The GroupLens Research Project: Collaborative Filtering Recommender Systems

Joseph A. Konstan
University of Minnesota
konstan@cs.umn.edu
http://www.grouplens.org

About me …

Associate 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
  • Web Automation
  • Visualization and Information Management

The Problem: Information Overload

Too many
  • research papers
  • books
  • movies
  • web pages
  • … even Usenet News articles!

Recommenders

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

History of Recommender Systems

The Early Years …

Why cave dwellers survived
How editors are like cave dwellers
Critics, critics, everywhere
Information Filtering

Information retrieval
- Dynamic information need
- Static content base

Information filtering
- Static information need
- Dynamic content base

Collaborative Filtering

Premise
- Information needs more complex than keywords or topics: quality and taste

Small Community: Manual
- Tapestry – database of content & comments
- Active CF - easy mechanisms for forwarding content to relevant readers

Automated CF

The GroupLens Project
(Resnick et al. CSCW '94)
- ACF for Usenet News
  - users rate items
  - users are correlated with other users
  - personal predictions for unrated items
- Nearest-Neighbor Approach
  - find people with history of agreement
  - assume stable tastes

Usenet Trial
(Miller et al. Usenet '97; Konstan et al. CACM Mar. '97)

Medium-scale Usenet trial
- seven weeks
- 250 users; 47,569 ratings; over 600,000 predictions
- variety of newsgroups
  - moderated and unmoderated
  - technical and recreational
  - gathered reading activity as well as ratings

Does it Work?

Yes: The numbers don’t lie!
- Usenet trial: rating/prediction correlation
  - rec.humor: 0.62 (personalized) vs. 0.49 (avg.)
  - comp.os.linux.system: 0.55 (pers.) vs. 0.41 (avg.)
  - rec.food.recipes: 0.33 (pers.) vs. 0.05 (avg.)
- Significantly more accurate than predicting average or modal rating.
- Higher accuracy when partitioned by newsgroup
**It Works Meaningfully Well!**

Relationship with User Behavior
- Twice as likely to read 4/5 than 1/2/3

Users Like GroupLens
- Some users stayed 12 months after the trial!

---

**ACF Blossomed**

1995
- Ringo (later Firefly)
- Bellcore Video Recommender

1996 Recommender Systems Workshop
- Agents Inc. (later Firefly)
- Net Perceptions

New issues of scale and performance!

---

**Today**

Broad research community
- Live research systems
- Substantial integration with machine learning, information filtering

Increasing commercial application
- Available commercial tools

---

Copyright 2001 Joseph A. Konstan
The GroupLens Research Project

Summer 2001

How It Works

C.F. Engine

Ratings

Correlations

Copyright 2001 Joseph A. Konstan
**GroupLens Model of Information Filtering**

- Users rate items.
- Users are correlated with other users.
- Predictions made for an item’s value to a particular user by combining ratings of highly correlated users who rated it.
- Recommendations for items for a particular user by identifying popular items among correlated users.

**Understanding the Computation**

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

**Understanding the Computation**

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

### Recent and Current Research

- Accuracy, Scale, and Sparsity
- Algorithm Performance and Metrics
- Filterbots
- Dimensionality Reduction Algorithms
Interfaces and User Experience
- Explaining Recommendations
- Ephemeral Recommendations
- PolyLens: Multi-User Recommendations
- MetaLens: Multi-Source Recommendations

Other Research (not covered in this talk)
- Distributed Recommenders (Sarwar)
- E-Commerce Recommender Applications (Schafer)
- User and Usage Studies

Algorithm Performance and Metrics
(Herlocker et al., SIGIR ’99; …)
Breese studied recommender algorithms
- k-nearest neighbors as good as any
We looked at relevant tuning parameters
- limiting neighborhood size important
- normalization of ratings very important
- most other parameters unimportant
  - correlation measure, weightings

Accuracy, Scale, and Sparsity

Which Metrics?
Many metrics used in published work
- Error metrics (MAE, MSE, RMSE)
- Decision-support metrics (ROC, errors)
- IR metrics (version of precision, recall)
We found that there are only two types
- Rank-sensitive, value-sensitive
  - All seem to work equally well and nearly identically, within type

Filterbots
The Inspiration: Need selfless, consistent raters
Humans?
No: ratings robots.
Filterbot Studies

Usenet Filterbots (Sarwar et al. CSCW 98)
- Simple, non-personalized filterbots
  - Spelling, Length, % new text
- One filterbot at a time

MovieLens Filterbots (Good et al. AAAI ’99)
- Personalized filterbots
  - Learned from genre, cast, descriptions
- Many filterbots per person

Lessons Learned

Even simple filterbots added value
C.F. best way to create a personal combination of filterbots
Filterbots better than a small community of users
Filterbots + users in CF better than either alone

Advantages of the FilterBot model

Combines best of agents and humans
- agents rate frequently, quickly, consistently
- humans add subjective taste and quality

Framework pulls out the best of each
- use only the bots that work; ignore the others
- use only the people who agree; ignore the others
- balance people and bots based on available ratings and agreement

Risks of Filterbots

- What if no humans read certain articles?
  - “voluntary” censorship or quality control?
- What about rogue filterbots?
- What if people “prefer” filterbots to humans?

New Algorithms

(Sarwar et al., EC 00 & WebKDD 01)

Significant challenges
Scale
- Number of users
- Number of items
Sparsity
- Small percentage of items experienced
- Hard to find overlap with other users
Example Challenge

Synonymy
- Similar products treated differently
- Increases sparsity, loss of transitivity
- Results in poor quality

Example
- C₁ rates recycled letter pads High
- C₂ rates recycled memo pads High
  → Both of them like Recycled office products

Idea: Dimensionality Reduction

Latent Semantic Indexing
- Used by the IR community for document similarity
- Works well with similar vector space model
- Uses Singular Value Decomposition (SVD)

Main Idea
- Find (latent) "taste space"
- Represent users and items as points (vectors) in taste space
- Reduced space is dense and less-noisy

SVD: Mathematical Background

\[ R_{k} = U_{k} S_{k} V'_{k} \]

The reconstructed matrix \( R_{k} \) 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

Experimental Setup

Data Sets
- MovieLens data (www.movielen.umn.edu)
  - 943 users, 1,682 items
  - 100,000 ratings on 1-5 Likert scale
  - Used for prediction and neighborhood experiments
- E-commerce data
  - 6,502 users, 23,354 items
  - 97,045 purchases
  - Used for neighborhood experiment

Results: Prediction

Graph showing MAE vs. training/test ratio for different models including SVD.
SVD Conclusions

Successful and promising approach

Several obstacles to overcome
- Incremental update
- Efficient “top-n” recommendations

Exploring SVD-based and other new algorithmic approaches

Interfaces and User Experience

Explaining Recommendations (Herlocker et al. CSCW 2000)

Challenge: Belief
- Why should users believe the recommendations?
- When should users believe the recommendations?

Approach
- Explain recommendations
  - Reveal data, process
  - Corroborating data, track record

Two Studies

Pilot study of explanation feature
- Users liked explain
- Unclear whether they become more effective decision-makers

Comprehensive study of different explanation approaches
- Wide variation of effectiveness
- Some explanations hurt decision-making

Most Compelling Interfaces

- Simple visual representations of neighbors ratings
- Statement of strong previous performance
  “MovieLens has predicted correctly 80% of the time for you”

Less Compelling Interfaces

- Anything with even minimal complexity
  - More than two dimensions
- Any use of statistical terminology
  - Correlation, variance, etc.
Addressing Ephemeral Needs (Herlocker)

What is an ephemeral interest need?

- Immediate, temporary, dynamic
- Current systems don’t support this
  - Assume interests will remain relatively constant
  - Recommendations are relative to all your interests as a whole

One Simple Approach

User submits “theme” query

- Theme contains examples of items similar to those desired by the user

Set of potentially similar items identified

- Using item-to-item correlation in ratings space
- Potentially similar items ranked based on traditional ACF predictions

Theme Creation

Theme Selection

Query Results

Results of Theme Study

Users were very positive about the theme query interface

- Relevance of results were dependent on the “support threshold”
  - Low support threshold => fewer relevant results

- When results were relevant, users were positive overall

Even the users in the low support threshold groups indicated they would like to have the interface added to MovieLens
PolyLens: A Group Recommender  
(O’Connor et al. Interact 2001)  
Challenge: People watch movies together  
Solution: A recommender for groups  
Issues  
- Group formation, rules, composition  
- Recommender algorithm for groups  
- User interface  

Goals  
Explore group recommender design space  
See if users would want and use a group recommender, at least for movies  
Study behavior changes in group members  
- group vs. other users  
- new users via groups vs. other new users  
Learn lessons about group recommenders  

Design Issues  
Characteristics of groups  
- public or private  
- many or few  
- permanent or ephemeral  
Formation and evolution of groups  
- joining policy  
- administration and rights  

PolyLens  
Design choices  
- private, small, administered, invited groups  
- combine individual recs with minimum misery  
- high-information interface with opt-out  
External invitations added by popular demand  

Field Testing  
PolyLens  

PolyLens Field Trial Timeline  
- Users  
- Groups
**Survey and Usage Results**

Satisfaction (95% like, 77% more useful)
Privacy not an issue (94% see, 93% share)
- individual recommendations “essential”
Groups reflect “real life” groups
New users via groups stayed 1.5x as often
- group vs. other users a wash
Many stillborn groups

**Field Test Results and Lessons**

Users like and use group recommenders
- groups have value for all members
- groups can help with outreach to new members
Users trade privacy for utility
Groups are both permanent and ephemeral
Users must be able to find each other

**MetaLens: A Meta-Recommender**

Integrating multiple sources of information into a single recommendation list

**What is the problem?**

- Genre
- MPAA ratings
- Film length
- Objectionable Content
- Distributor
- Release Date
- Start/End Time

**Sources of Data**

- Critical Reviews
- Average User Rating
- User’s personalized MovieLens prediction
- Distance to the Theater
- Special Accommodations
- Discounted Shows
What Have We Learned?

- Meta-recommenders can be built.
- Anecdotally, users like them.
- Some users make heavy use of them, and heavy users are most likely to make some use of them.

Conclusions

Collaborative filtering works!

Lots of important issues:
- Algorithms
- Interfaces and User Experience
- Privacy
- Applications

Future Work

Better integration of collaborative and content filtering

Better support for community

Better understanding of user rewards, social role of recommenders
CF Under Diminishing Returns

Original goal of CF was to help people sift through the junk to find the good stuff.

Today, there may be so much good stuff that you need to sift even more.

Certain types of content yield diminishing returns, even with high quality.

Portfolios of Content

What if my recommender knows which articles I’ve read, and can identify articles by topic?

What if it sees that I experience marginal returns from reading similar articles on a topic?

Could we downgrade some articles based on “lack of new content?” Could we discover which articles using collaborative filtering?

Temporal Collaborative Filtering

Today’s CF systems may expire or degrade ratings, but do little to detect or predict changes in preference.

Ripe area with lots of commercial applications …

Wine for the Time

Seasonal taste – can we detect that a particular customer shifts wine tastes during hot and cold weather? Can we learn either the content, or separate profiles, reflecting these different tastes?

Evolving taste – can we help a wine newcomer build her palate? Could we identify wines that take her a step or two beyond her current ones? Can we do so by augmenting regular collaborative filtering with temporal models?

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.

The GroupLens Research Project: Collaborative Filtering Recommender Systems

Joseph A. Konstan
University of Minnesota
konstan@cs.umn.edu
http://www.grouplens.org

Copyright 2001 Joseph A. Konstan