

Recommandation par filtrage collaboratif

L'objectif de ce tp est de pouvoir recommander des films à un utilisateur en évaluant la note que ce dernier lui donnerait, et ce en se basant sur les évaluations d'utilisateurs disposants d'un profil similaire.

Récupération des voisins possibles

Dans un premier temps, il s'agit de réduire la liste d'utilisateurs à un sous-ensemble représentant les voisins possibles. Pour ce faire, les seules personnes pouvant être possiblement proche de l'utilisateur doivent nécessairement partager un ou plusieurs films avec ce dernier.

Récupération des K plus proches voisins

Un fois le sous-ensemble de voisins possibles récupérés, il faut encore réduire l'ensemble au K plus proches voisins. Afin de limiter l'ensemble, l'utilisation de la *corrélation de Pearson* permet d'ordonner les utilisateurs afin de ne garder que les K meilleures candidatures.

Prédiction du *rating*

Finalement, le *rating* peut être estimé de la manière suivante: Pour chaque voisin faisant partie du sous-ensemble K, vérifier qu'il dispose du film que l'on tente d'évaluer. Si tel est le cas, ajouter le *rating* du voisin, pondéré par son poids (qui varie en fonction de son degré de similitude avec l'utilisateur pour lequel on veut faire une recommandation).

Mesure de la qualité de l'algorithme

Afin de mesurer la qualité de l'algorithme utilisé, l'estimation de toutes les données de test peut être effectuée. Pour chaque *rating* estimé, l'ajouter dans le calcul de la mesure de qualité qui somme la différence entre le *rating* réel et celui qui est nouvellement estimé. Le résultat sera une valeur comprise entre les valeurs minimum et maximum que peuvent prendre une évaluation.

Analyse des résultats

[TODO] : Pour l'instant, l'algorithme lancé deux fois sur les mêmes données ne donne pas un indice de qualité constant. De plus, il n'est pas optimisé -> sortir le calcul des corrélations de Pearson de la boucle (créer un dictionnaire et ne le faire qu'une fois par user!)