

UNIVERSITÉ DE SHERBROOKE
DÉPARTEMENT D'INFORMATIQUE

IFT 599/799 : Sciences de données

TP #4 — Automne 2019

Recommandation par filtrage collaboratif

Ce travail vise à développer une application du filtrage collaboratif pour construire un système de recommandation. Les données à utiliser sont celles fournies par le MovieLens 100K. Ces données sont disponibles dans le répertoire public du cours. Le but de ce TP est de se familiariser avec une méthode de recommandation de base et la méthode de cross-validation pour évaluer la performance d'un système d'apprentissage.

La date pour la remise de ce TP est le lundi 9 décembre 2019. Votre remise doit comprendre un rapport et les programmes que vous aurez développés pour ce TP, de même que des données modifiées ou nouvelles, s'il y a lieu. La remise doit se faire par la **turnin** sur opus.dinf.usherbrooke.ca.

Sujet : Recommandation par filtrage collaboratif

Un répertoire de données est fourni dans le répertoire public du cours (Travaux/TP4/MovieLens). Les données à utiliser sont les `u1.base`, `u1.test`, ..., `u5.base` et `u5.test`. Les fichiers `.base` contiennent les données d'apprentissage, alors que les fichiers `.test` contiennent les données pour tester les systèmes construits. Par souci de complétude, tous les fichiers de cet ensemble de données sont inclus dans le répertoire, bien que plusieurs ne soient pas utiles pour ce TP. Vous devez lire le fichier "readme" pour connaître plus de précisions sur ces données.

Les principales tâches du TP sont les suivantes.

- Planter un système de recommandation en se basant sur la méthode par voisinage, e.g. la méthode présentée dans la page 41 des transparents du thème 5 ;
- OU planter un système de recommandation en se basant sur la méthode par la projection dans l'espace latent, celle présentée aux pages 74-77.

Incluez UNE SEULE implantation dans votre remise.

- Tester votre système (vos systèmes) sur les 5 groupes de données `.base` et `.test`. Choisir l'une des deux mesures d'évaluation présentées à la page 37 des transparents du thème 5. Donner les résultats individuels (pour chaque ensemble de données de test) et la performance moyenne des 5 tests.

Pour ce TP, votre rapport doit être bref. Une page suffit. N'oubliez pas de remettre aussi vos données si elles sont modifiées des données originales.