



école doctorale **sciences pour l'ingénieur et microtechniques**



Titre de la thèse : Interprétation granulaire de données hétérogènes et multivariées / Granular interpretation of heterogeneous and multivariate data

Laboratoire d'accueil : LIB - EA 7534, 9 avenue Alain Savary, 21000 Dijon, FRANCE

Spécialité du doctorat préparé : Informatique

Mots-clefs : Intelligence artificielle, modélisation de connaissances, règles logiques, apprentissage automatique, réseaux artificiels de neurones, informatique granulaire, data science

Descriptif détaillé de la thèse :

Contexte : Priorité du gouvernement français, la lutte contre le trafic de stupéfiants est, d'une part, un enjeu de santé publique, avec chaque année, dans le monde, en moyenne 168 000 décès directement causés par l'usage de la drogue, et, d'autre part, un enjeu de sécurité, puisqu'il entraîne une dégradation des conditions de vie et des violences dans les quartiers touchés. Le ministre de l'Intérieur a placé, dès le mois de juillet 2020, ce combat contre la drogue parmi ses 3 priorités. Les connaissances des produits qui circulent en France sont rassemblées dans la base de données nationale STUPS© (Système de Traitement Uniformisé des Produits Stupéfiants) du ministère de l'Intérieur. Cette base contient des données hétérogènes et multivariées : des données macroscopiques (e.g. logos, dimensions), qualitatives (e.g. noms des agents de coupage), quantitatives (e.g. teneurs en principes actifs), mais également des données d'enquête non confidentielles (e.g. quantités saisies, date et lieu de saisie sur le territoire français). Créée en 1986, la base STUPS© est alimentée par les 5 laboratoires de Police Scientifique du Service National de Police Scientifique (SNPS) et par l'Institut de Recherche Criminelle de la Gendarmerie Nationale, et contient aujourd'hui environ 10 millions d'entrées. Présenté en septembre 2019, le Plan Stup français prévoit une série de 55 mesures, dont "La mise en place de nouveaux indicateurs pour connaître les usages des consommateurs, les méthodes des trafiquants et anticiper leurs évolutions". Or, la structure inhérente de la base STUPS© et les caractéristiques des données contenues ne permettent pas d'en extraire des connaissances (interprétation par une machine), afin de pouvoir identifier, expliquer et prédire les usages des consommateurs et les méthodes de trafiquants.

Travaux envisagés : Il s'agit de proposer un système intelligent pour répondre aux défis liés à l'interprétation de données hétérogènes et multivariées (modèles linéaires et non-linéaires) contenues dans la base de données STUPS© afin d'en décrire, comprendre et expliquer les connaissances implicites. Les travaux de recherche visés dans cette thèse concernent le domaine de l'Intelligence Artificielle (IA), et s'orienteront sur deux aspects fondamentaux : IA symbolique (modèles de connaissances définissant les sémantiques – Motik et al., 2012 – et autres aspects symboliques permettant d'interpréter et de raisonner sur ces connaissances – Motik 2006), d'une part, et, IA statistique (modèles d'apprentissage automatique de type réseaux artificiels de neurones – Bishop 1995 – permettant de construire des prédictions), d'autre part. Les recherches envisagées exploreront l'articulation de ces approches IA avec des approches granulaires (Mani 1998). En effet, selon Hobbs (Hobbs 1985), la capacité de conceptualiser le monde à différents niveaux et de bénéficier d'une mobilité totale entre ces niveaux est une caractéristique exclusive de la résolution humaine de problèmes. En effet, lorsque nous regardons le monde qui nous entoure nous n'en retirons que les choses qui servent nos intérêts du moment. Dans le cadre de cette thèse, nous investiguerons l'application de la théorie de la granularité de Hobbs au modèle de connaissances constitué, afin de permettre un raisonnement à différents niveaux de granularité.

Les problématiques de recherche adressées sont :

- comment intégrer de manière consistante et cohérente au sein d'une base de connaissances (ontologie) des données hétérogènes et inconsistantes dans le temps ?
- comment exploiter des résultats obtenus d'algorithmes d'apprentissage automatique pour améliorer la description des connaissances ?
- comment interpréter et raisonner sur les données ainsi intégrées de manière à déduire de nouvelles connaissances ?
- comment maximiser l'efficacité de l'approche ainsi spécifiée ?

Références bibliographiques :

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford University Press

Hobbs, J.R. (1985). Granularity. In Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Angeles, CA

Mani, I. (1998). A theory of granularity and its application to problems of polysemy and underspecification of meaning. In A. G. Cohn, L. K. Schubert, & S. C. Shapiro, eds. Principles of knowledge representation and reasoning, Proceedings of the Sixth International Conference (KR98). San Mateo: Morgan Kaufmann

Motik, B. (2006). Reasoning in description logics using resolution and deductive databases.

Motik, B., Patel-Schneider, P.F., Grau, B.C. (2012). OWL 2 Web Ontology Language Direct Semantics (Second Edition) W3C Recommendation. Available at: <https://www.w3.org/TR/owl-direct-semantics/>.

Profil demandé : Les candidats doivent être titulaires d'un diplôme d'ingénieur informatique ou d'un Master 2 en informatique. La maîtrise de la langue française est indispensable (niveau min. C1). Un bon niveau en communication anglaise est un plus. Les candidats doivent avoir un intérêt pour la recherche. Constituent un plus des compétences en ingénierie des connaissances (Web sémantique, ontologies) et/ou data science.

Financement : MESRI établissement

Dossier à envoyer pour le 30/06/2022

Période d'auditions: au fil de l'eau

Début du contrat : 1 octobre 2022

Salaire mensuel brut : 1975€

Direction / codirection de la thèse :

Ana ROXIN MCF-HDR (direction)

Ludovic JOURNAUX MCF (co-directeur)

Laurence DUJOURDY IR (co-directeur)

Contact:

Dr Ana ROXIN (directrice) : ana-maria.roxin@u-bourgogne.fr

Dr Ludovic JOURNAUX (co-directeur) : ludovic.journaux@agrosupdijon.fr

Dr Laurence DUJOURDY (co-directrice) : laurence.dujourdy@agrosupdijon.fr

Les candidats sont invités à soumettre leur candidature aux directeurs de thèse.

La candidature doit contenir les documents suivants :

- CV

- une lettre de motivation

- Au moins 1 lettre de référence



école doctorale **sciences pour l'ingénieur et microtechniques**

SPIM

PhD title: Granular interpretation of heterogeneous multivariate data

Host laboratory: LIB - EA 7534, 9 avenue Alain Savary, 21000 Dijon, FRANCE

The speciality of PhD: Computer science

Keywords: Artificial intelligence, knowledge modelling, logical rules, automatic learning, artificial neural networks, granular computing, data science

Job description:

Context: A priority of the French government, the fight against drug trafficking is, on the one hand, a public health issue, with each year, in the world, on average, 168,000 deaths directly caused by the use of drugs, and, on the other hand, a security issue, since it leads to a deterioration of living conditions and violence in the neighbourhoods affected. As of July 2020, the Ministry of the Interior has placed this fight against drugs among its three priorities. Knowledge of the products circulating in France is gathered in the national database STUPS© (Standardized Treatment System for Narcotic Products) of the Ministry of the Interior. This database contains heterogeneous and multivariate data: macroscopic data (e.g. logos, dimensions), qualitative (e.g. names of blending agents), quantitative (e.g. content of active ingredients), but also non-confidential survey data (e.g. quantities seizures, date and place of seizure on French territory). The STUPS© database was created in 1986 and is now supplied by the 5 Scientific Police laboratories of the National Scientific Police Service (SNPS) and by the Criminal Research Institute of the National Gendarmerie, and today contains around 10 million entries. Presented in September 2019, the French Stup Plan provides a series of 55 measures, including "The implementation of new indicators to know the uses of consumers, the methods of traffickers and anticipate their evolution". However, the inherent structure of the STUPS© database and the characteristics of the data contained do not make it possible to extract knowledge from it (ensuring a machine interpretation) nor to identify, explain and predict the uses of consumers and the methods of traffickers.

Research orientations: The aim is to propose an intelligent system to meet the challenges related to interpreting heterogeneous and multivariate data (linear and non-linear models) contained in the STUPS© database to extract, understand and explain the implicit knowledge. The research work targeted in this thesis concerns the field of Artificial Intelligence (AI) and will focus on two fundamental aspects: symbolic AI (knowledge models defining the semantics – Motik et al., 2012 – and other symbolic aspects allowing to interpret and reason on this knowledge – Motik 2006), on the one hand, and, statistical AI (automatic learning models of the artificial neural network type – Bishop 1995 – allowing to build predictions), on the other hand. The planned research will explore the articulation of these AI approaches with granular approaches (Mani 1998). Indeed, according to Hobbs (Hobbs 1985), the ability to conceptualize the world at different levels and to allow complete mobility between these levels is an exclusive feature of human problem-solving. Indeed, when we look at the world around us, we only take out the things that serve our interests at the moment. As part of this thesis, we will investigate the application of Hobbs' theory of granularity to the constituted knowledge model to allow reasoning at different levels of granularity.

The research questions addressed are:

- how to integrate consistently and coherently within a knowledge base (ontology) heterogeneous and inconsistent over time data?
- how to exploit the results obtained from automatic learning algorithms to improve the description of knowledge?
- how to interpret and reason on the data thus integrated to deduce new knowledge?
- how to maximize the effectiveness of the approach thus specified?

References:

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford University Press

Hobbs, J.R. (1985). Granularity. In Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Angeles, CA

Mani, I. (1998). A theory of granularity and its application to problems of polysemy and underspecification of meaning. In A. G. Cohn, L. K. Schubert, & S. C. Shapiro, eds. Principles of knowledge representation and reasoning, Proceedings of the Sixth International Conference (KR98). San Mateo: Morgan Kaufmann

Motik, B. (2006). Reasoning in description logics using resolution and deductive databases.

Motik, B., Patel-Schneider, P.F., Grau, B.C. (2012). OWL 2 Web Ontology Language Direct Semantics (Second Edition) W3C Recommendation. Available at: <https://www.w3.org/TR/owl-direct-semantics/>.

Candidate Profile: Applicants for this position must have either an engineer or a Master 2 diploma in computer science. Applicants must have a good level of French (min. C1). English is a plus.

Applicants should be interested in research activities in general. Skills in knowledge engineering (Semantic Web, ontologies) and/or data science are a plus.

Financing Institution: MESRI

Application deadline: 30th of June 2022

Starting date: October 2022

Gross monthly salary: 1975€

Supervisor(s):

Ana ROXIN MCF-HDR (lead), Ludovic JOURNAUX MCF, Laurence DUJOURDY IR

Contact:

Dr Ana ROXIN (director): ana-maria.roxin@u-bourgogne.fr

Dr Ludovic JOURNAUX (co-director): ludovic.journaux@agrosupdijon.fr

Dr Laurence DUJOURDY (co-director): laurence.dujourdy@agrosupdijon.fr

Applicants are invited to submit their application to the PhD supervisors.

Application must contain the following documents:

-CV

-Cover letter

-At least 1 reference letter