Recommender Systems: User Experience and System Issues

Joseph A. Konstan
University of Minnesota

konstan@cs.umn.edu
http://www.grouplens.org

About me …

• Professor of Computer Science & Engineering, Univ. of Minnesota
• Ph.D. (1993) from U.C. Berkeley
  • GUI toolkit architecture
• Teaching Interests: HCI, GUI Tools
• Research Interests: General HCI, and ...
  • Collaborative Information Filtering
  • Multimedia Authoring and Systems
  • Visualization and Information Management
  • Medical/Health Applications and their Delivery

A Quick Introduction

• What are recommender systems?
• Tools to help identify worthwhile stuff
  • Filtering interfaces
    • E-mail filters, clipping services
  • Recommendation interfaces
    • Suggestion lists, “top-n,” offers and promotions
  • Prediction interfaces
    • Evaluate candidates, predicted ratings

Scope of Recommenders

• Purely Editorial Recommenders
• Content Filtering Recommenders
• Collaborative Filtering Recommenders
• Hybrid Recommenders

Wide Range of Algorithms

• Simple Keyword Vector Matches
• Pure Nearest-Neighbor Collaborative Filtering
• Machine Learning on Content or Ratings

Classic Collaborative Filtering

• MovieLens*
• K-nearest neighbor algorithm
• Model-free, memory-based implementation
• Intuitive application, supports typical interfaces
  • *Note – newest releases use updated architecture/algorithm
Select Items; Predict Ratings

Understanding the Computation

<table>
<thead>
<tr>
<th></th>
<th>Hoop Dreams</th>
<th>Star Wars</th>
<th>Pretty Woman</th>
<th>Titanic</th>
<th>Blimp</th>
<th>Rocky XV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td>D A B D ?</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>A F D F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Susan</td>
<td>A A A A A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pat</td>
<td>D A C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jean</td>
<td>A C A C A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben</td>
<td>F A F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nathan</td>
<td>D A A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Understanding the Computation

<table>
<thead>
<tr>
<th></th>
<th>Hoop Dreams</th>
<th>Star Wars</th>
<th>Pretty Woman</th>
<th>Titanic Blimp</th>
<th>Rocky XV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td>D</td>
<td>A</td>
<td>B</td>
<td>D</td>
<td>?</td>
</tr>
<tr>
<td>John</td>
<td>A</td>
<td>F</td>
<td>D</td>
<td>F</td>
<td>?</td>
</tr>
<tr>
<td>Susan</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Pat</td>
<td>D</td>
<td>A</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jean</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>Ben</td>
<td>F</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nathan</td>
<td>D</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MovieLens

Freely accessible at: http://www.movielens.org
**Talk Roadmap**

- Introduction
- Algorithms
- Research Overview
- Influencing Users
- Recommending Research Papers
- Rethinking Recommendation

**Collaborative Filtering Algorithms**

- Non-Personalized Summary Statistics
- K-Nearest Neighbor
  - user-user
  - item-item
- Dimensionality Reduction
  - LSI
  - PLSI
  - Factor Analysis
- Content + Collaborative Filtering
- Burke's Survey of Hybrids
- Graph Techniques
- Horting
- Clustering
- Classifier Learning
  - Naive Bayes
  - Bayesian Belief Networks
  - Rule-induction
Collaborative Filtering Algorithms

- Non-Personalized Summary Statistics
- K-Nearest Neighbor
  - user-user
  - item-item
- Dimensionality Reduction
  - LSI
  - PLSI
  - Factor Analysis
- Content + Collaborative Filtering
  - Burke’s Survey of Hybrids
- Graph Techniques
- Clustering
- Classifier Learning
  - Naïve Bayes
  - Bayesian Belief Networks
  - Rule-induction

Item-Item Collaborative Filtering

Used for similarity computation
**Item-Item Matrix Formulation**

- Target item
- 5 closest neighbors
- Raw scores for prediction generation
- Approximation based on linear regression

**Item-Item Discussion**
- Good quality, in sparse situations
- Promising for incremental model building
  - Small quality degradation
  - Big performance gain

**Collaborative Filtering Algorithms**
- Non-Personalized Summary Statistics
- K-Nearest Neighbor
  - user-user
  - item-item
- Dimensionality Reduction
  - LSI
  - PLSI
  - Factor Analysis
- Content + Collaborative Filtering
  - Burke's Survey of Hybrids
  - Graph Techniques
  - Clustering
  - Classifier Learning
  - Naïve Bayes
  - Bayesian Belief Networks
  - Rule-induction

**Dimensionality Reduction**
- Latent Semantic Indexing
  - Used by the IR community
  - Worked well with the vector space model
  - Used Singular Value Decomposition (SVD)
- Main Idea
  - Term-document matching in feature space
  - Captures latent association
  - Reduced space is less-noisy

**SVD: Mathematical Background**

The reconstructed matrix $R_k = U_k S_k V_k^T$ is the closest rank-$k$ matrix to the original matrix $R$.

**SVD for Collaborative Filtering**

1. Low dimensional representation $O(m+n)$ storage requirement
2. Direct Prediction
**Singular Value Decomposition**
Reduce dimensionality of problem
- Results in small, fast model
- Richer Neighbor Network
Incremental Update
- Folding in
- Model Update

**Collaborative Filtering Algorithms**
- Non-Personalized Summary Statistics
- K-Nearest Neighbor
  - user-user
  - item-item
- Dimensionality Reduction
  - LSI
  - PLSI
  - Factor Analysis
- Content + Collaborative Filtering
  - Burke's Survey of Hybrids
  - Graph Techniques
  - Horting
  - Clustering
  - Classifier Learning
    - Naïve Bayes
    - Bayesian Belief Networks
    - Rule-induction

**Talk Roadmap**
- Introduction
- Algorithms
- Research Overview
- Influencing Users
- Recommending Research Papers
- Rethinking Recommendation

**Current and Recent Research**
User Experience
- Impact of Ratings on Users
- New User "Orientation"
- Confidence Displays
- Interface Design
- Human-Recommender Interaction
Algorithmic and Systems Issues
- Beyond Accuracy: Metrics and Algorithms
- Buddies and Multi-User Recommendations
- Influence and Shilling
Eliciting Participation in On-Line Communities
- Reinventing Conversation
- User-Maintained Communities
Extending Recommendation to New Domains
- Recommending Research Papers

**Talk Roadmap**
- Introduction
- Algorithms
- Research Overview
- Influencing Users
- Recommending Research Papers
- Rethinking Recommendation

**Does Seeing Predictions Affect User Ratings?**
- RERATE: Ask 212 users to rate 40 movies
  - 10 with no shown prediction
  - 30 with shown predictions (random order): 10 accurate, 10 up a star, 10 down a star
- Compare ratings to accurate predictions
  - "Prediction" is user's original rating
  - Hypothesis: users rate in the direction of the shown prediction
**The Study**

Please rate the movies listed below. These ratings will not be saved to your profile.

<table>
<thead>
<tr>
<th>Predicted Rating</th>
<th>Your Rating</th>
<th>Genre</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>Drama</td>
<td>The Godfather</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Comedy</td>
<td>The Shawshank Redemption</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Action</td>
<td>The Dark Knight</td>
</tr>
</tbody>
</table>

**Seeing Matters**

- **Prediction shown?**
  - Not shown
  - Shown

**Accuracy Matters**

- **Prediction manipulation**
  - Down: 20% Ratings %
  - Accurate: 60% Ratings %
  - Up: 20% Ratings %

**Domino Effects?**

- The power to manipulate?

**Rated, Unrated, Doesn’t Matter**

- Recap of RERATE effects:
  - Showing prediction changed 8% of ratings
  - Altering shown prediction changed 12%
- Similar experiment, UNRATED movies
  - 137 experimental users, 1599 ratings
  - Showing prediction changed 8% of ratings
  - Altering shown prediction changed 14%

**But Users Notice!**

- Users are often insensitive...
- UNRATED part 2: satisfaction survey
  - Control group: only accurate predictions
  - Experimental predictions accurate, useful?
  - ML predictions overall accurate, useful?
  - Manipulated preds less well liked
  - Surprise: 24 bad = MovieLens worse!
**Talk Roadmap**

- Introduction
- Algorithms
- Research Overview
- Influencing Users
  - Recommending Research Papers
  - Rethinking Recommendation

**Recommending Research Papers**

- Using Citation Webs
- For a full paper, we can recommend citations
  - A paper “rates” the papers it cites
  - Every paper has ratings in the system
- Other citation web mappings are possible, but many are have problems

---

**Pure Experiment Results -- Online**

- Worst algorithm returned good results over 25% of the time
- 76% of users got at least one good recommendation
- Users happy with one good recommendation in list of five

**What’s Next?**

- Short-Term Efforts
  - Task-specific recommendation
  - Understanding personal bibliographies
  - Privacy issues
- Longer-Term Efforts
  - Toolkits to support librarians and other power users
  - Exploring the shape of disciplines
  - Rights issues

**Task-Specific Recommendations**

- Many different user needs
  - Awareness in area of expertise
  - Find specific work in area of expertise
  - Explore peripheral or new area
  - Find people with relevant expertise
    - Reviewers, program committees, collaborators
  - Reading list for students, newcomers
    - Individuals or groups
- Different algorithms fulfill different needs

**Talk Roadmap**

- Introduction
- Algorithms
- Research Overview
- Influencing Users
- Recommending Research Papers
- Rethinking Recommendation
Evaluating Recommendations

- Prediction Accuracy
  - MAE, MSE
- Decision-Support Accuracy
  - Reversals, ROC
- Recommendation Quality
  - Top-n measures
- Item-Set Coverage

From Items to Lists

- Do users really experience recommendations in isolation?

Making Good Lists

- Individually good recommendations do not equal a good recommendation list
- Other factors are important
  - Diversity
  - Affirmation
  - Appropriateness
- Called the “Portfolio Effect”
  [Ali and van Stam, 2004]

Topic Diversification

- Re-order results in a rec list
- Add item with least similarity to all items already on list
- Weight with a ‘diversification factor’
- Ran experiments to test effects
**Experimental Design**

- Books from BookCrossing.com
- Algorithms
  - Item-based CF
  - User-based CF
- Experiments
  - On-line user surveys
  - 2125 users each saw one list of 10 recommendations

**Online Results**

![Graph showing comparison between Item-based CF and User-based CF](graph.png)

**Diversity is Important**

- User satisfaction more complicated than only accuracy
- List makeup is important to users
- 30% change enough to alter user opinion
- Change not equal across algorithms

**Human-Recommender Interaction**

- Three premises:
  - Users perceive recommendation quality in context; users evaluate lists
  - Users develop opinions of recommenders based on interactions over time
  - Users have an information need and come to a recommender as a part of their information seeking behavior


**HRI Pillars and Aspects**

**HRI Process Model**

- Makes HRI Constructive
  - Links Users/Tasks to Algorithms
  - Need New Metrics
**New Metrics**

- Benchmark a variety of algorithms
- Need several metrics inspired by different HRI Aspects
- Examples:
  - Ratability
  - Boldness
  - Adaptability

**Metric Experimental Design**

- ACM DL Dataset
  - Thanks to ACM for cooperation!
  - 24,000 papers
  - Have citations, titles, authors, & abstracts
  - High quality
- Algorithms
  - User-based CF
  - Item-based CF
  - Naïve Bayes Classifier
  - TF/IDF Content-based
  - Co-citation
  - Local Graph Search
  - Hybrid variants

**Ratability**

- Probability a user will rate a given item
  - “Obviousness”
  - Based on current user model
  - Independent of liking the item
- Many possible implementations
  - Naïve Bayes Classifier

**Ratability Results**

- Mean Ratability
  - Top-10
  - Top-20
  - Top-30
  - Top-40

**Boldness**

- Measure of “Extreme Predictions”
  - Only defined on explicit rating scale
  - Choose “extreme values”
  - Count appearance of “extremes” and normalize
- For example, MovieLens
  - 0.5 to 5.0 star scale, half-star increments
  - Choose 0.5 and 5.0 as “extreme”

**Boldness Results**

- Ratio to Expected
  - Top-10
  - Top-20
  - Top-30
  - Top-40
**Adaptability**

- Measure of how algorithm changes in response to changes in user model
  - How do users grow in the system?
  - Perturb a user model with a model from another random user
    - 50% each
    - See quality of new recommendation lists

**Adaptability Results**

![Adaptability Results](chart.png)

**Conclusions**

- From humble origins ...
  - Substantial algorithmic research
  - HCI and online community research
  - Important applications
  - Commercial deployment

**Acknowledgements**

- This work is being supported by grants from the National Science Foundation, and by grants from Net Perceptions, Inc.
  - Many people have contributed ideas, time, and energy to this project.
Recommender Systems: User Experience and System Issues

Joseph A. Konstan
University of Minnesota

konstan@cs.umn.edu
http://www.grouplens.org