

Projet MDI 343
Systèmes de recommandation

Nicolas Keriven et Jean-Baptiste Alayrac

29 avril 2014

Table des matières

Introduction	2
1 Présentation du problème et notations	3
1.1 Différentes méthodes	3
1.2 Factorisation de matrice	4
1.3 Notre approche et notations	4
2 Algorithmes	4
2.1 Descente de gradient stochastique simple	4
2.2 Algorithme parallélisé Jellyfish	4
3 Résultats	4

Introduction

1 Présentation du problème et notations

Le cadre général de la problématique de recommandation est de relier des utilisateurs à des produits. Dans cette partie nous présentons tout d'abord les méthodes les plus utilisées actuellement pour les système de recommandations avant de nous concentrer sur la méthode que nous avons approfondie.

1.1 Différentes méthodes

Actuellement, il existe deux principales approches afin de créer un système de recommandation. La première approche est appelée *content filtering* alors que la seconde est nommée *collaborative filtering*

Content filtering (ou filtrage par contenu)

Cette méthode vise à caractériser la nature de chaque utilisateur ainsi que de chaque produit afin d'en dégager un profil le plus précis possible. Pour un utilisateur, cela peut être son pays d'origine, son âge... Pour un produit, par exemple un film, cela peut être son genre, son box-office... Les algorithmes ont alors pour tâche d'associer des profils d'utilisateurs avec des profils de produits. Le projet Music Genome Project, utilisé notamment dans la radio Pandora.com utilise de telles méthodes afin de proposer du contenu aux utilisateurs. Chaque musique est alors caractérisée par des centaines d'attributs assimilable à des "gènes".

Collaborative filtering (ou filtrage collaboratif)

L'autre méthode se base essentiellement sur des avis passés (implicites ou explicites) d'utilisateurs à propos de produits. La méthode de *collaborative filtering* analyse les liens qui existent entre utilisateurs et produits via la connaissance de votes passés afin de proposer de nouvelles associations utilisateur/produit. L'avantage principal de cette méthode par rapport à celle de *content filtering* est qu'ici il n'y a nul besoin de connaître le produit ou les utilisateurs à priori. Par contre cette méthode connaît le problème du *cold start*, i.e., elle est incapable de traiter l'arrivée d'un nouveau produit ou d'un nouveau utilisateur. Afin de la faire fonctionner il faut posséder un nombre déjà conséquents d'avis de chaque utilisateurs et pour chaque produits.

Dans ce contexte de *collaborative filtering* il existe à nouveau deux principales méthodes. La première est connue sous le nom de *neighborhood method*, la deuxième se concentre autour de *latent factor models*.

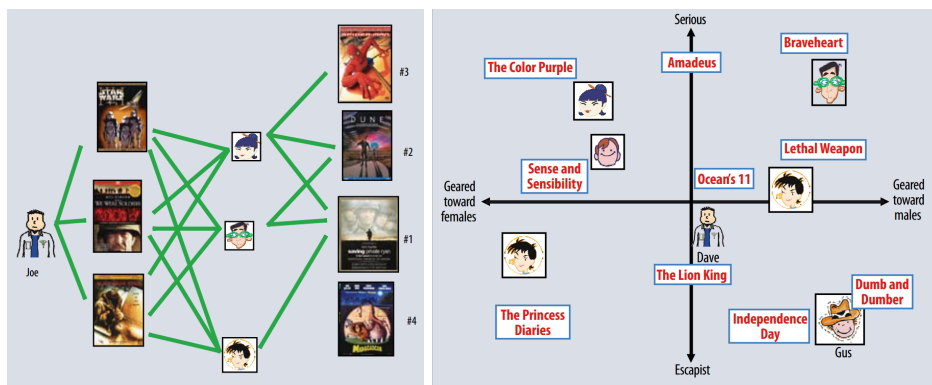


FIGURE 1 – A gauche : *neighborhood methods*, A droite : *Latent factor models*

Neighborhood methods

Latent factor model

1.2 Factorisation de matrice

1.3 Notre approche et notations

2 Algorithmes

2.1 Descente de gradient stochastique simple

2.2 Algorithme parallélisé Jellyfish

3 Résultats

Conclusion