Matrix Factorization Recommender Systems and Cold Start

Overview

Recommender Systems
Created to direct users to items of interest, given a dataset, e.g., user ratings or product descriptions.

Cold Start Problem
How to give ratings for new users and new items for which no ratings are recorded in the dataset?

Purpose
- Survey state of the art Matrix Factorization solutions to the Cold Start problem.
- Outline our on going current research.

User-Item Utility Matrix
Ratings for m users and n items in an m x n matrix, R. 
R_{i,j} = user i's rating of product j.

Matrix Factorization Model [1]
Assume low rank to complete R. R may be factored into two matrices, U and P, so that U P T \approx R where

U_{i,k} = user i's association with latent factor k
P_{j,k} = item j's association with latent factor k
P_{j,k} U_{i,k} \approx \text{User-Item interaction.}

Objective Function
Let \omega = \{\omega_j : user i has rated item j\}
\[ L(P, U) = \sum_{\omega_j \in \omega} (R_{i,j} - p_{i,j} u_{j,k})^2 + \lambda (\|p_i\|^2 + \|u_j\|^2) \]

Minimize L using Gradient Descent

User Cold Start Problem

How to predict ratings for new users in the MF framework?

Functional Matrix Factorization [2]
- Hybrid MF-Decision Tree Recommender System
- Uses MF to find best queries to estimate user profiles
- Recommender Decision Trees
  - Query new users about preferences for important items
  - a_u = user i's set of responses to queries.
  - T(a_u) \approx u_i (function for approx. user profiles)

Objective Function
\[ L(P, T) = \sum_{\omega_j \in \omega} (R_{i,j} - p_{i,j} T(a_u))^2 + \lambda (\|p_i\|^2 + \|u_j\|^2) \]

Fix T and Optimize w.r.t. P
\[ \forall j, p_j = \left[ \sum_{i \in \omega_j} T(a_u) T(a_u) \right]^{-1} \sum_{i \in \omega_j} R_{i,j} T(a_u) \]

Fix P and Optimize w.r.t. T
- Starting at root, partition users into sets by query \( h \).
- Find Optimal Profiles for users in Child Nodes
  - Let \gamma be a variable ranging over subscripts L, D, Q
  - u_\gamma = \arg\min \sum_{i \in \omega_j} \sum_{(h,j) \in \omega_j} (R_{i,j} - p_{i,j} u_{\gamma,j})

User Cold Start Summary
For new user z, use decision tree and queries to derive user profile p_{z,j}. User z's predicted rating for item j is then \( p_{z,j} u_{z,j} \).

Item Cold Start Problem

How to predict new unrated items in MF framework?

Item Cold Start Summary

For new item z and document-term vector q_{z,p}, derive item profile p_{z,j} by solving \( W_{z,p} = q_z \). User z's estimated rating for item j is then \( p_{z,j} u_{z,j} \).

Datasets

Amazon Movies Dataset (Leskovec, McAuley. 2013)
- 7,500,072 reviews (text & ratings)
- Plain text descriptions for all reviewed items

MovieLens Datasets (http://movielens.org)
- 20 million movie reviews
- Reviewed movies can be paired with plot descriptions from the Internet Movie Database (http://www.imdb.com/)
- Smaller Movielens dataset (100,000 reviews) has simple user demographic info (age, gender, occupation, zip)

Current Research

- Benchmark state of the art methods
- Implement and refine new models for item cold start.
- Explore models which handle both user and item cold start.
- Explore metrics for Coldish Start (when users or items have some but fewer ratings)

Model

Base Model for Item Cold Start

Objective Function
Let \( \Omega_{i,j} = 1 \) if \( (i, j) \in \omega \), 0 otherwise. Then
\[ L(U, W, Q) = \|R - U P\|^2 + \|T - U P\|^2 \]
\[ + \|R - U W T Q\|^2 + \|T - U W T Q\|^2 \]

Model Summary
The model learns user i's user-term vector \( u_i \). User i's estimated rating for new or established item \( z \) is then \( p_{z,j} u_{z,j} \).

Train and Test Split

Test set for users and items in "known" ratings set

Cold Start Metric (Basic MF version) [4]