

FOUILLE DE DONNÉES ET MEDIAS SOCIAUX

M2 DAC

TME. Recommendation Sociale

Ce TME a pour objectif de manipuler un modèle de recommandation sociale vu en cours, le modèle SoRec [ma et al., 2008]. Ce modèle fait l'hypothèse que les utilisateurs liés dans le réseau ont tendance à juger de la même manière les mêmes items. Il s'agit alors de tirer parti de cette information supplémentaire dans une approche classique de filtrage collaboratif à base de factorisation matricielle.

1 Données

On propose de s'intéresser à un dataset de notation de films. Les films sont notés de 1 à 5 par les utilisateurs. Les utilisateurs ont des liens explicites dirigés entre eux qui représentent des liens de confiance.

Le repertoire */Vrac/lamprier/FDMS/recodata* contient les fichiers suivants:

- *u.data*: Toutes les données de jugement au format "user id | film id | rating | timestamp", avec les champs séparés par des tabulations
- *u1.train* à *u5.train* et *u1.test* à *u5.train* : 5 fichiers d'entraînement avec les 5 fichiers de test correspondant. Les fichiers d'entraînement contiennent chacun 80% des données de *u.data* et les fichiers de test 20%. Attention à ne pas tester sur un fichier de test un modèle appris avec un fichier d'entraînement de numéro différent !
- *u.links*: Les liens entre utilisateur du réseau. Chaque ligne concerne liens sortants d'un utilisateur. Une ligne commence par le numéro d'un utilisateur, suivi par tous ses successeurs dans le réseau, séparés par des tabulation.

2 Modèle SoRec

Le modèle SoRec [ma et al., 2008] cherche à définir une représentation permettant d'expliquer à la fois les scores donnés aux items ainsi que les liens entre les utilisateurs.

