

Avis de Soutenance

Monsieur Mathieu GUILBERT

Informatique

Soutiendra publiquement ses travaux de thèse intitulés

Clustering sous contraintes et explicables fondés sur des méthodes ensemble et des modèles déclaratifs

dirigés par Madame Christel VRAIN et Madame Thi Bich Hanh DAO

Ecole doctorale : Mathématiques, Informatique, Physique Théorique et Ingénierie des Systèmes - MIPTIS

Unité de recherche : LIFO - Laboratoire d'Informatique Fondamentale d'Orléans

Soutenance prévue le **vendredi 09 janvier 2026** à 9h00

Lieu : 5 Rue du Carbone, 45100 Orléans

Salle : Amphi IRD

Composition du jury proposé

Mme Christel VRAIN	Université d'Orléans	Directrice de thèse
Mme Thi-Bich-Hanh DAO	Université d'Orléans	Co-directrice de thèse
M. Marcilio PEREIRA DE SOUTO	Université d'Orléans	Co-encadrant de thèse
Mme Céline ROUVEIRO	Université Sorbonne Paris Cité	Rapporteure
M. Julien ALIGON	Université Toulouse Capitole	Rapporteur
M. Ronan BUREAU	Université de Caen Normandie	Examinateur
M. Bruno CREMILLEUX	Université de Caen Normandie	Examinateur
Mme Marie-Jeanne LESOT	Sorbonne Université	Examinateuse

Mots-clés : Clustering, Explicabilité, Programmation par Contraintes, Méthodes ensemble, Machine Learning,

Résumé :

L'essor de l'Intelligence Artificielle Explicable (XAI) a récemment conduit à l'émergence du Clustering Explicable (XC), un sous-domaine visant à rendre les modèles de clustering plus transparents et plus interprétables. Étant donné un ensemble d'objets, le clustering consiste à regrouper des objets similaires en ensembles homogènes appelés clusters. Le clustering explicable cherche à fournir des explications du processus de clustering, que ce soit a posteriori ou lors de la conception, afin de renforcer la compréhension humaine des résultats produits et a fortiori la confiance. En parallèle, le clustering sous contraintes est une branche de l'apprentissage non supervisé permettant d'intégrer des connaissances ou préférences exprimées par l'utilisateur sous forme de contraintes. Cette supervision partielle guide le processus de clustering vers des résultats plus cohérents avec l'expertise humaine, tout en favorisant l'interactivité et la personnalisation du modèle. Ces deux paradigmes (explicabilité et contraintes) partagent un objectif commun : rapprocher la logique du modèle de celle de l'utilisateur. Pourtant, les approches existantes demeurent limitées dans leur expressivité et leur généralité. Les méthodes de clustering sous contrainte se restreignent souvent à des contraintes binaires simples, tandis que les modèles explicables par conception peinent à concilier qualité du clustering et qualité des explications. De plus, le potentiel des explications de clustering fondées sur des exemples ainsi que celui des représentations à deux espaces reste largement inexploré. Cette thèse aborde ces limitations à travers trois contributions complémentaires qui, ensemble, font progresser les domaines du clustering explicable et sous contraintes. Premièrement, notre méthode post-hoc Anchored Constrained Clustering Ensemble (ACCE) étend le cadre du clustering ensemble. Le clustering ensemble combine plusieurs clusterings pour produire une partition consensus plus robuste. Notre nouvelle méthode permet d'intégrer de nouvelles formes de contraintes jusqu'alors absentes du domaine. En formulant la satisfaction des contraintes comme un problème d'optimisation linéaire en nombres entiers, ACCE ajuste minimalement les partitions consensus. Nous introduisons le concept d'ancres, des objets représentatifs de leurs clusters, servant de points de référence lors du processus de modification minimale. Deuxièmement, la méthode Explainable Constrained Clustering with Frequent Patterns (ECS) introduit un cadre explicable par conception reposant sur une configuration à double représentation. Celle-ci sépare l'espace de caractéristiques utilisé pour le clustering de l'espace symbolique employé pour les explications. Grâce à une formulation en programmation par contraintes, ECS sélectionne conjointement les clusters et leurs explications exprimées sous forme de motifs booléens fréquents et discriminants, assurant ainsi la couverture, la compacité et la distinction des explications. Troisièmement, la méthode Example-based Explainable Constrained Clustering (EECS) est une approche de clustering explicable fondée sur les exemples. Comme la méthode précédente, elle sélectionne

conjointement les clusters et leurs explications. Elle identifie, pour chaque cluster, un ensemble restreint d'objets représentatifs (ou prototypes) servant d'explications naturelles et intuitives, tout en intégrant des contraintes utilisateur. Ces prototypes peuvent servir d'ancres telles que nous les avons définies dans ACCE. Ces approches établissent un cadre reliant intégration des contraintes et interprétabilité dans l'apprentissage non supervisé. Leur pertinence est démontrée sur des jeux de données synthétiques et réels, notamment à travers une application en chémoinformatiche portant sur le clustering de molécules.