

SOUTENANCE DE THÈSE

M. Gabriel FRISCH

Soutiendra sa thèse de doctorat sur le sujet :

Random graph clustering for collaborative filtering

Unité de recherche : Heudiasyc – UMR CNRS 7253

**Le lundi 18 octobre 2021 à 14h
A l'université technologique de Compiègne
Bâtiment Blaise Pascal - Salle GI042**

Devant le jury composé de :

M. Julien CHIQUET, directeur de recherches, membre examinateur
Institut national de la recherche agronomique, mathématiques et informatique
appliquées, Paris

M. Gérard GOVAERT, professeur, membre examinateur
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne

M^{me} Julie JOSSE, directrice de recherches, membre examinateur
Institut national de recherche en informatique et en automatique, UMR UA11
INSERM, Montpellier

M. Cédric FEVOTTE, directeur de recherches, membre rapporteur
Institut de Recherche en Informatique de Toulouse, ENSEEIHT, Toulouse

M^{me} Christine KERIBIN, maître de conférences HDR, membre rapporteur
Université Paris Saclay, institut de mathématiques, Paris

M. Yves GRANDVALET, directeur de recherche, directeur de thèse
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne

M. Jean-Benoist LEGER, maître de conférences, directeur de thèse
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne

Résumé :

Le filtrage collaboratif est une méthode qui vise à construire automatiquement des filtres personnalisés en utilisant les avis d'utilisateurs dans le but de leur proposer une liste restreinte d'objets qu'ils pourraient apprécier le plus. Les avis recueillis auprès de ces utilisateurs fournissent la base collaborative pour réaliser des prédictions sur les avis manquants.

La proportion de données manquantes, c'est à dire les avis non fournis par les utilisateurs, est généralement importante dans les systèmes de recommandation. Dans certains cas, ces données manquantes peuvent être informatives et l'ignorer peut mener à des mauvaises conclusions. La plupart des méthodes de filtrage collaboratif n'utilisent pas ces données manquantes pour en extraire de l'information. Dans cette thèse, nous voulons exploiter ces données manquantes en proposant des modèles qui supposent que les avis des utilisateurs ne manquent pas au hasard (MNAR). L'approche se base sur le modèle à blocs latents, un modèle de co-clustering génératif dont les régularités permettent de prédire les avis manquants.

L'équité de la recommandation est un des autres problèmes majeurs du filtrage collaboratif. L'équité est souvent vaguement définie comme la qualité de traiter les gens avec justesse et impartialité. Bien qu'imprécise, cette définition stipule que le traitement équitable fait référence à certains attributs protégés partagés par des groupes de personnes, tels que le sexe, l'âge, l'origine ethnique, le groupe socio-économique, etc. Ces dernières années, les travaux de recherche ont mis en évidence un manque d'équité dans les décisions prises par les algorithmes de filtrage collaboratif. Le modèle statistique que nous présentons dans cette thèse génère un co-clustering des utilisateurs et des objets du système tout en essayant de respecter une parité statistique des utilisateurs vis-à-vis de leurs attributs protégés. Nous donnons des garanties théoriques assurant une recommandation équitable, à condition que cette parité statistique des utilisateurs soit bien respectée.

Collaborative filtering is a method that aims at building automatically personalized filters by using feedback of people for the particular purpose of providing a shortlist of items that they might most enjoy. The reactions collected from people provide the collaborative basis for making predictions on other unseen items.

The datasets extracted from recommender systems have usually a high proportion of missing feedback, that is, feedback that were not provided by the users. In some situations, missing data can be informative and ignoring this information can lead to misleading conclusions. Most collaborative filtering methods do not have a principled method for extracting information from this missing data. In this thesis we take advantage of missing data by positing models of missingness that assume that the feedback are missing not at random (MNAR). The approach relies on the latent block model, a generative co-clustering model whose regularities allow to predict the missing entries.

An other issue in collaborative filtering is the fairness of recommendations. Fairness is often loosely defined as the quality of treating people with rightfulness. Although imprecise, this definition stipulates that fair treatment refers to certain sensitive attributes shared by groups of people, such as gender, age, ethnicity, socio-economic group, etc. In recent years, research has highlighted the lack of fairness in decisions made by recommendation system algorithms. The model we present in this thesis produces a co-clustering of users and items that respects a statistical parity of users with respect to the sensitive attributes. We give theoretical guarantees ensuring fair recommendations, provided that the statistical parity of users is respected.