# Collaborative Filtering with Constrained Aspect Model

Nicolas Delannay

Université catholique de Louvain (Belgium) – Machine Learning Group

www.ucl.ac.be/mlg

## Goal

Identify patterns by Collaborative Filtering in a user-item explicit rating matrix.



Rating scale
[1 = poor, 2 = …, 5 = excellent]
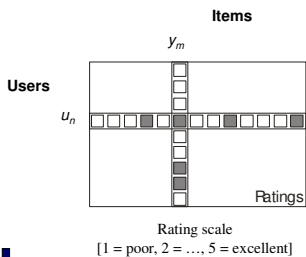
### Main application

Recommendation of new items to a user after seeing is previous ratings.
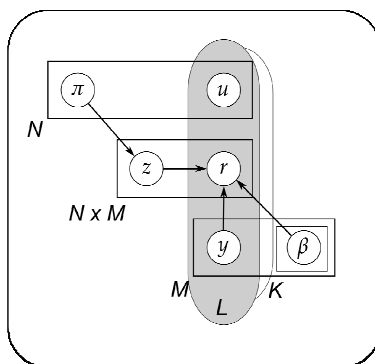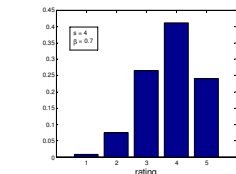
## Aspect Model

### Assumptions:

- There must be a set of rating profiles each characterizing an aspect of the rating procedure (e.g. an aspect for the movie database could stand for the quality of action scenes).
- Users have their own rating procedure represented by a mixture of the different aspects.

### Probabilistic modeling:

- K aspects (fixed a priori)

- For each user $u_n$ corresponds $\pi_n$, the aspect proportions.

- An aspect is defined by a distribution of rating for each items $y_m$. For example, a binomial distribution

$$P(r|y = m, z = k) \sim Bn(\beta_{mk}, s)$$



$z$ treated as hidden variable, $\{\pi, \beta\}$ adjusted by an EM algorithm

## Problem

- Typically, number of items > $10^2$ and rating matrix 95% sparse ➔ Model too flexible, overfitting.

- Aspects are difficult to interpret because they cover the entire item set.
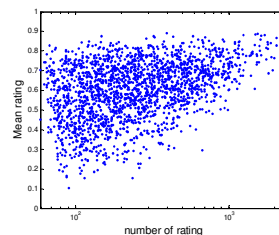
## Proposed solution

- Reduce the flexibility by adding a constraint on the rating distributions.

- Model the activity (frequency of observed ratings) of an item for an aspect.
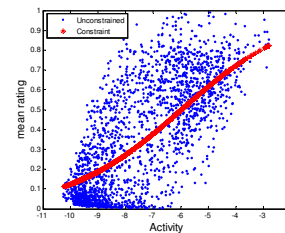
### Idea:

For each aspect, the satisfaction of a movie expressed by higher rating should be correlated with the activity

Observed correlation in movieLens database

Adjusted model parameters



## Experiment

Illustration on the movieLens dataset:

- Adjust the constrained model for 5 aspects.

- For each aspect, the 10 highest rated movies are given below (the first line qualifies the set of movies).

**Blockbuster**
Star Wars: Episode IV - A New Hope (1977)
Braveheart (1995)
Forrest Gump (1994)
Schindler's List (1993)
Terminator 2: Judgment Day (1991)
Toy Story (1995)
Shawshank Redemption, The (1994)
Silence of the Lambs, The (1991)
Jurassic Park (1993)
Fugitive, The (1993)

**Romance**
Four Weddings and a Funeral (1994)
Sense and Sensibility (1995)
Sleepless in Seattle (1993)
American President, The (1995)
Pretty Woman (1990)
Clueless (1995)
Dave (1993)
Much Ado About Nothing (1993)
Postino, Il (The Postman) (1994)
Piano, The (1993)

**Action – Sci-fi**
Independence Day (ID4) (1996)
Terminator 2: Judgment Day (1991)
Jurassic Park (1993)
True Lies (1994)
Rock, The (1996)
Stargate (1994)
Die Hard: With a Vengeance (1995)
Twelve Monkeys (1995)
GoldenEye (1995)
Heavy Metal (1981)

**Crime**
Fargo (1996)
Pulp Fiction (1994)
Usual Suspects, The (1995)
Dr. Strangelove or: How I Learned to…
Silence of the Lambs, The (1991)
Shawshank Redemption, The (1994)
Schindler's List (1993)
Blade Runner (1982)
Taxi Driver (1976)
Close Shave, A (1995)

**Thriller**
Seven (Se7en) (1995)
Clerks (1994)
Heat (1995)
True Romance (1993)
Professional, The (a.k.a. Leon: The Professional) (1994)
Kingpin (1996)
Casino (1995)
Usual Suspects, The (1995)
Dazed and Confused (1993)
Primal Fear (1996)

Clearly, aspects in the model are interpretable entities !

Note also that they are discovered by an unsupervised procedure

## Main References

- Thomas Hofmann, *Latent semantic models for collaborative filtering*, ACM Trans. Inf. Syst. 22(1): 89-115 (2004).

- Benjamin Marlin, *Modeling user rating profiles for collaborative filtering*. NIPS 16, (2004).