An Introduction to Recommender Systems

Huizhi (Elly) Liang
The University of Melbourne
26 May, 2015
Agenda

1. Introduction
2. Recommendation Approaches
3. Evaluation Approaches
4. Recommender Systems in Web 2.0
5. Conclusion
Introduction
Information Overload

- Catalog
  - Yahoo
- Search Engine
  - Google, Bing
- Recommender Systems
  - Amazon, Netflix
  - Google personalized news

Question 1: Search Engine V.s. Recommender Systems
Search Engine vs. Recommender System

- User will try search engine if
  - they have specific needs
  - they can use keywords to describe needs

- User will try recommender system if
  - they do not know what they want now
  - they can not use keywords to describe needs
Are they effective?

(Celma & Lamere, ISMIR 2007)

- Netflix
  - 2/3 rated movies are from recommendation
- Google News
  - 38% more click-through are due to recommendation
- Amazon
  - 35% sales are from recommendation

Question 2: What can be recommended?
What can be recommended?

- Advertising messages
- Investment choices
- Restaurants
- Cafes
- Music tracks
- Movies
- TV programs
- Books
- Cloths
- Supermarket goods
- Tags
- News articles
- Online mates (Dating services)
- Future friends (Social network sites)
- Courses in e-learning
- Drug components
- Research papers
- Citations
- Code modules
- Programmers
2 Recommendation Approaches
Recommender systems

- Recommendation approaches
  - Content based
    - Term vector model
  - Collaborative Filtering (CF)
    - User-KNN & Item-KNN
    - Matrix Factorization techniques
  - Hybrid
Content-based Method

- **Text information**
  - Web pages, blogs, micro-blogs, reviews, descriptions, comments, title, abstract, keywords

- **Multi-media information**
  - Audio, Video, Image

- **Main Idea**
  - Content matching between the topics of item and user

- **Steps**

  ![Diagram with steps](Image)

  - **User Profiling**
  - **Cosine Similarity**
  - **Rank**
  - **Term Vectors**
Example

A reluctant hero. An epic journey. A Choice between the life he left behind and the incredible new world he’s learned to call home. James Cameron’s Avatar – the greatest adventure of all time.

Term Vectors

<table>
<thead>
<tr>
<th>hero</th>
<th>...</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dream</th>
<th>...</th>
<th>crime</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>...</td>
<td>1</td>
</tr>
</tbody>
</table>

Cosine Similarity

Recommendation

0

Question 3: Strength & Weakness

Acclaimed filmmaker Christopher Nolan directs an international cast in this sci-fictioner that travels around the globe and into the world of dreams. Dom Cobb (Leonardo DiCaprio) is the best there is at extraction: stealing valuable secrets inside the subconscious during the mind’s vulnerable dream state. His skill has made him a coveted player in industrial espionage but also has made him a fugitive and cost him dearly. Now he may get a second chance if he can do the impossible: inception, planting an idea rather than stealing one. If they succeed, Cobb and his team could pull off the perfect crime. But no planning or expertise can prepare them for a dangerous enemy that seems to predict their every move. An enemy only Cobb could have seen coming.
Content-based Method

- **Strength**
  - Easy for text
  - Efficient & Scalable
  - Only rely on target user’s history or information needs

- **Weakness**
  - Difficult for multimedia & Noise
  - Always recommend similar items & Poor Novelty
  - Attacks
K-Nearest Neighborhood

  - Rating information
  - Many people liked “Avatar”
  - Can you predict how much I like it?

- **Main idea**
  - Pick 20-50 people that have similar taste with me
  - How much I like depends on how much they like

- **Steps**
  
  ![Diagram](Diagram.png)

  - **User Profiling**
  - **Cosine Similarity**
  - **Rating Prediction**

  
  ![Customer Reviews](Customer_Reviews.png)

  - [Write a customer review](#)
Example

- **Step 1**
  - Rating vector

- **Step 2**
  - Similarity of Elly and all other users

- **Step 3**
  - Select 2-nearest users: user ?&?

- **Step 4**
  - Average ratings of all neighbor users
  - ?
Item-based CF (2001, deployed at Amazon)

- I have watched so many good & bad movies (e.g., Avatar)
- Would you recommend me watching “Inception”?

The idea

- pick from my previous list 20-50 movies that share similar audience with “Avatar”
- how much I will like depend on how much I liked those early movies
- People who have watched those movies also liked this movie
Example

- **Step 1**
  - Rating vector

- **Step 2**
  - Similarity of Elly’s items and all other items

- **Step 3**
  - Select 2-nearest items: item 6 & ?

- **Step 4**
  - Average ratings of all similar items (rank)
  - ?

**Question 4: Strength & Weakness**
K-Nearest Neighborhood

- **Strength**
  - Content free
  - Good Novelty
  - Easy to understand

- **Weakness**
  - Cold start
    - New user
    - New item
  - Scalability
Matrix Factorization

Collaborative filtering (3)

~ [0.1 0.3 0.2 0.9 0.5 0.4 0.7 0.3 0.8 1.5]

Matrix Factorization (2006, Netflix challenge)

- You may have watched thousands of movies
- But perhaps I can tell these movies belong to 10 groups, like Action, Sci-Fi, Animation, etc, ...
- So 10 numbers are enough to describe your taste
- Likewise, “Titanic” has been watched by millions of people, but perhaps …10 numbers are enough to describe its features
- Magic: these hidden aspects can be discovered automatically by Matrix Factorization!
Singular Value Decomposition (SVD)

- Handy mathematical technique that has application to many problems
- Given any $m \times n$ matrix $R$, find matrices $U$, $I$, and $V$ such that
  \[ I = U I V^T \]
  - $U$ is $m \times r$ and orthonormal
  - $I$ is $r \times r$ and diagonal
  - $V$ is $n \times r$ and orthonormal
- Discard the smallest $I$ values to get $R_{m,k}$ with $k \ll r$
  - $R_{m,k}$ is an $k$-approximation of $R$ based on $k$ most important latent feature
  - Recommendation will be given based on $R_k$

\[
\begin{pmatrix}
R
\end{pmatrix} = 
\begin{pmatrix}
\vdots \\
U
\end{pmatrix} \begin{pmatrix}
I_1 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & I_r
\end{pmatrix} \begin{pmatrix}
V \\
\end{pmatrix}^T
\]
Step 1
- User’s rating vector

Step 2
- Predictor

Step 3
- Minimize cost function using gradient descent
Hybrid approaches

- **Weighted**
  - scores of several recommenders are combined together

- **Switching**
  - switch between recommenders according to the current situation

- **Mixed**
  - present recommendations that are coming from several recommenders

- **Cascade**
  - one recommender refines the recommendations given by another
Practice at home

- Large rating dataset
  - MovieLens Dataset
  - Netflix Challenge
  - KDD Cup 2011&2012

- Tips
  - Open source recommender systems
    - Mahout Taste
    - MyMediaLite Recommender System Library
Evaluations Approaches
Evaluation Metrics

- **User Satisfaction**
  - Subjective metric
  - Measured by user survey or online experiments

- **Prediction Accuracy**
  - Measured by offline experiments
    - Training (80%) & Test split (20%)
  - Top-N Recommendation
    - Precision: size (hit set)/size (recommendation list)
    - Recall: size(hit set)/size(test set)
  - Rating Prediction
    - Mean Absolute Error (MAE)
    - Root Mean Absolute Error (RMSE)
Evaluation Metrics

- Diversity
- Coverage
- Novelty
- Trust & Robust
- Real-time

- Making new algorithms by 3 steps experiments:
  - Offline testing
  - User survey
  - Online testing
4 Recommender Systems in the Web 2.0
Web 2.0 and Social Media

- **Web 2.0 [Oreiley03]**
  - It’s all about **people**
    - Joining online communities
    - Connecting to each other on social networks
    - Creating content as in wikis and blogs
    - Annotating content with tags, comments, ratings

- **Social Media**
  - Refers to Web 2.0 sites that allow users to share and interact
  - Characterized by
    - User-centered design
    - User-generated content (e.g., tags)
    - Social networks and online communities

Question 5: new user information?
A question: Can we make use of these information sources?

- Blogs
- Social Media
- Online comments
- Online stores
- Review sites
- Locations
- Mobility
Amazon @ Web 2.0

- **Tags**
  - avatar (256)
  - james cameron (154)
  - science fiction (135)
  - action thriller (84)
  - environmentalism (82)
  - 3d (75)
  - sam worthington (70)
  - zoe saldana (57)
  - cameron (22)
  - gaia (12)

- **Social Networks**

  29,156,163 people like the Avatar movie page on Facebook.
  Khoi-Nguyen Tran, Launa Lee and 31,439,318 others like this.

- **Reviews**

  **🌟🌟🌟🌟🌟 Great Transfer to Blu-Ray** January 28, 2011
  By Thomas Reed

  I got this as soon as I found it available on the net. It will not be available commercially for some time and that, of course, means the price is WAY to high for most viewers. I was willing to be taken for a ride but if you do not just have to have it now I would recommend waiting until it is available everywhere.

  The video quality is fantastic. I have a Samsung 40" 3D setup and the movie was just beautiful to watch. Not quite the same as IMAX but very close (size of screen being the only difference that I could see). The 3D is, to my eyes, exactly as good as the IMAX on-screen version. I am a huge fan of the movie but believe me I would tell you if the video quality was not great.

  I would not hesitate to do the purchase again (even considering the huge rip-off in price at this time) but advise others to consider if you really have to have it right now or can wait awhile.

  Great movie, almost unbelievable video transfer quality, and a price that is just not right!

  Hope this helps.

  21 Comments  |  Was this review helpful to you? [Yes] [No]
Practice at home

- **Recommender System based on tweets**
  - Recommend tweets, users, news, topics

- **Tips:**
  - Content, KNN, Hybrid
  - Text, friendship, social interaction
  - Twitter API for implementation

"PM JG on SKY confirms the carbon tax will only apply to 500 companies in Australia”. @Latika Bourke

✓ “The more we here about this #carbon tax the more it sounds like a camel - a horse designed by a committee. #AUSPOL ”@demonspofforth

✓ “#Carbon price revealed: Finally, here's what it will cost auspol Latest Business & Austr. [http://bit.ly/ngTS6V ] ” @1southerncross
Conclusions
Conclusions

- **Recommender Systems**
  - Content Based
  - Collaborative Filtering
    - KNN based approach
    - Matrix factorization

- **Evaluation**

- **Recommender Systems in Web 2.0**

- **Open Research Questions**
  - Cold start
  - Scalability in the era of Big Data
  - More accurate & better performance
  - Recommender systems & Search Engine
    - Query recommendation
Further Exploration

- **Recommended Readings:**
  - **Books**
    - data mining, information retrieval, machine learning, social science
  - **Classic papers**
    - Item-based Collaborative Filtering Recommendation Algorithms
    - Matrix Factorization Techniques for Recommender Systems.
  - **Conferences**
    - SIGKDD, WWW, SIGIR, ICDM, CIKM