INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES

Campagne « Contrats Doctoraux Ordinaires » - INSA RENNES 2022

Catégorie demande	✠ contrat doctoral 100 % ☐ contrat doctoral 100 % situation handicap (compléter et joindre le dossier de candidature joint) ☐ contrat doctoral 50 %
Si cofinancement nature du cofinancement**	**(partenaire/organisme, montant cofinancement, nature du contrat, durée, calendrier)**
Titre du sujet de thèse	Indiquez-moi vos préférences et je vous montrerai ce qui vous intéresse dans les données
Unité/équipe	☐ Inst. FOTON ☐ ISCR-CSM ☐ ISCR-CM ☐ ISCR-COrint ☐ IETR ✠ IRISA ☐ IRMAR ☐ LGCGM ☐ LS2N/MECAPROCE
Direction de la thèse/Codirection (HDR)	Peggy Cellier (INSA) Alexandre Termier (UR1)
Co-encadrement	Tassadit Bouadi (UR1) Bruno Crémilleux (Unievrsité de Caen Normandie)
Cotutelle / co-direction internationale envisagée* ?	**non**
Description des objectifs scientifiques du projet de thèse	Les méthodes de fouille de données ont pour objectif d'aider l'utilisateur dans sa compréhension des données en découvrant des modèles utiles, inattendus et intéressants pour celui-ci. Pour une recherche fructueuse de tels modèles, il est nécessaire de prendre en compte les attentes de l'utilisateur. Il existe plusieurs formalismes pour capturer ces attentes. Dans cette thèse, nous nous intéressons aux notions de préférences et d'intérêt subjectif de l'utilisateur, deux formalismes encore peu étudiés en fouille de données. Ce travail porte plus particulièrement sur la découverte de modèles particuliers : les sous-groupes exceptionnels/inattendus. Un sous-groupe est un ensemble de données caractérisé par un comportement spécifique. Un sous-groupe exceptionnel est un sous-groupe dont le comportement est considéré comme surprenant car il dévie de la norme (p. ex. comportement global d'une population) ou parce qu'il contredit les attentes et croyances de l'utilisateur (p. ex. un groupe de consommateurs ayant un comportement d'achat distinct de celui de consommateurs au profil similaire). Dans la littérature, deux familles de méthodes portent sur l'extraction de sous-groupes exceptionnels. L'Exceptional Preference Mining (EPM) [1] recherche, à partir de préférences individuelles d'une population donnée, des sous-groupes dont les préférences s'écartent de celles de l'ensemble de la population. Par exemple, identifier des sous-groupes d'individus dont les préférences en terme de consommation d'une catégorie de produit divergent ou contredisent les préférences de la population globale. Les préférences peuvent être exprimées sous de multiples formes [2] (quantitative comme des notes/scores relatives à des mesures d'intérêt, comparaisons de paires ou d'autres formalismes plus sophistiqués). Dans ce travail, nous partons de préférences exprimées sous formes de relations d'ordres (par exemple, si on considère cinq boissons, l'utilisateur les a ordonnées de celle qu'il préfère le plus à celui qu'il préfère le moins). La deuxième famille est la fouille de données selon l'intérêt subjectif de l'utilisateur [3] : cette approche consiste à rechercher des sous-groupes

surprenants par rapport à une connaissance du domaine qui est un a priori sur les données. Par exemple, on s'attend à ce que la température du lendemain soit proche de celle du jour courant car un brusque changement de température n'est pas la situation usuelle. Si il fait aujourd'hui une température de 23 degrés et si on indique à l'utilisateur que la température prévue pour le lendemain est 35 degrés, celui-ci est alors fortement intéressé. Autrement dit, on maximise la préférence de l'aspect inattendu de l'information découverte par rapport à la connaissance du domaine.

L'objectif de cette thèse est double. D'une part, il s'agit d'étudier ce qui caractérise et différencie ces deux approches et de proposer une méthode les combinant afin de découvrir des sous-groupes maximisant l'intérêt subjectif de l'utilisateur tout en intégrant les préférences explicites exprimées par ce dernier. Cette combinaison est nouvelle et permettra d'offrir un formalisme plus général d'expression de l'intérêt de l'utilisateur. D'autre part, il s'agira d'inscrire ce formalisme d'expression de l'intérêt de l'utilisateur dans un procesus de fouille interactive [4]. Ce dernier, en prenant en compte un retour utilisateur sur l'information extraite, permet un processus plus fructueux d'exploration des données. Pour cela, il sera nécessaire de réaliser une méthode prenant en compte des retours utilisateurs pour l'intérêt subjectif et l'exceptional model mining. Dans le cas de l'intérêt subjectif, on sait réévaluer sa valeur dans la cadre d'une démache itérative [5] ce qui est une première étape pour développer une telle méthode.

Références

- [1] Cláudio Rebelo De Sá, Wouter Duivesteijn, Paulo Azevedo, Alípio MárioJorge, Carlos Soares, Arno Knobbe, "Discovering a taste for the unusual:exceptional models for preference mining", Machine Learning, 2018.
- [2] Meltem Öztürké, Alexis Tsoukiàs, and Philippe Vincke. "Preference modelling." Multiple criteria decision analysis: State of the art surveys. Springer, New York, NY, 2005. 27-59.
- [3] Tijl De Bie "Maximum entropy models and subjective interestingness: an application to tiles in binary databases" Data Mining Knowledge Discovery 2011.
- [4] Matthijs Van Leeuwen "Interactive Data Exploration using Pattern Mining. In: Holzinger, A & Jurisica, I (eds) Interactive Knowledge Discovery and Data Mining: State-of-the-Art and Future Challenges in Biomedical Informatics, LNCS 8401, Springer, 2014.
- [5] Matthijs Van Leeuwen, Tijl De Bie, Eirini Spyropoulou, Cédric Mesnage "Subjective interestingness of subgraph patterns" de Machine Learning 2016.

**** joindre justificatifs associés***

Je déclare exactes toutes les informations contenues dans ce document
Date et signature du responsable scientifique du projet :

Avis du directeur d'unité :	Avis du responsable de composante :
Date et signature :	Date et signature :

Classement du responsable de composante