Recommender System Incorporating User Personality Profile through Analysis of Written Reviews

Peter Potash & Anna Rumshisky
Department of Computer Science, University of Massachusetts Lowell

Motivating Example

- Two users gave the film ‘Inception’ a 10/10 rating
- User A’s review:
  - “My sister has been bothering me to see this movie for more than two months, and I am really glad that she did, because this movie was excellent, E-X-C-E-L-L-E-N-T, EXCELLENT!”
- User B’s review:
  - “So far, Christopher Nolan has not disappointed me as a director, and `Inception' is another good one.”

Motivation

- Standard Matrix Factorization ignores external information about users and items
- By analyzing users' written reviews, we have a better understanding of users
- This understanding leads to better recommendations
  - Even if users give same numeric rating for item, an analysis of the written review helps to understand how preferences diverge in the future

Leveraging Text

- Different ways of representing users based on text
  - Bag-of-Words
  - Dimensionality Reduction (PCA, LSA, LDA)
- Personality Profile
  - Personality profile provides interpretable representation of users

Presentation Outline

- Previous Approaches
- Kernelized Probabilistic Matrix Factorization (KPMF)
- Creating a Personality Model
- Experimental Design
- Results
- Discussion

Related Work: Matrix Factorization with NLP

- Topic Modeling
  - McAuley and Leskovec. *Hidden factors and hidden topics: understanding rating dimensions with review text*. 2013.
  - Bao et al. *TopicMF: Simultaneously Exploiting Ratings and Reviews for Recommendation*. 2014.
- Distributed Representations
  - Almahairi et al. *Learning distributed representations from reviews for collaborative filtering*. 2015.
- These works primarily seek to understand items/categories

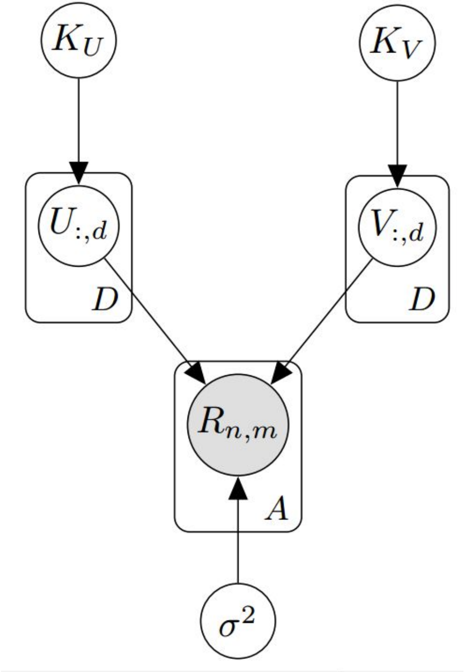
Related Work: Rec Systems with Personality

- Personality from explicit test
  - Tkalcic et al. *Personality based user similarity measure for a collaborative recommender system*. 2009.
  - Nunes. *Recommender systems based on personality traits*. 2008.
- Personality from text
  - Roshchina et al. *A comparative evaluation of personality estimation algorithms for the twin recommender system*. 2011

KPMF

- Assumes generative process for rows of latent user/item matrices
- KPMF uses Gaussian Process parameterized by zero mean function and covariance matrices, K_u and K_v
- Rating for user i and item j drawn from Normal distribution, centered at u_i and v_j with standard deviation σ, a hyperparameter of the model

KPMF



MyPersonality Dataset (2013)

- Facebook activity of various users, including posts
- Users separately took personality test to determine **Big 5 personality score** (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism)
- Sample dataset available for 250 users, roughly 10,000 status updates
- Shared task was held to develop systems that could predict personality based on facebook activity
- Systems developed for shared task inspired features for our system

Personality Features

- Punctuation count
- Part-of-speech tag count
- Affin count
    - Lexicon assigns emotional valence score from -5 to 5 for a given token
- “To” count
- General Inquirer tags
    - Lexicon has 182 possible topic tags for input tokens

Personality Model

- Train 5 different personality models, one for each personality trait
- Use linear regression to predict continuous values in range [1,5]
- Result is 5-dimensional vector

Creating User Covariance Matrix

- Given personality vectors for two users, covariance is calculated as cosine similarity between personality vectors
- Resulting cosine similarities are projected to interval [0,0.4]

Experimental Design

- Used crawler to create dataset of 2,087 Users and 3,500 movies from imdb
- Rating matrix is 1.55% dense
- Randomly create 3/5,1/5,1/5 training, validation, test split of data
- Trained model with gradient descent, use performance on validation set as stopping criteria
- Ran experiment 5 times, with 5 random splits of dataset

Comparative Models

- Matrix Factorization without optimization
- KPMF with diagonal user covariance matrix
- KPMF using textual feature representation
- KPMF using personality vector
- KPMF using personality vector concatenated with text features
- KPMF using dimensionality reduction of feature vector

Results

Model	RMSE	% Improvement
KPMF Personality	0.2006	11.3
KPMF Text Features	0.1980	12.5
KPMF Personality + Text Features	**0.1901**	**16.0**
KPMF Dimensionality Reduction	0.2087	7.7
KPMF Diagonal User Covariance	0.2122	6.2
MF	0.2262	-

Discussion

- Best performance achieved when using both features and personality vector
- Raw features alone better than personality vector
- Personality vector is better dimensionality reduction
- Future work:
  - Improved personality prediction model
  - Compute weighted similarity of personality vectors, learning weights during training
  - Filtering item-specific content from reviews

Questions?