User Modeling and Recommendations

– Part 2

Many slides adapted from Lora Aroyo
http://de.slideshare.net/laroyo
User Modeling Basic Concepts

- **User Profile**: a data structure that represents a characterization of a user at a particular moment of time represents what, from a given (system) perspective, there is to know about a user. The data in the profile can be explicitly given by the user or derived by the system.

- **User Model**: contains the definitions & rules for the interpretation of observations about the user and about the translation of that interpretation into the characteristics in a user profile.
  - *user model* is the recipe for obtaining and interpreting user profiles.

- **User Modeling**: the process of representing the user.
USER ADAPTATION

Knowing the user - this knowledge - can be applied to adapt a system or interface to the user to improve the system functionality and user experience.
User-Adaptive Systems

user profile

user modeling

profile analysis

observations, data and information about user

adaptation decisions

Last.fm: Adapts to your music taste

user profile
interests in
genres, artists, tags

user modeling
(infer current
musical taste)

compare profile
with possible next
songs to play

next song to be played

history of
songs, like,
ban, pause,
skip
Eli Pariser: Beware online “filter bubbles”

As web companies strive to tailor their services (including ads and search results) to our personal habits, there’s a dangerous unintended consequence: we’re spending more and more time online in a world that is tailored to fit our own personal interests. The more we spend time in these filter bubbles, the more insulated we become from reality. Pariser argues powerfully that this will ultimately prove to be bad for us and bad for democracy.

Pariser is the co-founder of the social networking site Friendster and the author of "The Filter Bubble: How the New Personalized Web Is Changing What We Buy, Who We Vote For, and Our Ideas of News and Culture."
Issues in User-Adaptive Systems

- **Overfitting, “bubble effects”, loss of serendipity problem:**
  - systems may adapt too strongly to the interests/behavior
  - e.g., an adaptive radio station may always play the same or very similar songs
  - We search for the right balance between *novelty* and *relevance* for the user (Diversity!)

- **“Lost in Hyperspace” problem:**
  - when adapting the navigation – i.e. the links on which users can click to find/access information
  - e.g., re-ordering/hiding of menu items may lead to confusion
What is good user modeling & personalization?

http://www.flickr.com/photos/bellarosebyliz/4729613108
Success perspectives

- From the consumer perspective of an adaptive system:
  
  Adaptive system maximizes satisfaction of the user
  
  hard to measure/obtain

- From the provider perspective of an adaptive system:

  Adaptive system maximizes the profit

  influence of UM & personalization may be hard to measure/obtain
Evaluation Strategies

- **User studies**: Clean-room study: ask/observe (selected) people whether you did a good job
- **Log analysis**: Analyze (click) data and infer whether you did a good job, e.g., cross-validation by “Leave-one-out”
- **Evaluation of user modeling**:
  - measure quality of profiles directly, e.g. measure overlap with existing (true) profiles, or let people judge the quality of the generated user profiles
  - measure quality of application that exploits the user profile, e.g., apply user modeling strategies in a recommender system
    (not trivial to evaluate recommenders
    -> next lecture topic, work by Dellschaft)
Measuring success?

Is strategy X better than the baseline?
Possible metrics

- The usual IR metrics:
  - Precision: fraction of retrieved items that are relevant
  - Recall: fraction of relevant items that have been retrieved
  - F-Measure: (harmonic) mean of precision and recall

- Metrics for evaluating recommendation (rankings):
  - Mean Reciprocal Rank (MRR) of first relevant item
  - Success@k: probability that a relevant item occurs within the top k
  - Precision@k, Recall@k & F-Measure@k
  - If a true ranking is given: rank correlations

- Metrics for evaluating prediction of user preferences:
  - MAE = Mean Absolute Error
  - True/False Positives/Negatives
[Rae et al.] shows a typical example of how to investigate and evaluate a proposal for improving (tag) recommendations (using social networks)

- Task: test how well the different strategies (here different tag contexts) can be used for tag prediction/recommendation
  - Given two tags used already for a photo predict five more tags

Steps: ...

[Rae et al. Improving Tag Recommendations Using Social Networks, RIAO’10]
[Rae et al.] shows a typical example of how to investigate and evaluate a proposal for improving (tag) recommendations (using social networks)

- Task: test how well the different strategies (here different tag contexts) can be used for tag prediction/recommendation

- PC: Personal context
- SCC: social contact context
- SGC: social group context
- CC: collective/global context
Example Evaluation

- Task: test how well the different strategies (here different tag contexts) can be used for tag prediction/recommendation

Steps:
1. Gather a dataset of tag data part of which can be used as input and aim to test the recommendation on the remaining tag data
2. Use the input data and calculate for the different strategies the predictions
3. Measure the performance using standard (IR) metrics:
   - Precision of the
     - top 5 recommended tags (P@5), Mean Reciprocal Rank (MRR), Mean Average Precision (MAP)
4. Test the results for statistical significance using Student’s T-test, relative to the baseline (e.g. existing approach, competitive approach)
[Guy et al.] shows another **example** of a similar evaluation approach.

Here, the different **strategies** differ in the way people and tags are used in the strategies: with these tag-based systems, there are complex relationships between users, tags and items, and strategies aim to find the relevant aspects of these relationships for modeling and recommendation.

Here, their **baseline** is the strategy of the ‘**most popular**’ tags: this is a strategy **often used**, to compare the globally most popular tags to the tags predicted by a particular personalization strategy, thus investigating whether the personalization is worth the effort and is able to outperform the easily available baseline.

[Guy et al. Social Media Recommendation based on People and Tags, SIGIR’10]
user interactions (level & type) instead of general social context - better for recommendations?

does hybrid always work worse?
Recommendation Systems

Predict items that are relevant/useful/interesting (and to what extent) for given user (in a given context)

it’s often a ranking task
MONDAY, JUNE 23, 2008

New Personalized Homepage and Improved Email Notifications

The new features included in our most recent site update are geared toward a more personalized experience for you. Check out the details...

NEW PERSONALIZED HOMEPAGE

Our goal is to give a simple answer to the question, “What should I watch today?” with a personalized, customizable homepage that makes it easy to find the videos and people that you care about. After reviewing results from our beta test and feedback from the community, we're happy to now release the new personalized homepage to all logged-in users. Since launching the beta version in February, data has shown an increase in the number of users visiting the homepage, the frequency of visits, and the number of subscriptions users make over time. So, this optimized version of the homepage not only means a customizable experience for you, but more exposure for your cool videos and channels on YouTube (Note: if you are logged out or do not have an account you won’t be able to see the personalized homepage).

IMPROVED EMAIL NOTIFICATIONS

The email messages you receive from YouTube have been updated with a new layout. In addition to a streamlined design and many wording changes, any comments or messages are now included directly within the body of the email, saving you time.

We're eager to hear your feedback, so please let us know what you think of this latest round of improvements.

The YouTube Team
Collaborative Filtering: Problem

→ u1 likes Pulp Fiction?
Collaborative Filtering

- Typical assumption:
  - It is too difficult to represent content and your content preferences
    - Or, do you know the difference between
      - White metal
      - Black metal
      - Thrash metal, speed metal
      - Death metal
      - Power metal
      - Doom and gothic metal
      - ....

?  ➔ Don’t even try
Representing content in collaborative filtering

- An object is represented by who likes it how much
- $\text{PulpFiction}^T = (\text{null}, 5, 1, \text{null}, \text{null}, 2, 5, \ldots)$

Cold Start Problem: New movie
- $\text{Oblivion}^T = (\text{null}, \text{null}, \text{null}, \ldots)$
- No one has rated it yet because it will only be released in 2013
Representing users in collaborative filtering

- A user is represented by what he likes
- JohnSmith^T=(null,5,1,null,null,2,5,....)

- Cold Start Problem: New user
  - SteffenStaab^T=(null,null,null,....)
  - I have not rated any movie yet
Collaborative Filtering

- **Memory-based**: User-Item matrix: ratings/preferences of users => compute similarity between users & recommend items of similar users

- **Model-based**: Item-Item matrix: similarity (e.g. based on user ratings) between items => recommend items that are similar to the ones the user likes

- **Model-based**: Clustering: cluster users according to their preferences => recommend items of users that belong to the same cluster

- **Model-based**: Bayesian networks: \( P(u \text{ likes item B} | u \text{ likes item A}) = \) how likely is it that a user, who likes item A, will like item B learn probabilities from user ratings/preferences

- **Others**: rule-based, other data mining techniques
Social networks & interest similarity

- Limitations of collaborative filtering:
  - ‘cold start’ and
  - ‘sparsity’
  - the lack of control (over people who share some, but not all of my interests) is also a problem, i.e. cannot add ‘trusted’ people, nor exclude ‘strange’ ones

- ‘social recommenders’: presence of social connections defines the similarity in interests (e.g. social tagging CiteULike):
  - Rationale: homophily = birds of a feather flock together
  - does a social connection indicate user interest similarity?
  - how much users interest similarity depends on the strength of their connection?
  - is it feasible to use a social network as a personalized recommendation?

[Lin & Brusilovsky, Social Networks and Interest Similarity: The Case of CiteULike, HT’10]
Conclusions

- pairs unilaterally **connected** have more common information items, metadata, and tags than non-connected pairs.
- the similarity was largest for **direct connections** and
- decreased with the increase of distance between users in the social networks
- users involved in a **reciprocal** relationship exhibited significantly larger similarity than users in a unidirectional relationship

- *traditional item-level similarity may be less reliable way to find similar users in social bookmarking systems*
- *items collections of peers connected by self-defined social connections could be a useful source for cross-recommendation*