CSE 8803RS: Recommendation Systems

Summary of Topics

Hongyuan Zha

School of Computational Science & Engineering
College of Computing
Georgia Institute of Technology

Surveys of Recommenders

We focused on two survey papers on recommender systems
— it seems that a more up to date survey paper will be useful

- Examples of Recommenders
- Categorizations of Recommenders
  — Content-based
  — Collaborative filtering: memory-based; model-based
- Challenging issues for Recommenders

Classical Recommenders

We study several classical papers on early recommender systems

- Tapestry
  — emails and news
- GroupLens
  — NetNews
- Virtual Community
  — Movies
- Ringo
  — Music and artists

Memory-Based CF Methods

Two classical approaches for CF

- User-based methods vs. Item-based methods
- Duality of users and items; asymmetry in real-world applications
- Similarity measures and extensions
- Fast methods for computing item similarity tables
- Experiments

Recommendation Justification

This is a very practical issue in CF, and its effective resolution will impact user experience/retaintion

- The idea of user feature profiles and coverage
- Biclustering based users' ratings
- Approaches based on case-based reasoning

Matrix Factorization Methods

We focused on the workhorse of CF using matrix factorization methods — all those methods are related to SVD and low-rank factorization

- SVD and regularized SVD
  — simple idea of shrinking to zero
- Several other tricks for real-world applications
  — bias removal
  — predicted score trimming
  — representing user profiles using rated item profiles
- Incorporating memory-based methods

Matrix Factorization Methods

Several more disciplined, model-based formulations of matrix factorization

- Maximum-margin matrix factorization
  — a classical paper, leading to SDP
- Probabilistic and Bayesian formulation
  — User/item profiles modeled by Gassians, add priors to those Gaussians
  — Gibbs sampling for posteriors and predictive distributions

Matrix Factorization Methods

Theoretical analysis: random missing model, low-rank assumption, exact recovery of missing entries

- Matrix completion and compressive sensing
  — several other types of matrix completion problems
- Theoretical bounds
  — random missing mechanism
  — several different styles of analysis
  — upper bounds on fraction of missing entries

Multiple Data Sources

Assuming availability of user/item meta-features

- Incorporate user and/or item features
  — regression prior based methods
  — a unified model
- Probabilistic models
  — two-stage hierarchical models
  — Monte Carlo EM

Multiple Data Sources

Data in the form of multiple entities in multiple relations

- Joint matrix factorization
  — More general framework for simultaneous low-rank factorization of multiple matrices
  — Not as many real use cases
  — Still some technical issues need to addressed: entity profiles involved in multiple relations
- Application in document recommendation
  — documents, authors, publication venues
  — joint factorization

Temporal Recommendation

Capture temporal variations of user/item characteristics

- Temporal dynamics
  — Modifications of MF-based methods
  — Improvements?
- Transient vs. recurrent interests

Temporal Recommendation

Next-basket problems

- Each user is identified with a transition matrix on items, rather than a simple rating vector on items
  — Tensor factorization
  — Simplifications done on pair-wise factors

Cold-Start Problems

Assuming no meta-features: designing interview process by asking a sequence of questions

- Simple questions
- Information criteria
- Static seed approaches
- Interview process based on decision trees

Online Recommendation

Multi-armed bandit problem

- Upper confidence bound method (UCB)
- Exploration and exploitation
  — items constantly change
  — need to present new items to user to get feedback
  — long term reward

Long-Tail Problems in CF

Long-tail phenomenon

- Long Tail phenomenon in Netflix data
  — head and tail measured in relative sense
- How to leverage long tail for CF?
  — user clustering: related to group recommendation

Transfer Learning for Recommendation

Cross-domain recommendation: leverage data for mature domains

- Matrix factorization based methods
  — what factors are shared?
- Multiple domain problems
  — Correlated user profiles for different domains

Content based Recommendation and User Modeling for Community Conversations

- Search and recommendation agent in a 'conversational' community
- User models (short term and long term)
- NLP and IE for query reformulations and conversational/semantic search

Recommendation and Social Networks

Recommendations exploring social networks. Can be considered as a special case of joint matrix factorization

- Trust and distrust
- Interests and friendship propagation
- Positive and negative influences

Evaluation for Recommenders

Discuss several conventional evaluation metrics for CF

- Evaluation metrics
  — RMSE
  — MAE
  — DCG-like metrics for ranking
- Diversity

Computational Issues

- Reduce memory footprint using hashing
  — matrix factorization on reduced space
- Scalability
- Parallel implementations

Non-standard Items and Users

- Group recommendation
  — convert to ranking problem
  — rank aggregation methods
- Subset of items

Trust and Distrust

- Distrust propagation

Privacy

- Netflix data privacy