Avis de Soutenance

Madame Ryma BOUMAZOUZA

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés

Modèles prédictifs & raisonnement avec les explications

dirigés par Monsieur Bertrand MAZURE et Monsieur Karim TABIA

Soutenance prévue le ***jeudi 08 décembre 2022*** à 13h30
Lieu : UFR des Sciences Jean Perrin Rue Jean Souvraz SP 18 F-62307 Lens Cedex France
Salle : des thèses

Composition du jury proposé

M. Bertrand MAZURE	Université d'Artois	Directeur de thèse
M. Karim TABIA	Université d'Artois	Co-directeur de thèse
Mme Marie-Jeanne LESOT	Sorbonne Université - LIP6	Rapporteure
Mme Fahima CHEIKH- ALILI	Université d'Artois	Examinatrice
M. Sylvain LAGRUE	Université de Technologie de Compiègne	Rapporteur
Mme Christine SOLNON	INSA de Lyon	Examinatrice

Résumé :
Cette thèse étudie une méthode d'explicabilité qui allie à la fois le caractère "agnostique" des méthodes numériques et qui propose des explications plus "rigoureuses" qui caractérisent les explications symboliques. Le but étant d'expliquer les prédictions des techniques de classification mono-étiquette et multi-étiquettes. Plusieurs contributions sont apportées dans cette thèse. Premièrement, nous avons travaillé sur le cas mono-étiquette. Nous avons proposé une approche qui va de l'encodage en représentation symbolique du modèle dont on souhaite expliquer les prédictions à la génération d'explication basée sur un oracle SAT. L'idée est de prendre un classificateur, avec une instance, et de produire une formule propositionnelle que nous utiliserons pour générer nos explications. L'inconsistance de cette formule permet d'expliquer les prédictions négatives. Nous considérons les deux cas où nous pouvons avoir la représentation logique du modèle dans son ensemble ou une approximation basée sur un modèle de substitution. Nous nous intéressons à deux types complémentaires d'explications symboliques : les raisons suffisantes qui correspondent à un sous-ensemble minimal de l'entrée conduisant à une prédiction spécifique et les contrefactuelles qui correspondent à un sous-ensemble de l'entrée permettant de connaître les modifications minimales à apporter pour obtenir une prédiction différente. Deuxièmement, nous avons proposé des propriétés à considérer afin de prioriser et sélectionner les explications en évaluant leur pertinence ainsi que celle des variables les composants. Par la suite, nous nous sommes intéressés à l'explication des prédictions multi-étiquettes. Nous avons proposé des explications multi- étiquettes à différents niveaux de granularité et étudié la combinaison d'explications mono-étiquette ainsi que les relations structurelles entre classes comme moyen de les générer. Enfin, nous nous sommes intéressés aux scores d'importance au niveau des caractéristiques pour déterminer dans quelle mesure chacune contribue à la sortie d'un modèle multi-étiquettes. Cette contribution examine deux possibilités différentes d'utiliser des méthodes existantes pour le single-label comme oracles ou d'utiliser des attributions de caractéristiques obtenues à partir d'explications symboliques. Afin d'évaluer la qualité des attributions de caractéristiques, nous étendons les propriétés de sensibilité, de stabilité des données au cas multi-étiquettes en plus d'une nouvelle propriété spécifique à la classification multi- étiquettes que nous appelons corrélation label-explication.