

Généricité et explicabilité dans les systèmes de recommandation.

Localisation : Université Paris-Dauphine, LAMSADE

Encadrante : Elsa Negre, MCF HDR (elsa.negre@lamsade.dauphine.fr)

Mots-clés : Systèmes de recommandation, *Machine learning*, Systèmes d'Informations.

Financement : Ce sujet est prioritaire pour une demande de bourse de thèse du ministère.

Candidatures : Les candidats intéressés sont invités à envoyer une lettre de motivation, un CV, leurs relevés de notes (Licence + Master) avec classements (Master 2 également, éventuellement partiel), et une ou plusieurs lettres de recommandation, le plus tôt possible, **avant le 27 mars 2020**. Une seconde sélection sera ensuite faite par l'école doctorale et le laboratoire sur la base d'une audition en mai 2020.

Contexte :

La recommandation peut se résumer par le problème d'estimation de scores pour des entités qui n'ont pas encore été vues/évaluées par un utilisateur. En effet, le nombre d'entités ainsi que le nombre d'utilisateurs du système peuvent être très importants ; il est, de ce fait, difficile que chaque utilisateur voit toutes les entités ou que chaque entité soit évaluée par tous les utilisateurs. Il est donc nécessaire d'estimer les scores pour les entités non encore évaluées. Cette évaluation est habituellement basée sur les scores donnés par un utilisateur à d'autres entités. Lorsqu'il est possible d'estimer les scores pour les entités non encore évaluées, les entités ayant les scores les plus élevés peuvent être recommandées à l'utilisateur.

Les systèmes de recommandation s'appliquent dans de nombreux domaines, que ce soit dans le cadre d'applications commerciales, industrielles ou académiques. En fait, parmi les systèmes de recommandation les plus connus, nous pouvons citer ceux utilisés dans le commerce électronique : Amazon.com pour la recommandation de produits et Netflix pour la recommandation de films. Mais de nos jours, de nombreux systèmes que nous utilisons tous les jours proposent des recommandations à leurs utilisateurs (des groupes de travail ou des individus dans LinkedIn, des amis sur Facebook, de la musique pour last.fm ou des nouvelles pour Forbes.com). Les systèmes de recommandation existants diffèrent donc par leur portée, leur contexte d'application ainsi que par les données qu'ils manipulent. Mais ils ont également de nombreux points communs : (i) le besoin : aider à la prise de décision ; (ii) l'objectif de recommander des éléments/utilisateurs ; (iii) la formalisation : la fameuse matrice d'utilité ; (iv) l'algorithme : prédire les scores. Pourtant, malgré leurs similitudes, les systèmes de recommandation existants sont spécifiques à une application et sont développés/mis en œuvre via des cadres ad hoc.

Or, les systèmes de recommandation, comme tout système informatisé, sont régis par les principes de l'ingénierie et de la qualité logicielle. Le génie logiciel repose sur sept principes : la rigueur, la décomposition en sous-problèmes, la modularité, l'abstraction, l'anticipation des évolutions, la généricité (selon laquelle un système réutilisable/adaptable a beaucoup plus de valeur qu'un système dédié) et une construction incrémentale. De plus, la norme ISO 9126 définit six groupes d'indicateurs de qualité logicielle : la capacité fonctionnelle, la facilité d'utilisation, la fiabilité, la performance, la maintenabilité et la portabilité. Actuellement, les systèmes de recommandation sont définis pour des cas d'application spécifiques, ce qui limite leur adaptabilité, leur réutilisabilité et leur généricité. Ainsi, la spécialisation des systèmes de recommandation va à l'encontre du principe de généricité et donc, de l'ingénierie et de la qualité logicielle. Par conséquent, tendre vers la généricité des systèmes de recommandation, c'est-à-dire un système de recommandation qui fonctionne quel que soit le cas d'application afin d'avoir un système adaptatif, avec un niveau d'abstraction, favorisant, entre autres, l'interopérabilité et la réutilisabilité est un enjeu important. Or, à notre connaissance, il n'existe pas, à ce jour, un tel système de recommandation.

Par ailleurs, l'interaction du système de recommandation avec l'utilisateur est importante. En général, les utilisateurs souhaitent avoir un contrôle sur les recommandations faites et pouvoir indiquer si une recommandation ne leur convient pas. Cette démarche s'inscrit dans la recherche de systèmes plus centrés sur l'utilisateur. De plus, le fait de donner une explication personnalisée à l'utilisateur peut également nous aider à gagner sa confiance. En effet, réussir à indiquer la raison

pour laquelle une recommandation est proposée à l'utilisateur peut contribuer à améliorer la confiance qu'il a du système mais aussi, si l'utilisateur peut interagir avec le système de recommandation, améliorer le système lui-même en mettant en adéquation les raisons d'une telle recommandation et les raisons d'accord ou de refus de celle-ci par l'utilisateur. Des travaux se sont déjà intéressés à l'explication des recommandations. Cependant, nombre d'entre eux se limitent à afficher des valeurs numériques ou à donner une explication du type « Ceux qui ont acheté A ont aussi acheté B ». Être capable d'expliquer la recommandation soulève des problèmes tels qu'expliquer comment fonctionne le système de recommandation, permettre aux utilisateurs d'indiquer au système de recommandation que la recommandation ne convient pas, convaincre les utilisateurs de suivre la recommandation, améliorer la facilité d'utilisation. Une explication du type « Ce livre appartient à votre genre préféré : Thriller/Policiers, il a obtenu une note moyenne de 4/5 auprès des autres utilisateurs et il est disponible à la librairie de votre quartier » pourrait être une meilleure explication. Or, à notre connaissance, il n'existe pas, à ce jour, un tel système de recommandation.

L'objectif principal de cette thèse est d'étudier la diversité des **systèmes de recommandation** , leurs points communs et différences (d'un point de vue algorithmique mais aussi applicatif) dans un contexte de **grande masse de données en constante évolution** , ainsi que de comprendre de tels systèmes dans leur contexte. Il s'agira ensuite de tendre vers un modèle de système générique de recommandation capable d'expliquer à l'utilisateur les recommandations retournées.

Remarque :

Une piste possible serait, par exemple d'incorporer différentes informations complémentaires à partir de l'analyse des médias sociaux (entre autres), comme cela se fait dans les systèmes de recommandation multi-domaines. En effet, la prolifération des sites de commerce électronique, des médias sociaux, ... a permis aux utilisateurs de fournir des commentaires, d'exprimer leurs préférences/intérêts et de maintenir des profils utilisateurs dans de multiples systèmes, reflétant la variété de leurs goûts/intérêts. Tirer parti de toutes ces informations disponibles dans différents systèmes et relatives à différents champs/spécialités peut être bénéfique pour générer des profils utilisateurs plus complets et de meilleures recommandations, par exemple, en atténuant les problèmes de démarrage à froid ou de faible densité dans un champs cible ou en proposant des recommandations personnalisées « croisées » pour des éléments de champs différents. Les systèmes de recommandation multi-domaines (*cross-domain recommender systems*) visent à générer ou à améliorer des recommandations pour un champs particulier en exploitant les profils utilisateurs (ou toutes autres données/informations) issues d'autres champs.

Implications :

Ce travail pourra s'appuyer sur les travaux de thèse de Zahra Vahidi Ferdousi (systèmes de recommandation contextuels), sur les travaux de recherche de l'encadrante ainsi que de certains membres du LAMSADE (notamment les membres de l'équipe « Science des Données », spécialisés en *Machine learning* et systèmes de recommandation).

Publications de l'encadrante en rapport avec le sujet :

- Zahra Vahidi Ferdousi, Dario Colazzo, Elsa Negre. *CBPF: Leveraging Context and Content Information for Better Recommendations*. ADMA 2018: 381-391
- Elsa Negre, Franck Ravat, Olivier Teste. *OLAP Queries Context-Aware Recommender System*. DEXA (2) 2018: 127-137
- Imène Brigui-Chtioui, Philippe Caillou, Elsa Negre. *Intelligent Digital Learning: Agent-Based Recommender System*. ICMLC 2017: 71-76
- Elsa Negre. *Systèmes de recommandation : généralité, évaluation et améliorations* Habilitation à Diriger des Recherches, Université de Toulouse, Novembre 2017.
- Ning Wang, Marie-Hélène Abel, Jean-Paul A. Barthès, Elsa Negre. *An answerer recommender system exploiting collaboration in CQA services*. CSCWD 2016: 198-203
- Elsa Negre. *Recommender Systems – Introduction*. Livre - WILEY, 2016.
- Elsa Negre. *Towards a Knowledge (Experience)-Based Recommender System for Crisis Management*. 3PGCIC 2013: 713-718