Partnership for International Research and Education
A Global Living Laboratory for Cyberinfrastructure Application Enablement

Enhancing Item-Based Collaborative Filtering by Incorporating Tags

Student: Scott Roepnack, M.Sc. Student, Florida Atlantic University
FAU Advisor: Dr. Shihong Huang, Florida Atlantic University
PIRE International Partner Advisor: Dr. Yong Zhang, Tsinghua University, Beijing, China

National Science Foundation

I. Research Overview and Outcome

Problem Statement

Item-based Collaborative Filtering (IBCF) has become more extensively used because the challenges of User-based Collaborative Filtering (UBCF) have come to a pinnacle in regards to the web, due to the problems, such as scalability. Nearest neighbor algorithms require computation that grows with both the number of users and the number of items. With millions of users and items, a typical web-based recommender system running existing algorithms will suffer serious scalability problems [1]. IBCF does not have as an extensive scalability problem that UBCF does, so it has become more widely used. Even though, IBCF performs the function of recommending items for users, it is not a perfect system.

→ Rating Predictions for user-item pairs are normally on average 25% off from the actual ratings users would provide.

The main objective of this research is to increase recommender system accuracy.

Background

Collaborative Filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple users, viewpoints, and data sources. Collaborative Filtering typically targets application domains that have very large data sets. Collaborative Filtering has several application domains, such as sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in *electronic commerce* and web 2.0 applications where the focus is on user data [2][3].
Making predictions or "filtering", about the interests of a user by collecting small amounts of information from many users (collaborating). The underlying assumption of CF approach is that those who agreed in the past tend to agree again in the future [3].

Research Results

After running several tests we kept having results that showed no changes. Meaning the results were the same if the data set had been run against a *pure* item-based collaborative filtering scheme and when the data set was run against our undirected item-tag recommender.
So what happened? Our problem was concerning the tag density. The one data set provided by Group Lens that contains tags is unbalanced. In other words, the tag density was two orders of magnitude smaller than that of the ratings. Sample data example:

Item Count = n(I) = 10,681
User Count = n(U) = 71,567
Tag Count = n(T) = 100,000
Rating Count = n(R) = 10 million

Tag Density = n(T) / [n(U) * n(I)] = 0.00013082 = ~0.01%

Rating Density = n(R)] / [n(U) * n(I)] = 0.013082 = ~1%

This pitfall has shown us that we need to focus directly on the tag recommender portion of our design. Thus, we have refocused our work to tags. We propose a recommender model that allows users to rate tags (not just items), correct tag syntactical errors, and combine similar tags. We believe this will allow tags to hold more weight in the long run and provide more accurate recommendations because the rated tags could then be evaluated more appropriately. This is a different way of looking at items, ratings, and tags.

Research Objectives

→ Enhance Item-Based Collaborative Filtering by incorporating user defined tags
→ Identify any weaknesses found in recommender system technology

PIRE Objectives

→ Staying safe while abroad
→ Provide international collaboration experience

Undirected Item-Tag Recommender

Our research team decided to integrate tags by creating an undireded hybrid recommender. The undirected system was chosen over directed and cascade because, item-based collaborative filtering is in itself a decent recommender algorithm, and the output of one recommender was not required as input for the other. So the process for a recommendation under our system is as follows:

Step 1: Receive input of a user id and an item id.
Step 2: Use the standard item-based recommender system to predict the rating that the user would provide for the inputted item.
Step 3: Use the inputted user and item again on a tag-based recommender that utilizes clustering. (This scheme will output a weight).
Step 4: Combine the rating generated from Step 2 to the weight generated from Step 3, without exceeding the minimum or maximum rating.
Step 5: Output the finalized rating.

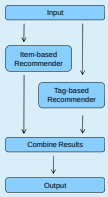


Figure 1: Item-Tag Recommender Flow Chart.

Research Tools

Taste Framework

Taste is a flexible, fast collaborative filtering engine for Java. The engine takes users' preferences for items ("tastes") and returns estimated preferences for other items. For example, a site that sells books or CDs could use Taste to figure out, from past purchase data, which CDs a customer might be interested in listening to [4] using the basic algorithms already Implemented in the Taste framework.
Taste provides a rich set of components from which you can construct a customized recommender system from a selection of algorithms. Taste is designed for performance, scalability and flexibility. It supports a standard EJB interface for J2EE-base applications [4].

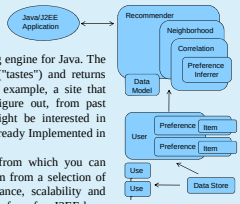


Figure 2: Taste Framework Flow Chart [4].

Sample Data Sets: Group Lens Movie Data Set

The GroupLens Research Project is a research group in the Department of Computer Science and Engineering at the University of . Members of the GroupLens Research Project are involved in many research projects related to the fields of information filtering, collaborative filtering, and recommender systems [5].

Available Data Sets:

→ 100,000 ratings for 1682 movies by 943 users
→ 1 million ratings for 3900 movies by 6040 users
→ 10 million ratings and 100,000 tags for 10681 movies by 71567 users

Future Work

The future will focus on two aspects.

1. Rating Tags

Up until now users have always rated items and those ratings are used to provide recommendations that the user may find appealing. We suggest a recommender system that users will have the ability to:

→ Rate tags
→ Correct tag syntactical errors
→ Combine similar tags

The proposed tagging system will provide more options to the recommendation algorithm.

2. Ripple Recommendation (Graph Recommender)

Current recommender systems rely almost exclusively on:

→ Mathematical algorithms
→ User provided ratings of items
→ Tags

Some research has been done in the area of graph recommendation. This research is still not complete and we believe that by incorporating the rating tag structure into a graph recommender, the system will be able to provide more accurate predictions in the form of ratings for use in the overlying recommendation structure.
Once the graph structure is created, it can be queried for item suggestions for a particular user. Beginning at the node that represents the user in question and *rippling* outward, as if a stone had been tossed into a pond. Therefore, the problem itself would then become a graph search problem, which has already been extensively researched.

II. International Experience

Tsinghua University (清华大学)

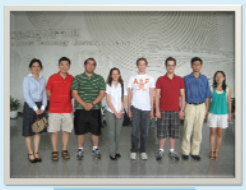
Tsinghua University was established in 1911, originally under the name "Tsinghua Xuetang". Tsinghua University was built on the site of a former royal garden belonging to a prince [6].
The faculty values the interaction between Chinese and Western cultures, the sciences and humanities, the ancient and modern. Most national and international university rankings place Tsinghua as one of the best universities in Mainland China [7].
Personally, I thought the university was wonderful, the size of the campus was daunting, however we did find our way around. Staying at the university gave us an experience like no tourist experiences.

Personal Benefits

The time I spent at Tsinghua University through the PIRE program benefitted me both academically and professionally. In the academic sense I was exposed to new viewpoints, problem solving techniques and a project that guided me towards my thesis. Professionally, my resume has grown in substance, I met new colleagues to network with and I've had an international collaboration experience that most people my age do not even think is possible to achieve.

Tsinghua University's Campus

Tsinghua University's Old Gate

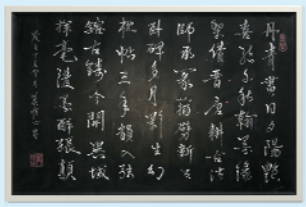


PIRE China at Tsinghua

Research Team

Mandarin Chinese (普通話)

The most common dialect in China is Mandarin. There are other dialects, such as Cantonese. English speakers have a hard time learning Mandarin because of its four tones and complex calligraphy.
While I was there I learned the basics, such as, "Hello" (你好), "Good-bye" (再见), and "Thank-you" (谢谢). I also took Mandarin lessons to better understand the complex language.
During my nine weeks in China, I learned more than I would have learned if I had not been immersed in the language and culture. I learned more practical things than a news version of the language, which seems to be a very common complaint from people who learn a foreign language here in the USA. While I was not able to understand news broadcasts, I did start to pick up daily conversations happening around me.

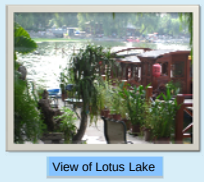


Traditional Chinese Writing

Hou Hai (后海)

One of the oldest areas of Beijing is Hou Hai. In recent years it has become famous for nightlife because it is home to several popular restaurants, bars, and cafes.
Until recently Hou Hai only consisted of a few restaurants. Hou Hai first became popular with modern establishments on the newly constructed Lotus Lake in 2003.
Hou Hai is very popular to both the younger generations and older ones.

Playing Dominos



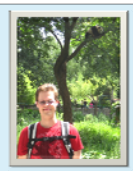
View of Lotus Lake

Beijing, China (北京 , 中国)

As part of the PIRE program we spent nine weeks living in Beijing, China in the foreigner's dorms at Tsinghua University. Beijing is the capital of China and is home to the Summer Palace, Tiananmen Square, Mao's Tomb, the Forbidden City, the 2008 Olympic Park, and many other cultural landmarks.
During our nine weeks stay, we had the opportunity to experience Chinese culture and its excellent food. Some of my favorite Chinese foods were Beijing roast duck (北京烤鸭), delicious chicken and peanuts, hot pot (火锅), and Tsingtao beer (青岛啤酒).
The students from Tsinghua University were very helpful in acclimating us to life in China. Without them we would not have survived the first few weeks when we were completely unable to communicate with most people, let alone read anything. The students showed us several cultural sites, as well as places to just relax.

Tiananmen Square (天安门广场)

The Summer Palace (颐和园)



Beijing Zoo (北京动物园)

The Forbidden City (故宫博物院)

III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

IV. References

[1] Item-based Collaborative Filtering Recommendation Algorithms http://www10.org/cdrom/papers/519/
[2] Collaborative Filtering Explained http://web4.cs.ucl.ac.uk/staff/jun.wang/blog/topics/research-resources/collaborative-filtering/
[3] Collaborative Filtering Wiki http://en.wikipedia.org/wiki/Collaborative_filtering#Item_based_filtering
[4] Taste Framework http://taste.sourceforge.net/old.html
[5] Group Lens Research http://www.grouplens.org/
[6] Tsinghua University History http://www.tsinghua.edu.cn/eng/about.jsp?boadid=32&bid2=3201&pageno=1
[7] Chinese University Rankings http://www.chinaeducenter.com/en/universityranking1.php
[8] Tsinghua University Wiki http://en.wikipedia.org/wiki/Tsinghua_University#cite_note-chinaeducenter.com-1