big Data (Big Data 2018), Seattle, WA, USA, December 10-13, 2018*, pages 1925–1930, 2018.
- [CI-33] E. Kalinicheva, J. Sublime et M. Trocan. Neural Network Autoencoder for Change Detection in Satellite Image Time Series. *The 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2018), Bordeaux, France, December 9-12, 2018*, pages 641–642, 2018.
- [CI-32] S. Legras, C. Rouveiro et V. Ventos. The Game of Bridge : A Challenge for ILP. *The 28th International Conference on Inductive Logic Programming (ILP 2018), Ferrara, Italy, September 2-4, 2018*, volume 11105 de *Lecture Notes in Artificial Intelligence*, pages 72–87, 2018.
- [**CI-31**] P.-A. Murena, J. Sublime, B. Matei et A. Cornuéjols. An Information Theory based Approach to Multisource Clustering. *The 27th International Joint Conference on Artificial Intelligence (IJCAI 2018), July 13-19, 2018, Stockholm, Sweden*, pages 2581–2587, 2018.
- [CI-30] L. H. Ngo, M. Luong, N. Metodiev, T. Le-Tien, S. Guérif et E. Viennet. Sparse Representation Wavelet Based Classification. *The 2018 IEEE International Conference on Image Processing (ICIP 2018), Athens, Greece, October 7-10, 2018*, pages 2974–2978, 2018.
- [CI-29] A. Osmani et M. Hamidi. Hybrid and convolutional neural networks for locomotion recognition. *The 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18), Singapore, October 9-11, 2018*, pages 1531–1540, 2018.

- [CI-28] P. Rastin, G. Cabanes, B. Matei et J.-M. Marty. Change Detection in Individual Users' Behavior. *The 27th International Conference on Artificial Neural Networks and Machine Learning (ICANN 2018)*, Rhodes, Greece, October 4-7, 2018, pages 501–510, 2018.
- [CI-27] H. Soldano, S. Bary, G. Santini et D. Bouthinon. Core Stratification of Two-Mode Networks. *The 7th International Conference on Complex Networks and their Applications (COMPLEX NETWORKS VII)*, Cambridge, UK, December 11-13, 2018, volume 812 de *Studies in Computational Intelligence*, pages 130–142, 2018.
- [CI-26] H. Soldano, G. Santini, D. Bouthinon, S. Bary et E. Lazega. Bi-Pattern Mining of Two Mode and Directed Networks. *The 2018 World Wide Web Conference 2018 (WWW'18)*, Lyon, France, April 23-27, 2018, pages 1287–1294, 2018.
- [CI-25] J. Sublime et S. Lefebvre. Collaborative Clustering through Constrained Networks using Bandit Optimization. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018)*, Rio de Janeiro, Brazil, July 8-13, 2018, pages 1–8, 2018.
- [CI-24] J. Sublime, D. Maurel, N. Grozavu, B. Matei et Y. Bennani. Optimizing exchange confidence during collaborative clustering. *The 2018 International Joint Conference on Neural Networks (IJCNN 2018)*, Rio de Janeiro, Brazil, July 8-13, 2018, pages 1–8, 2018.
- [CI-23] L.-M. Veillon, G. Bourgne et H. Soldano. Better Collective Learning with Consistency Guarantees. *The 21th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2018)*, Tokyo, Japan, October 29-November 2, 2018, volume 11224 de *Lecture Notes in Artificial Intelligence*, pages 671–679, 2018.
- [CI-22] S. Zouinina, N. Grozavu, Y. Bennani, A. Lyhyaoui et N. Rogovschi. A Topological k-Anonymity Model Based on Collaborative Multi-view Clustering. *The 27th International Conference on Artificial Neural Networks (ICANN 2018)*, Rhodes, Greece, October 4-7, 2018, volume 11141 de *Lecture Notes in Computer Science*, pages 817–827, 2018.
- [CI-21] S. Zouinina, N. Grozavu, Y. Bennani, A. Lyhyaoui et N. Rogovschi. Efficient k-Anonymization through Constrained Collaborative Clustering. *The 2018 IEEE Symposium Series on Computational Intelligence (SSCI 2018)*, Bangalore, India, November 18-21, 2018, pages 405–411, 2018.
- 2017 [CI-20] S. Bouveret, Y. Chevaleyre, F. Durand et J. Lang. Voting by sequential elimination with few voters. *The 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, Melbourne, Australia, August 19-25, 2017, pages 128–134, 2017.
- [CI-19] C. Cérin, J.-L. Gaudiot, M. Lebbah et F. Yuehgoh. Return of experience on the mean-shift clustering for heterogeneous architecture use case. *The 2017 IEEE International Conference on Big Data (BigData 2017)*, Boston, MA, USA, December 11-14, 2017, pages 3499–3507, 2017.
- [CI-18] Z. C. Dagdia, C. Zarges, G. Beck et M. Lebbah. A Distributed Rough Set Theory based Algorithm for an Efficient Big Data Pre-processing under the Spark Framework. *The 2017 IEEE International Conference on Big Data (Big Data 2017)*, Boston, MA, USA, December 11-14, 2017, pages 911–916, 2017.
- [CI-17] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. ANCA : Attributed Network Clustering Algorithm. *The 6th International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS VI)*, Lyon, France, November 29 - December 1, 2017, volume 689 de *Studies in Computational Intelligence*, pages 241–252, 2017.
- [CI-16] M. Hmimida et R. Kanawati. A graph-based meta-approach for tag recommendation. H. Cherifi, S. Gaito, W. Quattrociocchi et A. Sala, éditeurs, *The 5th International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS 2016)*, Milano, Italy, Novembre 21-15, 2016, volume 693 de *Studies in Computational Intelligence*, pages 309–320. Springer, 2017.
- [CI-15] A. Lachaud, N. Grozavu, B. Matei et Y. Bennani. Collaborative clustering between different topological partitions. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, AK, USA, May 14-19, 2017, pages 4111–4117. IEEE, 2017.
- [CI-14] C. Laclau, I. Redko, B. Matei, Y. Bennani et V. Brault. Co-clustering through Optimal Transport. *The 34th International Conference on Machine Learning (ICML 2017)*, Sydney, NSW, Australia, 6-11 August 2017, volume 70 de *Proceedings of Machine Learning Research*, pages 1955–1964, 2017.

3.2 Domaine 3. Production scientifique

- [CI-13] H. Léger, D. Bouthinon, M. Lebbah et H. Azzag. An Instance Based Model for Scalable θ -Subsumption. *The 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2017)*, Boston, MA, USA, November 6-8, 2017, pages 846–852, 2017.
- [CI-12] F. Lezzar, D. Benmerzoug, I. Kitouni et A. Osmani. A GRC-Centric Approach for Enhancing Management Process of IoT-Based Health Institution. *The 3rd International Conference on Mobile, Secure, and Programmable Networking (MSPN 2017)*, Paris, France, June 29-30, 2017, Revised Selected Papers, volume 10566 de *Lecture Notes in Computer Science*, pages 207–221, 2017.
- [CI-11] D. Maurel, J. Sublime et S. Lefebvre. Incremental Self-Organizing Maps for Collaborative Clustering. *The 24th International Conference on Neural Information Processing (ICONIP 2017)*, Guangzhou, China, November 14-18, 2017, volume 10634 de *Lecture Notes in Computer Science*, pages 497–504, 2017.
- [CI-10] A. Osmani, M. Hamidi et A. Chibani. Machine Learning Approach for Infant Cry Interpretation. *The 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2017)*, Boston, MA, USA, November 6-8, 2017, pages 182–186, 2017.
- [CI-9] A. Osmani, M. Hamidi et A. Chibani. Platform for Assessment and Monitoring of Infant Comfort. *The 2017 AAAI Fall Symposia*, Arlington, Virginia, USA, November 9-11, 2017, pages 36–44, 2017.
- [CI-8] P. Rastin, B. Matei et G. Cabanes. Signal-Based Autonomous Clustering for Relational Data. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, Alaska, USA, May 14-19, 2017, pages 4103–4110, 2017.
- [CI-7] M. Rifi, M. Hibti et R. Kanawati. Exploring network metrics for accident scenarios : a Case of study of the Uncontrolled Level Drop. *The 2017 International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2017)*, Pittsburgh, Pennsylvania, USA, September 24–28, 2017, Proceedings.com, pages 1177–1184, 2017.
- [CI-6] N. Rogovschi, J. Kitazono, N. Grozavu, T. Omori et S. Ozawa. t-Distributed stochastic neighbor embedding spectral clustering. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, Alaska, USA, May 14-19, 2017, pages 1628–1632, 2017.
- [CI-5] N. Sokolovska, Y. Chevaleyre et J.-D. Zucker. The Fused Lasso Penalty for Learning Interpretable Medical Scoring Systems. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, AK, USA, May 14-19, 2017, pages 4504–4511, 2017.
- [CI-4] H. Soldano, G. Santini, D. Bouthinon et E. Lazega. Hub-Authority Cores and Attributed Directed Network Mining. *The 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2017)*, Boston, MA, USA, November 6-8, 2017, pages 1120–1127, 2017.
- [CI-3] J. Sublime, B. Matei et P.-A. Murena. Analysis of the influence of diversity in collaborative and multi-view clustering. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017)*, Anchorage, AK, USA, May 14-19, 2017, pages 4126–4133, 2017.
- [CI-2] L.-M. Veillon, G. Bourgne et H. Soldano. Effect of Network Topology on Neighbourhood-Aided Collective Learning. *The 9th International Conference on Computational Collective Intelligence (ICCCI)*, Nicosia, Cyprus, September 27-29, 2017, volume 10448 de *Lecture Notes in Artificial Intelligence*, pages 202–211, 2017.
- [CI-1] L.-M. Veillon, G. Bourgne et H. Soldano. Waves : a Model of Collective Learning. *2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI'17)*, Leipzig, Germany, August 23-26, 2017, pages 314–321, 2017.

Autres articles dans des conférences ou workshops à comité de lecture

- 2023** [CO-32] R. Khoufache, M. D. Dilmi, H. Azzag, M. Lebbah et E. Goffinet. Propriétés émergentes du multi-clustering bayésien non paramétrique : Application aux données d'images multivues. *20ème Journées Francophones Extraction et Gestion des Connaissances (EGC 2023)*, Lyon, France, 16-20 Janvier 2023, volume E-39 de *Revue des Nouvelles Technologies Informatiques*, pages 639–640, 2023.

- 2022** [CO-31] M. Atzmueller et R. Kanawati. Explainability in Cyber-Security using Complex Network Analysis : A Brief Methodological Overview. *The European Interdisciplinary Cybersecurity Conference (EICC 2022)*, Barcelona, Spain, June 15 - 16, 2022, pages 49–52, 2022.
- [CO-30] R. Khoufache, M. D. Dilmji, H. Azzag, E. Goffinet et M. Lebbah. Emerging properties from Bayesian Non-Parametric for multiple clustering : Application for multi-view image dataset. *2022 IEEE International Conference on Data Mining Workshops (ICDM Workshops 2022)*, Orlando, FL, USA, November 28-December 1, 2022, pages 31–38, 2022.
- [CO-29] E. Leclercq, J. Rivalan, F. Roupin et C. Rouveirol. Allocation de ressources par une méthode hybride machine learning - optimisation en contexte de conteneurs. *Extraction et Gestion des Connaissances (EGC 2022)*, Blois, France, 24-28 janvier 2022, volume E-38 de *Revue des Nouvelles Technologies de l'Information*, pages 499–500, 2022.
- [CO-28] J.-V. Loddo et R. Kanawati. Mariotel : A virtual remote computer science lab. *The 1st International Conference on Online-Labs in Education*, Stuttgart, Germany, March 10-12, 2022, pages 485–494, 2022.
- 2021** [CO-27] M. Chevallier, N. Rogovschi, F. Boufarès et N. Grozavu. Techniques de génération de population initiale d'algorithmes génétiques pour la sélection de caractéristiques. *9ème Conférence Internationale Francophone sur la Science des Données (CIFSD 2021)*, Marseille, France, 9-11 Juin 2021, pages 1–12, 2021.
- [CO-26] E. Goffinet, M. Lebbah, H. Azzag, L. Giraldi et A. Coutant. Multiple Co-clustering de séries temporelles. *52èmes Journées de Statistique de la Société Française de Statistique (SFdS 2021)*, Nice, France, 7-11 Juin 2021, 2021.
- [CO-25] M. Hamidi et A. Osmani. Modélisation du processus de génération de données pour la reconnaissance d'activités. *Extraction et Gestion des Connaissances (EGC 2021)*, Montpellier, France, 25-29 janvier 2021, volume E-37 de *Revue des Nouvelles Technologies de l'Information*, pages 405–412, 2021.
- [CO-24] T. Mecharnia, L. Chibout Khelifa, F. Hamdi, N. Pernelle et C. Rouveirol. Découverte de règles contextuelles pour prédire la présence d'amiante dans les bâtiments. *32èmes Journées francophones d'Ingénierie des Connaissances (IC 2021)*, Bordeaux, France, June 30 - July 2, 2021, pages 81–88, 2021.
- [CO-23] J. Schoenke, N. Aschenbruck, R. Interdonato, R. Kanawati, A.-C. Meisener, F. Thierart, G. Vial et M. Atzmueller. Gaia-AgStream : An Explainable AI Platform for Mining Complex Data Streams in Agriculture. *The First International Conference on Smart and Sustainable Agriculture (SSA 2021)*, Virtual Event, June 20-21, 2021, volume 1470 de *Communications in Computer and Information Science*, pages 71–83, 2021.
- 2020** [CO-22] M. O. Attaoui, H. Azzag, M. Lebbah et N. Keskes. Multi-Objective Data Stream Clustering. *The 2020 Genetic and Evolutionary Computation Conference (GECCO'20)*, Cancún, Mexico, July 8-12, 2020, pages 113–114, 2020. (Poster).
- [CO-21] M. O. Attaoui, H. Azzag, M. Lebbah, N. Keskes et M. Ghesmoune. Soft Subspace Growing Neural Gas pour le Clustering de Flux de Données. *Extraction et Gestion des Connaissances (EGC 2020)*, Brussels, Belgium, January 27-31, 2020, volume E-36 de *Revue des Nouvelles Technologies de l'Information*, pages 441–448, 2020.
- [CO-20] H. N. Benkhaled, D. Berrabah et F. Boufarès. Block Sizes Control For an Efficient Real Time Record Linkage. *The 5th IEEE International Conference on Cloud Computing and Artificial Intelligence : Technologies and Applications (CloudTech 2020)*, Marrakesh, Morocco, November 24-26, 2020, pages 1–6, 2020.
- [CO-19] C. Deheeger, J. Pierre, J. Baptiste, S. Legras, A. Rimbaud, C. Rouveirol, H. Soldano et V. Ventos. Une expérience d'élicitation de connaissances expertes dans le domaine du Bridge. A. Cornuéjols et E. Cuvelier, éditeurs, *Extraction et Gestion des Connaissances (EGC 2020)*, Brussels, Belgium, January 27-31, 2020, volume E-36, de *Revue des Nouvelles Technologies de l'Information*, pages 97–108. Éditions RNTI, 2020.
- [CO-18] F. Forest, M. Lebbah, H. Azzag et J. Lacaille. Carte SOM profonde : Apprentissage joint de représentations et auto-organisation. *Conférence d'Apprentissage (CAp 2020)*, Vannes, France, June 23-26, 2020, 2020.
- [CO-17] E. Goffinet, M. Lebbah, H. Azzag et L. Giraldi. Clustering de séries temporelles par construction de dictionnaire. *20ème Edition des Journées Francophones Extraction et Gestion des Connaissances (EGC 2020)*, Bruxelles, Belgique, 27-31 Janvier 2021, volume E-36 de *Revue des Nouvelles Technologies*, pages 181–192, 2020.

3.2 Domaine 3. Production scientifique

- [CO-16] D. Maurel, S. Lefebvre et J. Sublime. Deep Cooperative Reconstruction with Security Constraints in Multi-View Environments. *The 20th International Conference on Data Mining Workshops (ICDM Workshops 2020)*, Sorrento, Italy, November 17-20, 2020, pages 579–588, 2020.
- [CO-15] S. Zevio, G. Santini, H. Soldano, H. Zargayouna et T. Charnois. A Combination of Semantic Annotation and Graph Mining for Expert Finding in Scholarly Data. *The 2020 Workshop on Graph Embedding and Mining Workshop (GEM 2020) co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2020)*, Virtual Event, September 18, 2020, page 16, 2020.
- 2019** [CO-14] P. Alizadeh, A. Osmani et E. Traversi. Calcul d'une politique déterministe dans un MDP avec récompenses imprécises. *Extraction et Gestion des connaissances (EGC 2019)*, Metz, France, January 21-25, 2019, volume E-35 de *Revue des Nouvelles technologies*, pages 45–56, 2019.
- [CO-13] H. N. Benkhaled, D. Berrabah et F. Boufarès. A novel approach to improve the Record Linkage process. *The 6th IEEE International Conference on Control, Decision and Information Technologies (CoDIT 2019)*, Paris, France, April 23-26, 2019, pages 1504–1509, 2019.
- [CO-12] M. Hamidi et A. Osmani. HAR Data Generation Model to Highly Reduce Learning Space. *Knowledge Representation & Reasoning Meets Machine Learning workshop (K2RML@NeurIPS 2019)*, Vancouver, Canada, September 19, 2019, pages 1–11, 2019.
- [CO-11] M. Hamidi et A. Osmani. Improving Human Activity Recognition with Data Sources Integration. *Deep Learning for Human Activity Recognition workshop @IJCAI 2019*, Macao, China, August 10-16, 2019, 2019.
- [CO-10] P. Rastin, B. Matei et G. Cabanes. Apprentissage non-supervisé relationnel dans l'espace des coordonnées barycentriques. *Extraction et Gestion des Connaissances (EGC 2019)*, Metz, France, January 21-25, 2019, volume E-35 de *Revue des Nouvelles Technologies de l'Information*, pages 413–418, 2019.
- [CO-9] P. Rastin, B. Matei et G. Cabanes. Détection de changement dans les profils en ligne d'utilisateurs. *Extraction et Gestion des Connaissances (EGC 2019)*, Metz, France, January 21-25, 2019, volume E-35 de *Revue des Nouvelles Technologies de l'Information*, pages 419–424, 2019.
- [CO-8] S. Zevio, G. Santini, H. Zargayouna et T. Charnois. Fouille de texte et fouille de graphe appliquées à la recherche d'experts. *30èmes journées francophones d'Ingénierie des Connaissances (IC 2019)*, Toulouse, France, 2-4 juillet 2019, pages 222–223, 2019.
- 2018** [CO-7] M. Atzmueller, H. Soldano, G. Santini et D. Bouthinon. MinerLSD : Efficient Local Pattern Mining on Attributed Graphs. *The 2018 IEEE International Conference on Data Mining Workshops (ICDMW 2018)*, Singapore, November 17- 20, 2018, pages 219–228, 2018.
- [CO-6] D. Maurel, J. Sublime et S. Lefebvre. Cartes Auto-Organisatrices Incrémentales appliquées au Clustering Collaboratif. *Extraction et Gestion des Connaissances (EGC 2018)*, Paris, France, January 23-26, 2018, volume E-34 de *Revue des Nouvelles Technologies de l'Information*, pages 403–408, 2018.
- [CO-5] M. Rifi, M. Hbiti et R. Kanawati. A Complex Network Analysis Approach for Risk Increase Factor Prediction in Nuclear Power Plants. *The 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2018)*, Funchal, Madeira, Portugal, March 20-21, 2018, 2018.
- [CO-4] S. Zevio, H. Zargayouna, G. Santini et T. Charnois. Vers une cartographie automatique des thématiques et profils d'experts associés à une conférence scientifique : 9 ans d'ateliers recherche d'information sémantique. *10ème Atelier Recherche d'Information Sémantique (RISE@CORIA/TALN 2018)*, Rennes, France, 15 Mai 2018, rennes, france, 2018.
- 2017** [CO-3] G. Beck, H. Azzag, M. Lebbah et T. Duong. Mean-shift : Clustering scalable avec les plus proches voisins approximés. *Extraction et Gestion des Connaissances (EGC 2017)*, Grenoble, France, 24-27 Janvier 2017, volume E-34 de *Revue des Nouvelles Technologies de l'Information*, pages 415–425, 2017.
- [CO-2] A. Coutant et C. Rouveirol. Network Inference of Dynamic Models by the Combination of Spanning Arborescences. *Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM 2017)*, Lille, France, 3-6 juillet 2017, pages 45–52, 2017.

- [CO-1] H. Léger, D. Bouthinon, M. Lebbah et H. Azzag. Nouveau modèle pour un passage à l'échelle de la 0-subsumption. *17ème Journées Francophones Extraction et Gestion des Connaissances (EGC 2017)*, 24-27 Janvier 2017, Grenoble, France, volume E-33 de *Revue des Nouvelles Technologies de l'Information*, pages 339-344, 2017.

Direction d'ouvrages (DO)

Livres

- 2019 [LI-1] I. Redko, A. Habrard, E. Morvant, M. Sebban et Y. Bennani. *Advances in Domain Adaptation Theory*. Elsevier, 2019.

Édition d'ouvrages collectifs

- [ED-2] E. Banissi, A. Ursyn, M. W. M. Bannatyne, N. Datia, R. Francese, M. Sarfraz, T. G. Wyeld, F. Bouali, G. Venturini, H. Azzag, M. Lebbah, M. Trutschl, U. Cvek, H. Müller, M. Nakayama, S. Kernbach, L. Caruccio, M. Risi, U. Erra, A. Vitiello et V. Rossano. *The 23rd International Conference on Information Visualisation (IV 2019)*, Paris, France, July 2-5, 2019, Part I. IEEE, 2019.
- 2018 [ED-1] M. Lebbah, C. Llargeron et H. Azzag. *18ème Conférence Internationale sur Extraction et Gestion des Connaissances (EGC 2018)*, Paris, France, January 22-26, 2018, volume E-34 de *Revue des Nouvelles Technologies de l'Information*. Hermann-Éditions, 2018.

Logiciels et Ressources (LO)

- 2022 [LO-13] M. Hamidi et A. Osmani. Reduction of the Position Bias via Multi-Level Learning for Activity Recognition. <https://github.com/alphaequivalence/positionAbstraction>, 2022.
- [LO-12] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. EvalSeg : Multi Modal Evaluation For Medical Image Segmentation, 2022.
- [LO-11] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. UnifiedAR : A Modular Platform for Activity Recognition from Sensor Data, 2022.
- 2021 [LO-10] F. Forest, M. Lebbah, H. Azzag et J. Lacaille. DESOM : Deep Embedded Self-Organizing Map model. <https://github.com/FlorentF9/DESOM>, 2021.
- [LO-9] E. Goffinet, M. Lebbah et H. Azzag. MultiCocustering : 3 méthodes de CoClustering. <https://github.com/EtienneGof/MultiCocustering>, 2021.
- [LO-8] M. Hamidi et A. Osmani. Augmented Experiments in Material Engineering Using Machine Learning. <https://github.com/alphaequivalence/nx-project>, 2021.
- [LO-7] M. Hamidi et A. Osmani. Hierarchical Learning of Dependent Concepts for Human Activity Recognition. <https://github.com/alphaequivalence/hierarchicalSHL>, 2021.
- [LO-6] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. Auto Profiler : Automatic Interactive Tree-based Profiling of Python Scripts in Jupyter Notebook. https://github.com/modaresimr/auto_profiler, 2021.
- [LO-5] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. ConfJournalRank : Integrated Multi SourcesConference and Journal Information". <https://rank.modaresi.eu.org>, 2021.
- [LO-4] S. M. R. Modaresi, A. Osmani, M. Razzazi et A. Chibani. JCV2 : Enabling Compatibility of OpenCV UI Components with Jupyter Notebooks. https://github.com/modaresimr/opencv_jupyter_ui, 2021.
- 2020 [LO-3] A. Coutant, E. Goffinet, M. Lebbah et H. Azzag. spark-tss : une API pour le prétraitement de séries temporelles multi-variables. <https://github.com/spark-tss/spark-tss>, 2020.
- [LO-2] M. Hamidi et A. Osmani. Learning Mobility-Related Human Activities from Sensor-Rich Environments. <https://github.com/HamidiMassinissa/shl-nas>, 2020.
- 2018 [LO-1] G. Beck, M. Lebbah, H. Azzag et A. Coutant. Clustering4Ever : une API pour le clustering. <https://github.com/Clustering4Ever>, 2018.

Thèses et Habilitations (TH)

Thèses

- 2022** [TH-18] M. Chevallier. L'Apprentissage Artificiel au service de la qualité de données : détection de contraintes. Thèse, Université Sorbonne Paris Nord, 2022.
- [TH-17] M. El Hamri. Structural Optimal Transport for Domain Adaptation with Theoretical Guarantees. Thèse, Université Sorbonne Paris Nord, 2022.
- [TH-16] Y. Foucade. Learning from Data and Learners. Thèse, Université Sorbonne Paris Nord, 2022.
- [TH-15] M. Hamidi. Metaapprentissage guidé par les connaissances du domaine. Thèse, Université Sorbonne Paris Nord, 2022. Sous la direction de Aomar OSMANI.
- [TH-14] A. Zaiou. Quantum Machine Learning Approaches for Graphs and Sequences : Application to Nuclear Safety Assessment. Thèse, Université Sorbonne Paris Nord, 2022.
- 2021** [TH-13] M. O. Attaoui. Vers de Nouvelles Méthodes de Clustering de Flux de Données. Thèse, Université Sorbonne Paris Nord, 2021.
- [TH-12] F. Forest. Apprentissage non supervisé de représentations de données et structures de partitionnement : applications à la surveillance à grande échelle de turbofans. Thèse, Université Sorbonne Paris Nord, 2021.
- [TH-11] E. Goffinet. Clustering Multi-Blocs et Visualisation Analytique de Données Séquentielles Massives Issues de Simulation du Véhicule Autonome. Thèse, Université Sorbonne Paris Nord, 2021.
- 2020** [TH-10] F. E. Ben Bouazza. Multi-models clustering through optimal transport theory. Thèse, Université Sorbonne Paris Nord, 2020.
- [TH-9] K. Benlamine. Apprentissage non supervisé profond quantique de représentations de données évolutives. Thèse, Apprentissage non supervisé profond quantique de représentations de données évolutives, 2020.
- [TH-8] S. Zouinina. Data Anonymization through Unsupervised Learning. Thèse, Université Sorbonne Paris Nord, 2020.
- 2019** [TH-7] G. Beck. Scalable Clustering Applying Local Accretions. Thèse, Université Sorbonne Paris Nord, 2019.
- [TH-6] H. Léger. Passage à l'échelle de la theta-subsumption. Thèse, Université Sorbonne Paris Nord, 2019.
- [TH-5] M. Rifi. Modélisation et Analyse des Réseaux Complexes : application à la sûreté nucléaire. Thèse, Université Sorbonne Paris Nord, 2019.
- 2018** [TH-4] I. Falih. Attributed Network Clustering : Application to recommender systems. Thèse, Université Sorbonne Paris Cité, 2018.
- [TH-3] P. Rastin. Apprentissage non-supervisé à partir de données évolutives. Thèse, Université Sorbonne Paris Cité, 2018.
- [TH-2] T. Sarazin. Apprentissage distribué de grandes masses de données. Thèse, Université Sorbonne Paris Cité, 2018.
- [TH-1] L.-M. Veillon. Apprentissage artificiel collectif, aspects dynamiques et structurels. Thèse, Université Sorbonne Paris Cité, 2018.

Habilitations

- 2021** [HR-3] N. Rogovschi. Contributions to data dimensionality reduction and co-clustering. HdR, Université Sorbonne Paris Nord, 2021.
- [HR-2] J. Sublime. Contributions to modern unsupervised learning : Case studies of multi-view clustering and unsupervised Deep Learning (Contributions à l'apprentissage non-supervisé moderne : Applications aux cas du clustering multi-vue et de l'apprentissage profond non-supervisé). HdR, Université Sorbonne Paris Nord, 2021.
- 2020** [HR-1] N. Grozavu. Contributions on Collaborative Unsupervised. HdR, Université Sorbonne Paris Nord, 2020.

Publications antérieures à l'arrivée au LIPN

Les références comportent « ↑ » et correspondent à des publications antérieures à l'arrivée au LIPN, non référencées par ailleurs.

Signification des soulignements :

- *simple* : enseignant-chercheur ou chercheur titulaire,
- *double* : doctorant,
- *pointillé* : post-doc, ingénieur ou autre situation.

3.3 Domaine 4. Inscription des activités de recherche dans la société

3.3.1 Référence 1. L'équipe se distingue par la qualité et la quantité de ses interactions avec le monde non-académique.

L'équipe a plusieurs collaborations industrielles qui permettent à ces membres de financer les doctorants et les diverses actions pour la recherche. Durant la période entre 2017 et 2022, l'équipe a eu 19 projets avec des acteurs socio-économiques variés (SAFRAN, Renault, EDF, FaceBook, Nukkai, etc.). L'ensemble de ces contrats représente plus de 2 millions d'euros (voir section 3.1.2).

3.3.2 Référence 2. L'équipe développe des produits à destination du monde culturel, économique et social.

Le laboratoire met tout en œuvre pour accompagner les initiatives de valorisation et de transfert comme il a pu le prouver avec l'accompagnement actif de la startup HephIA.

Suite au succès de plusieurs collaborations industrielles et à la participation à plusieurs projets investissement d'avenir (PIA), ANR et CIFRE, fin 2020, un membre de l'équipe A³ ainsi qu'un ancien doctorant et un ancien postdoctorant de l'équipe ont porté un projet de création de startup devant la Société d'Accélération du Transfert de Technologies (SATT) Erganeo. L'obtention d'un financement de maturation (470k€) a permis de concrétiser cet objectif. L'objectif de HephIA est de cartographier les données massives de façon automatisée en détectant les informations les plus pertinentes à extraire. A la fin 2022, la taille de la startup HephIA hébergée par l'équipe A³ a atteint 14 membres.

Chapitre 4

Trajectoire de l'équipe A3

Sommaire

| | |
|---|-----------|
| 4.1 Organisation et ambitions de l'équipe | 53 |
| 4.2 Analyse AFOM/SWOT | 54 |
| 4.3 Perspectives et projet de recherche | 55 |
| 4.3.1 Axe ADA : Apprentissage à partir de Données et d'Apprenants | 55 |
| 4.3.2 Axe MAR : Métaapprentissage et Apprentissage de Structures | 56 |
| 4.3.3 Axe ARG : Apprentissage Relationnel et Graphes | 59 |

4.1 Organisation et ambitions de l'équipe

La structuration de l'équipe, sous sa forme actuelle, est récente. Les objectifs attendus sont une meilleure coordination des membres pour créer des synergies autour des nouveaux axes structurants de l'équipe. Cette coordination doit se centrer sur l'amélioration de la visibilité de nos activités, des recrutements centrés sur ces ambitions et la recherche de collaborations, notamment, internationales.

Des collaborations locales doivent aussi être renforcées. L'intelligence artificielle est devenue incontournable pour beaucoup d'activités de recherche, l'équipe renforcera

4.2 Analyse AFOM/SWOT

naturellement ses collaborations avec équipes RCLN, AOC et dans une moindre mesure avec les équipes LOVE et CALIN. L'équipe collabore aussi avec plusieurs équipes au sein de la fédération Mathstic à travers des coencadrement de doctorants et la participation à plusieurs projets de recherche. L'équipe souhaite, également, développer un projet interdisciplinaire d'intérêt général au niveau de l'université USPN. Nous discutons actuellement sur un projet autour de l'IA, l'apprentissage automatique et l'éthique. Ce type de projet nécessite des disponibilités de quelques membres de l'équipe.

L'équipe a redéfini trois axes de recherche et souhaite limiter la diversification des thématiques afin de contribuer significativement sur ces thèmes et renforcer son expertise théorique. Dans chacun des axes, des problèmes fondamentaux sont étudiés en lien avec des domaines d'applications qui répondent aux besoins de l'environnement socio-économique essentiellement à travers la formation par la recherche de doctorants en CIFRE.

Malgré l'ambition affichée, le risque principal est la poursuite de l'affaiblissement de l'équipe avec des départs non compensés, notamment celui d'un rang A en 2017.

4.2 Analyse AFOM/SWOT

• Forces et Atouts

- *Positionnement original et large spectre scientifique de l'équipe*, lui assurant une bonne visibilité extérieure sur ses thèmes de recherches ;
- *Attractivité du thème et de l'équipe dans le monde académique et industriel* : nombreux doctorants, bon taux d'insertion académique et industriel à l'issue de la thèse, nombreux partenariats industriels (ANR) et industriels (FUI, PIA, CIFRE) ;
- *Formation à la recherche* : adossement à un master très visible (EID2) ;
- *Valorisation de la recherche* : transfert technologique par la création de start-ups ;
- *Nombreuses initiatives d'animation de la recherche* : organisation de conférences françaises, d'écoles thématiques, de tutoriels, de sessions spéciales dans des conférences internationales.

• Faiblesses

- *Absence de chercheurs permanents* : la politique de recrutement de chercheurs CNRS doit être dynamisée, notamment à travers le recrutement de postdocs et un recours plus fréquent à des mois de chercheurs invités ;
- *Relations internationales à intensifier*, notamment au niveau des contrats internationaux ;
- *Diffusion de logiciels à renforcer* : des logiciels sont développés dans l'équipe, mais leur diffusion est encore trop restreinte, faute de personnel dédié au développement ;
- *Concurrence* : la concurrence avec les unités parisiennes et les centres de recherche privés fragilise l'équipe (difficulté à recruter doctorants et permanents), malgré un positionnement fort et original ;
- *arité de genre* : L'équipe compte peu de femmes.

• Opportunités

- *Visibilité accrue de l'apprentissage (et de l'Intelligence Artificielle en général)* dans le grand public, ce qui peut susciter des vocations d'étudiants ;
- *Renforcement de notre position à l'international* : les liens avec d'autres équipes et chercheurs à l'international doivent s'intensifier ;
- *Large couverture des thèmes de l'apprentissage avec des positionnements originaux* ;

- *Développement de recherches transverse en apprentissage à travers des projets interdisciplinaires notamment avec les laboratoires MathSTIC et L2TI.*

- **Menaces**

- *Malgré l'adossement à un M2, les entreprises attirent de plus en plus de bons étudiants qui se détournent des carrières universitaires sur le domaine de l'apprentissage artificiel et de la fouille de données ;*
- *Les départs d'HDR déstabilisent certains thèmes de recherche qui risquent de disparaître ;*
- *L'affaiblissement de l'équipe du fait du non-remplacement des départs.*

4.3 Perspectives et projet de recherche

Le projet de l'équipe sur la prochaine période poursuivra les travaux en cours en allant vers une meilleure intégration des activités et un renforcement de l'expertise en fondements théoriques de l'apprentissage. La réalisation de ce projet dépendra de l'évolution de l'équipe, notamment en termes de recrutement d'enseignants chercheurs permanents.

4.3.1 Axe ADA : Apprentissage à partir de Données et d'Apprenants

Transport Optimal et Analyse Topologique pour l'Apprentissage Multi-modale : La théorie du transport optimal offre un cadre formel très puissant pour l'apprentissage artificiel. Trois thèses de doctorat sur cette thématique ont été soutenues. Le départ de ces diplômés ainsi que certains membres permanents de l'axe ADA, vers d'autres laboratoires académiques va réduire les activités autour de ce thème pourtant pertinent et présentant de nombreuses perspectives. Par ailleurs, suite au manque de forces locales, le thème portant sur l'apprentissage quantique ne sera plus poursuivi dans l'axe ADA. Néanmoins, nos contributions sur cette thématique très prometteuse a permis de proposer des approches très pertinentes et innovantes dans le domaine, à travers notamment deux thèses de doctorat soutenues au LIPN, dont une dans le cadre d'une CIFRE avec EDF. Par conséquent, et suite à l'évolution des questions scientifiques autour de l'apprentissage, le projet que nous présentons pour l'axe ADA sera resserré uniquement autour d'un seul thème, ce qui permettra d'animer plus efficacement l'axe. Les données sont souvent présentées dans un format qui couvre plusieurs modalités, comme le matériel vidéo, qui comprend normalement au moins une image et du son (voix, musique), ainsi que du texte sous forme de sous-titres, qui sont souvent dans une langue autre que les données parlées. D'autres situations, comme la détection humaine, fournissent des données dans plusieurs modalités, par exemple des données sur l'expression du visage sous forme de photographies mélangées à des données auditives (voix, son), haptiques (toucher) ou autres données sensorielles. Le défi de la recherche sur l'analyse des données multimodales est donc axé sur l'intégration et l'interprétation des données dans et entre les modalités, ainsi que les connaissances et idées qui en découlent. Les résultats de cette recherche permettront d'améliorer la compréhension et la modélisation de sources de données riches dans des domaines tels que les affaires, la santé, l'environnement et l'éducation. Le problème fondamental est d'apprendre à décrire et à résumer les données multimodales d'une manière qui exploite la complémentarité et la redondance des différentes modalités. Pour ce faire, des approches fondées sur le transport optimal et l'analyse topologique des données (TDA : Topological Data Analysis) seront développées.

4.3.2 Axe MAR : Métaapprentissage et Apprentissage de Structures

Modèles génératifs et Multi-Cocustering : Nous souhaitons poursuivre l'étude des modèles génératifs pour les données temporelles multivariées [CO-26, CO-18]. Pour résoudre le problème de données manquantes, nous proposons d'utiliser des modèles génératifs (GAN) ceci dans le contexte d'une thèse *CIFRE* en cours avec *Safran*. Des travaux récents proposent des TTS-GAN basés sur les transformers et constituent une autre méthode permettant de résoudre en grande partie le problème de disparition du gradient qui entrave les RNNs en utilisant des neurones basés sur l'attention. Dans le même sens, dans le cadre du pôle *MathStic*, en collaboration avec le laboratoire L2ti (laboratoire de Traitement et Transport de l'Information de Sorbonne Paris Nord) renforcée par le LabCom *IRISER*, nous étudions l'apport des mécanismes d'attention dans le traitement des données multimodales (images et textes). Une des techniques permettant de prendre en compte la multimodalité est d'agir au niveau du processus de normalisation. Ces travaux seront également renforcés par un projet *IHU* qui a été accepté récemment (courant Mai 2023) dont l'objectif est de créer un écosystème multidisciplinaire alliant des forces de recherche fondamentale et clinique pour analyser et réduire l'impact (décès et séquelles) de la maladie du sepsis.

En ce qui concerne le *Multi-Cocustering*, pour tenir compte des données non euclidiennes, nous souhaitons utiliser une approche à bloc latent bayésien non paramétrique. Nous envisageons d'étendre nos activités dans le domaine du *Multi-Cocustering* [CI-87] pour prendre en compte des données multivues et multimodales provenant de différentes sources (image, texte, etc). Nous proposons également de prendre en compte le principe de l'adaptation des modèles topologiques aux données multivues, nous proposons de travailler sur de nouvelles approches topologiques du *Multi-CoClustering (MCC Topologique)* qui regroupe les variables partageant la même partition d'observations et la même vue [18]. Un de nos principaux objectifs sera d'étudier l'explicitation des résultats du *Multi-CoClustering Topologique*, c'est-à-dire permettre un retour visuel le plus efficace du *clustering* [TH-7]. Ces travaux seront soutenues par une thèse financée par le programme *DIM A4IDF* en collaboration avec l'UVSQ Paris-Saclay qui a débutée en Janvier 2023.

Par ailleurs, les applications basées sur l'apprentissage profond ont connu un fort succès ces dernières années. Poursuivre nos activités en apprentissage profond et mécanismes d'attention est une des perspectives à court termes de l'équipe. Notamment, pour encoder/vectoriser les données brutes non formatées (séries temporelles, images, etc.). Les données disponibles pour le clustering sont issues de sources hétérogènes et déséquilibrées que nous souhaitons explorer pour la détection de fraude dans les données bancaires dans le cadre d'une collaboration *CIFRE* avec la *BPCE*.

Métaapprentissage : En métaapprentissage, la minimisation du nombre d'exemples pour apprendre, la quantification de tous les biais d'apprentissage et la certification ou la quantification de la qualité de l'apprentissage font partie des objectifs des activités de l'équipe. L'intégration des connaissances du domaine passent par la recherche de prédicats génériques et d'invariants (plus généralement régularités) génériques indépendants du domaine [CI-97]. Il peut s'agir de modèles équationnels, de topologies ou de toute forme de régularités sur les processus ou les structures qui minimisent l'ensemble des fonctions admissibles [21]. On utilise des applications à base de capteurs, car elles présentent un cadre complet pour l'étude des problèmes du métaapprentissage. Ces problèmes, comme nous l'avons montré dans le papier qui synthétise les biais en apprentissage [RI-18] sont relativement complets. Ils portent sur diverses multimodalités (images, son, accélération, ...), la relativité des données [CI-98], les combinaisons de capteurs [CI-52], le transfert des données (traitement séquentiel, parallèle des données ou synthèse du transfert de l'apprentissage), la coexistence de multiples théories d'apparence contradictoires (correspondant aux points de vue différents) ou

aux mécanismes de conciliation [CI-97], la structuration et le traitement des données, le temps réel et l'intégration des questions liées à la consommation d'énergie, au débit et aux capacités de traitement locales, à l'écologie, à la protection des données et à l'éthique qui doivent aussi être traités par les approches du métaapprentissage. Deux véhicules, conçus par deux constructeurs différents, peuvent, par exemple, utiliser les mêmes données pour aboutir à des théories inconsistantes, mais qui coexistent. La collaboration dans ce cas n'est pas centrée sur la théorie générale à apprendre mais sur d'autres objectifs externes : ici, éviter les accidents et fluidifier la circulation, par exemple. L'expressivité du modèle de métaapprentissage doit être capable d'englober la formalisation de ce type de problèmes.

Il en est de même pour des questions qui portent, par exemple, sur l'interprétation, l'explication de modèles d'apprentissages ou par la nécessité qu'ils respectent des normes (loi, éthique, etc.). Il s'agit de concepts à traiter par l'apprentissage, mais à d'autres niveaux d'abstraction, comme c'est le cas, par exemple, des modèles MOF en métamodélisation. La modélisation du métaapprentissage doit gérer les hyperparamètres qui les définissent et gèrent leurs liens avec les algorithmes d'apprentissage. Prenons l'exemple de l'explication : au niveau du modèle abstrait du métapprentissage, l'explication peut se définir comme un problème d'optimisation et d'apprentissage de représentations entre trois espaces : la représentation de la théorie apprise, la représentation de l'acteur à qui l'explication est destinée et la représentation d'un espace intermédiaire qui est un simple changement de représentation de la théorie apprise avec le moins de perte possible tout en s'approchant le plus possible de la représentation de l'acteur. Du point de vue de la spécification concrète, il s'agit d'un jeu d'optimisation de métriques entre ces trois espaces de représentations. Le métaapprentissage doit donner un cadre pour expliciter ces objets et les leurs interrelations. Il en est de même pour l'éthique. Dans le cas de la recherche de solutions d'apprentissage automatique qui respectent l'éthique, la décomposition du problème par le métaapprentissage le rend plus facile à résoudre. Initialement, il convient de respecter un ensemble de principes (implémentés par des règles) décrit dans un langage de représentation donné pour lesquels il convient de trouver une bonne représentation qui minimise la perte par rapport au langage initial et qui permet d'être traduite dans la représentation de la théorie apprise, par les modèles d'apprentissage, pour contraindre les réponses à celles admises par l'éthique. Du point de vue opérationnel, il existe plusieurs manières de faire. L'une d'elle consiste à incorporer ces règles lors de la conception du système. Soit via des coefficients de régularisation (solution souvent utilisée par les réseaux de neurones profonds), soit via des modèles proches de ceux sur lesquels nous travaillons : recherche de la modularité et d'invariants dans les frontières de décisions de sorte à définir des composants autonomes évaluables et composables. Ce procédé permet d'assurer une transparence des boîtes noires en contrôlant les entrées et sorties des composants élémentaires les constituant, et en intégrant les contraintes (dont celles de l'éthique, dans ce cas) par le contrôle des échanges entre ces composants/Services élémentaires. La deuxième manière de faire est totalement exogène, elle consiste à considérer la théorie apprise comme une boîte noire et à développer des outils d'audit externes et indépendants permettant de garantir le test et la validation des règles que l'on souhaite faire imposer au système.

Ces types de structures et de contrôles nécessitent de construire un modèle de métaapprentissage global permettant de formaliser, décrire, instancier et exécuter ce type de scénarios. Construire une telle architecture est complexe. Elle impose la recherche d'une simplicité de la métamodélisation qui doit comporter très peu de prédicats. Cette recherche de la simplification passe par un élément fondamental, la définition du système d'évaluation qui structure la définition du pipeline de métaapprentissage de bout en bout. Dans cette définition du métaapprentissage, l'algorithme (ou les algorithmes) n'est qu'un hyperparamètre parmi d'autres. La recherche d'une "bonne" architecture nécessite de se poser les questions non limitatives à propos des problèmes suivants (1) les données qui sont rarement des faits, (2) les propriétés intrinsèques du problème et à

4.3 Perspectives et projet de recherche

la structure des solutions disponibles pour le résoudre, (3) la modélisation explicite de tous les biais et leur quantification et (4) la structure de l'évaluation qui impose le plus de biais implicites. En plus de ces objectifs à court terme, l'équipe souhaite traiter des questions plus théoriques liées au métaapprentissage telles que :

Caractérisation du paysage d'optimisation Cette perspective regroupe les travaux sur l'étude des régularisations dans l'espace des biais qui permettent de structurer les solutions à la fois pour faciliter le transfert, permettre une modularité (ou un apprentissage par composants) facilitant ainsi la transparence des résultats et ouvrant une voie pour l'explication, l'interprétation et le contrôle explicite des résultats (imposer le respect de la réglementation, de l'éthique, de standards, etc). En particulier,

- La connaissance du domaine a un impact important sur le paysage de l'optimisation pour faciliter l'exploration. C'était le cas dans notre travail sur l'utilisation des invariants du domaine sous forme de symétries entre les différentes vues fournies par les capteurs ayant une distribution dynamique, mais structurée et régulière. Comme le rappelle [23], *Les symétries des modèles d'observation deviennent des symétries du problème d'optimisation*. De plus, comme nous l'avons montré [**CI-97**, **RI-26**], la manière dans le problème d'apprentissage est structurée en sous-problèmes en interaction via le transfert, le partage et la réutilisation, se traduit également dans le paysage d'optimisation par des propriétés qui semblent intéressantes qu'il convient d'explorer. Une voie possible est d'explorer le lien qui existe entre les limites de décisions du paysage d'optimisation et les invariants explicites du domaine.
- Les connaissances a priori ajoutées à une représentation explicite du paysage d'optimisation peuvent-elles conduire à des sauts inductifs substantiels dans l'espace d'optimisation? Peut-on utiliser les approches d'optimisation globales en présence de connaissances structurelles à la place des méthodes locales? Ceci permettrait une meilleure structuration des biais des algorithmes et faciliterait le transfert d'hypothèses,
- Les travaux de Vapnik sur les prédicats à travers les invariants suggèrent qu'il existerait un petit nombre de prédicats qui structureraient les connaissances du domaine. C'est le cas notamment en utilisant les groupes de symétries comme les groupes de Lie que nous comptons poursuivre pour encoder des connaissances du domaine [19]. Des travaux récents généralisent ce problème d'apprentissage au calcul variationnel nécessitant une optimisation des paramètres dans l'espace courbe [19]. Une voie prometteuse est l'optimisation sur des variétés lisses ou Riemanniennes, ce qui revient à faire l'optimisation sur une structure de variété connue [16]. Ceci pourrait avoir comme conséquence d'améliorer le processus d'optimisation car il existe une vaste littérature sur les méthodes d'optimisation plus adaptées à ces variétés [1, 2, 3, 22, 13, 12] qui étendent les algorithmes de convergence classiques du cadre Euclidien au cadre Riemanien.

Trajectoire de
l'équipe A3

Recherche de correspondances entre connaissances du domaine et les paramètres des modèles Dans le cas des modèles symboliques, cette correspondance est plus simple à établir. Elle est plus difficile dans le cas des approches connexionnistes. Le choix de certaines architectures imposent des contraintes aux modèles et les formes à apprendre dans des sous-espaces contraints comme nous l'avons fait dans nos travaux [**CI-98**]. Dans les derniers travaux de V. Vapnik [21], il résume cette tâche à la recherche de prédicats encodant des régularités (invariants, etc.), à la recherche de fonctions admissibles et à la convergence faible. Les méthodes qui permettent de capturer naturellement des aspects précis du domaine permettront de construire des apprenants robustes et évolutifs. Par exemple, existe-t-il une correspondance entre les modèles d'apprentissage et leurs capacités de calcul et la connaissance du domaine exprimée dans un langage spécifique, à l'image de la correspondance Curry et Howard [10] entre les modèles de calcul et les systèmes de preuves? Peut-on concevoir des mo-

dèles qui encodent, dans leurs mécanismes et composants internes, des biais inductifs correspondant aux connaissances du domaine ? Il existe plusieurs travaux dans ce sens pour tenter de combler l'aspect boîte noire des modèles neuronaux [6, 7, 9]. Ces systèmes suggèrent, comme c'est le cas par exemple, dans les architectures logicielles, que chaque composant neuronal soit responsable d'une connaissance spécifique, cette modularité, comme nous avons pu le faire dans [CI-86, CI-97] et [CI-98] est une voie possible pour la transparence et l'explication.

4.3.3 Axe ARG : Apprentissage Relationnel et Graphes

Fouille de graphes attribués, caractérisation des structures au-delà des énumérations Nos travaux portant sur la fouille de graphes attribués nous ont permis de proposer i) une approche d'énumération réduisant l'espace d'exploration aux seuls motifs clos supportés par un ensemble de sommets dont le sous-graphe induit vérifie une propriété de coeur et ii) une approche de sélection réduisant le nombre de motifs en assurant que les motifs présentés soient suffisamment différents. Si ces restrictions sont intéressantes du point de vue de l'exploitation humaine des résultats en limitant le nombre de motifs à considérer, elles induisent nécessairement une perte d'information par rapport au graphe attribué. Dans un premier temps, nous chercherons à quantifier et à limiter cette perte d'information en proposant des représentations augmentées des motifs actuels qui puissent par exemple prendre en compte des informations statistiques sur les motifs exclus lors des sélections. Nous chercherons ensuite à mettre en place des méthodes d'exploitation des motifs énumérés, par exemple, pour des systèmes de recommandation en considérant les distances en intention et en extension des points exclus de l'ensemble support des motifs retenus. Nous chercherons à valider ces approches dans le cadre de collaborations en cours de constitution en les appliquant à des graphes de terrain élaborés à partir de données d'enquêtes de type ethnographiques.

Explications pour la prise de décision séquentielle et en monde partiellement observable Nous nous sommes intéressés dans des travaux récemment acceptés à la notion d'explication dans des langages expressifs (restrictions de la logique des prédicats) pour problèmes de décision séquentiels modélisés par des processus de décision de Markov (MDP)[17]. Nous avons proposé une formalisation et des algorithmes de Programmation Logique Inductive [5] pour réinterpréter la notion d'explication *abductive* de P. Marquis [14] et proposer des explications communes minimales pour un groupe d'exemples partageant une même classe. Ces travaux peuvent être étendus dans de nombreuses directions, une première étant de relever les défis liés aux différents aspects à la complexité de calculer de telles explications en logique d'ordre un (borner le nombre d'explications abductives/ la complexité du calcul de ces explications). Une autre extension que nous envisagerons est de produire des explications de trajectoires dans un cadre partiellement observable. Cette extension ouvre de nombreuses perspectives d'applications, car dans de nombreuses situations réelles, la prise de décision séquentielle se fait dans un environnement partiellement observable (POMDP)[17], c'est-à-dire que tout ou partie des états sur les trajectoires étudiées – dont des éléments éventuellement essentiels pour la prise de décision – ne sont pas observés. Par exemple, dans un jeu de cartes, les mains des adversaires ne sont pas observées. Elle pose également de nombreux défis théoriques et algorithmiques.

Dans le cadre qui nous intéresse, une théorie du domaine (modèle partiel de l'environnement) nous permettra de raisonner et de calculer un *état de croyance*, i.e. une distribution de probabilité sur les états possibles satisfaisant les contraintes de la théorie du domaine. De nombreux formalismes et outils ont été conçus et développés ces dernières années pour faire du raisonnement et de l'apprentissage dans un cadre incertain (mentionnons [15] et plus récemment [20]), nous allons pouvoir nous appuyer sur ces formalisations et outils développés en logique des prédicats probabiliste pour étendre notre cadre logique d'explications en logique des prédicats pour des POMDPs.

Au-delà des aspects techniques, construire une explication dans un cadre POMDP est un problème ouvert. On peut l'exprimer comme un problème apprentissage d'une politique explicite dans un cadre POMDP et tenter de généraliser des travaux sur l'élicitation de politiques dans un cadre MDP [8, 4] à un cadre partiellement observable. Un article de J. Lang et B. Zanuttini [11] propose de représenter des politiques dans un cadre partiellement observable avec des *Knowledge Based Probabilistic programs*, qui référencent explicitement à un état de croyance de l'agent "décideur". De tels *Knowledge Based Probabilistic programs* exprimés sous la forme de programmes logiques probabilistes peuvent servir à exprimer des politiques à suffisamment haut niveau pour être compréhensibles et transférables. Apprendre de tels programmes est donc une voie prometteuse pour construire des explications que nous souhaitons. Nous évaluerons les approches proposées notamment dans le domaine des jeux de cartes et du Bridge, en collaboration avec la société Nukkai, avec laquelle ces approches ont été initiées.

Références complémentaires

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization algorithms on matrix manifolds. In *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.
- [2] S. Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Trans. on Automatic Control*, 58(9), 2013.
- [3] N. Boumal, P.-A. Absil, and C. Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1) :1–33, 2019.
- [4] Y. Coppens, D. Steckelmacher, C. M. Jonker, and A. Nowé. Synthesising reinforcement learning policies through set-valued inductive rule learning. *CoRR*, abs/2106.06009, 2021.
- [5] A. Cropper and S. Dumancic. Inductive logic programming at 30 : A new introduction. *J. Artif. Intell. Res.*, 74 :765–850, 2022.
- [6] L. De Raedt, A. Kimmig, and H. Toivonen. Problog : A probabilistic prolog and its application in link discovery. In *IJCAI*, volume 7, pages 2462–2467. Hyderabad, 2007.
- [7] L. De Raedt, R. Manhaeve, S. Dumancic, T. Demeester, and A. Kimmig. Neuro-symbolic= neural+ logical+ probabilistic. In *NeSy'19@ IJCAI, the 14th International Workshop on Neural-Symbolic Learning and Reasoning*, 2019.
- [8] S. Džeroski, L. De Raedt, and K. Driessens. Relational reinforcement learning. *Machine Learning*, 43(1) :7–52, Apr 2001.
- [9] P. Hitzler, A. Eberhart, M. Ebrahimi, M. K. Sarker, and L. Zhou. Neuro-symbolic approaches in artificial intelligence. *National Science Review*, 9(6) :nwac035, 2022.
- [10] W. A. Howard. The formulae-as-types notion of construction. *To HB Curry : essays on combinatory logic, lambda calculus and formalism*, 44 :479–490, 1980.
- [11] J. Lang and B. Zanuttini. Probabilistic knowledge-based programs. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 1594–1600, 2015.
- [12] M. Lezcano-Casado and D. Martinez-Rubio. Cheap orthogonal constraints in neural networks : A simple parametrization of the orthogonal and unitary group. In *International Conference on Machine Learning*, pages 3794–3803. PMLR, 2019.
- [13] Y. Liu, F. Shang, J. Cheng, H. Cheng, and L. Jiao. Accelerated first-order methods for geodesically convex optimization on riemannian manifolds. *Advances in Neural Information Processing Systems*, 30, 2017.
- [14] P. Marquis. Extending abduction from propositional to first-order logic. In *FAIR*, volume 535, pages 141–155. Springer, 1991.
- [15] L. D. Raedt, K. Kersting, S. Natarajan, and D. Poole. *Statistical Relational Artificial Intelligence : Logic, Probability, and Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2016.
- [16] W. Ring and B. Wirth. Optimization methods on riemannian manifolds and their application to shape space. *SIAM Journal on Optimization*, 22(2) :596–627, 2012.
- [17] R. S. Sutton and A. G. Barto. *Reinforcement learning - an introduction, Second Edition*. MIT Press, 2018.
- [18] T.-J.-Y.-S.-G. O.-M. Takamura-Y-Okamoto-S-Yamawaki-and-K-Doya. Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions. In *PLOS ONE 12(10)*, oct-2017.
- [19] M. Tao and T. Ohsawa. Variational optimization on lie groups, with examples of leading (generalized) eigenvalue problems. In *International Conference on Artificial Intelligence and Statistics*, pages 4269–4280. PMLR, 2020.
- [20] G. Van den Broeck, K. Kersting, S. Natarajan, and D. Poole. *An Introduction to Lifted Probabilistic Inference*. MIT Press, aug 2021.

RÉFÉRENCES COMPLÉMENTAIRES

- [21] V. Vapnik and R. Izmailov. Complete statistical theory of learning : learning using statistical invariants. In *Conformal and Probabilistic Prediction and Applications*, pages 4–40. PMLR, 2020.
- [22] H. Zhang, S. J Reddi, and S. Sra. Riemannian svrg : Fast stochastic optimization on riemannian manifolds. *Advances in Neural Information Processing Systems*, 29, 2016.
- [23] Y. Zhang, Q. Qu, and J. Wright. From symmetry to geometry : Tractable nonconvex problems. *arXiv preprint arXiv :2007.06753*, 2020.

**Trajectoire de
l'équipe A3**



Partie III
Acronymes

RÉFÉRENCES COMPLÉMENTAIRES

| | |
|----------------------|---|
| A³ | Apprentissage Artificiel et Applications |
| ANR | Agence Nationale de la Recherche |
| AOC | Algorithmes et Optimisation Combinatoire |
| ASPC | Alliance Sorbonne-Paris-Cité |
| CALIN | Combinatoire, Algorithmique et Interactions |
| CD | Contrat Doctoral |
| CIFRE | Conventions Industrielles de Formation par la Recherche |
| CNRS | Centre National de la Recherche Scientifique |
| Complexités | Axe transverse Complexités |
| CR | Chargé de Recherche |
| DR | Directeur de Recherche |
| DIM | Domaine d'Intérêt Majeur |
| EA | Equipe d'Accueil |
| EDF | Electricité de France |
| EFL | Labex <i>Empirical Foundations of Linguistics</i> |
| FR | Fédération de Recherche |
| FUI | Fonds Unique Interministériel |
| HDR | Habilitation à Diriger des Recherches |
| IG | Institut Galilée |
| IUT | Institut Universitaire Technologique |
| IUTV | IUT de Villetaneuse |
| L2TI | Laboratoire de traitement et transport de l'information, USPN - EA 3043 |
| LabCom | Laboratoire commun (Appel PIA 4) |
| LabEx | Laboratoire d'Excellence (appel Grand Emprunt) |
| LAGA | Laboratoire Analyse, Géométrie et Applications, USPN - UMR 7539 |
| LIPN | Laboratoire d'Informatique de Paris-Nord, Université Sorbonne Paris Nord - UMR 7030 |
| LoVe | Logique et Vérification |
| MathSTIC | Fédération de Recherche MathSTIC, Université Sorbonne Paris Nord - FR 3734 |
| MCF | Maître de Conférences |
| PIA | Programme d'Investissements d'Avenir |
| PU | Professeur des Universités |
| RCLN | Représentation des Connaissances et Langage Naturel |
| RFSI | Réseau Francilien en Sciences Informatiques |
| UMR | Unité Mixte de Recherche |
| USPN | Université Sorbonne Paris Nord |
| VP | Vice-Président (de USPN) |