Online Comment Analysis for Recommendation

Tushar Mehmi

Dronachaya College of Engineering, Gurgaon

Abstract— **This Research paper analysis the use of soft computing techniques to develop recommendation systems. Information on the Internet grows rapidly and users should be directed to high quality Websites those are relevant to their personal interests. However, there is no way to Judge these web pages. Displaying quality content to users based on ratings or past Search results are not adequate. There's a lacking of powerful automated process combining human opinions with machine learning of personal preference. The ongoing rapid expansion of the Internet greatly increases the necessity of effective recommender systems for filtering the abundant information. Extensive research for recommender systems is conducted by a broad range of communities including social and computer scientists, physicists, and interdisciplinary researchers. Despite substantial theoretical and practical achievements, unification and comparison of different approaches are lacking, which impedes further advances.**

I. INTRODUCTION

The study of recommender systems and information filtering in general is no exception with the interest of physicists steadily increasing over the past decade. The task of recommender systems is to turn data on users and their preferences into predictions of users' possible future likes and interests. The goal of this paper is to study recommendation engines and identify the shortcomings of traditional recommendation engines and to develop a web based recommendation engine by making use of user based collaborative filtering (CF) engine and combining context based results along with it. The system makes use of numerical ratings of similar items between the active user and other users of the system to assess the similarity between users' profiles to predict recommendations of unseen items to active user.

A. Recommendation Systems

Recommendation system is an information filtering technique, which provides users with information, which he/she may be interested in. Most of the recommendation systems can be classified into either User based collaborative filtering systems or Item based collaborative filtering systems (Billsus, 1998). In user based collaborative filtering a social network of users sharing same rating patterns is created. Then the most similar user is selected and a recommendation is provided to the user based on an item rated by most similar user. In item based collaborative filtering relationship between different items is established then making use of the active user's data and the relationship between items a prediction is made for the active user (Machine, 2008).

B. Collaborative filtering

Collaborative filtering techniques collect and establish profiles, and determine the relationships among the data according to similarity models. The possible categories of the data in the profiles include user preferences, user behavior patterns, or item properties. Collaborative filtering solves several limitations in content-based filtering techniques (Balabanovic & Shoham, 1997), which decides user preference only based on the individual profile. Collaborative filtering has been expressed in different terminologies in literatures. Social Filtering and Automated Collaborative Filtering (ACF) are two frequently referred terminologies. Collaborative-filtering-based recommendation systems are also referred as Collaborative Filtering Recommender systems and Automated Collaborative Filtering systems.

### C. Applications of Recommender Systems

C.1 Netflix Prize

There are several lessons that we have learned in this competition. Firstly, the company gained publicity

and a superior recommendation system that is supposed to improve user satisfaction. Secondly, ensemble methods showed their potential of improving accuracy of the predictions. Thirdly, we saw that accuracy improvements are increasingly demanding when RMSE drops below a certain level. Finally, despite the company's effort, anonymity of its users was not sufficiently ensured. As a result, Netflix was sued by one of its users and decided to cancel a planned second competition.

II. FIGURES

A new approach is designed to comprise both content-based and collaborative filtering techniques in order to provide the accurate prediction on user preferences. The decisions of how accurate the predictions are depend on the subjective opinions from the users. A recommendation system including both technologies is a hybrid recommendation system (Balabanovic & Shoham, 1997). Hybrid methods solve the problem of extreme case coverage that collaborative filtering techniques unable to handle.
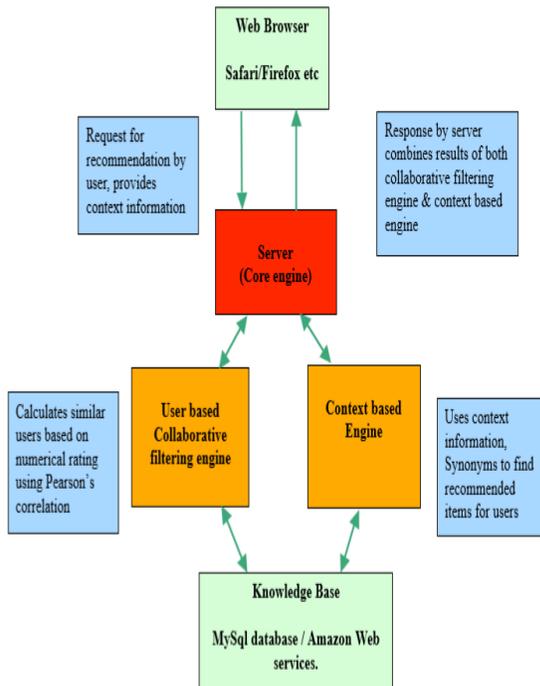
III. METHODOLOGIES

The proposed system makes use of Pearson's correlation to implement User based collaborative filtering, and context, Synonym finder to implement Context based filtering techniques to generate recommendations for the active user.

A. Tags (free tagging)

In order to calculate similar users for the active user we first reduce the three ratings for any movie to a single movie rating between zero and one, after that we generate a user/movie matrix(Pereira, 2006) as shown in the following fig
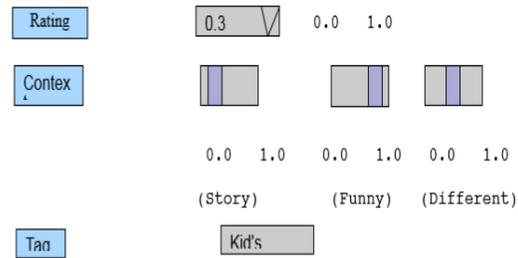
Figure 4: Movie rating parameters

The below figure shows the user similarity matrix in which the ratings between different users are listed. Now in order to calculate similar users we define to be a partition set where, $\alpha>0$ for example let α ={0.4,0.5,0.8,0.9,1.0}.
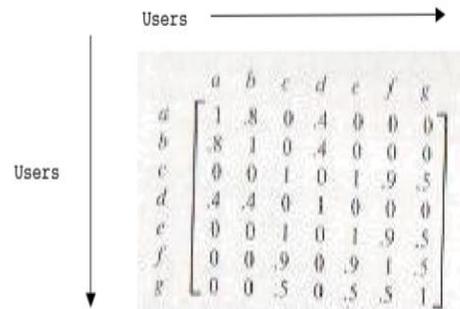

Figure 6: User similarity matrix(Klir, 1988)

B. Pearson's Correlation

The way to find out similar users. The correlation is a way to represent data sets on graph. Pearson's correlation   is x-y axis graph where we have a straight line known as the best fit as it comes as close to all the items on the chart as possible. If two users rated the books identically then this would result as a straight line (diagonal) and would pass through every books rated by the users. The resultant score is this case is 1. The more the users disagree from each other the lower their similarity score would be from 1. Pearson's Correlation helps correct grade inflation. Suppose a user 'A' tends to give high scores than user 'B' but both tend to like the book they rated. The correlation could still give perfect score if the differences between their scores are consistent.

$$r = \frac{\sum_{i=1}^{n} \left((x_i - \bar{x})(y_i - \bar{y}) \right)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Pearson's Correlation Formula

C. Context Engine

This scheme was initiated with an item based collaborative filtering approach example: Amazon related books etc. The item based collaborative filtering approach was build using Pearson's correlation, but instead of calculating similarity between users here we calculated similarity between items. The results were good but it did not meet the goals set for the context-based engine initially. The system did not give good results due to lack of ratings, the system did not fill up the deficiencies of the CF based engine, the system did not do justice to the word 'related' items, because of all these reasons the below approach was followed. This engine makes use of contextual information provided by the user, synonyms, metadata about the products to find recommended items.
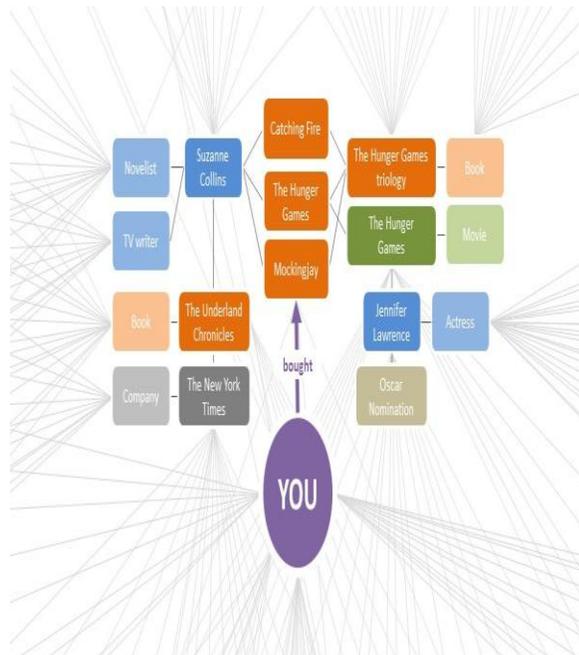
Illustrative image for Context Engine

IV. EVALUATION METRICS FOR RECOMMENDATION

A. Rating and Ranking Correlations

to calculate the correlation between the predicted and the true ratings. There are three well known correlation measures, namely the Pearson product-moment correlation, the Spearman correlation and Kendall's Tau. The Pearson correlation measures the extent to which a linear relationship is present between the two sets of ratings.
It is defined as

$$PCC = \frac{\sum_{\alpha}(\tilde{r}_{\alpha} - \bar{\tilde{r}})(r_{\alpha} - \bar{r})}{\sqrt{\sum_{\alpha}(\tilde{r}_{\alpha} - \bar{\tilde{r}})^2}\sqrt{\sum_{\alpha}(r_{\alpha} - \bar{r})^2}},$$

where $r\alpha$ and $r\tilde{\alpha}$ are the true and predicted ratings, respectively. The Spearman correlation coefficient ρ is defined in the same manner as the Pearson correlation, except that $r\alpha$ and $\tilde{}~r\alpha$ are replaced by the ranks of the respective objects. Similarly to the Spearman correlation, Kendall's Tau also measures the extent to which the two rankings agree on the exact values of ratings. It is defined as
$\tau = (C - D)/(C + D)$ where C is the number of concordant pairs—pairs of objects that the system predicts in the correct ranked order and D is the number of discordant pairs—pairs that the system predicts in the wrong order. $\tau = 1$ when the true and predicted ranking are identical and $\tau = -1$ when they are exactly opposite. For the case when objects with equal true or predicted ratings exist, a variation of Kendall's Tau was proposed in

$$\tau \approx \frac{C - D}{\sqrt{(C + D + S_T)(C + D + S_P)}},$$

where S_T is the number of object pairs for which the true ratings are the same, and S_P is the number of object pairs for which the predicted ratings are the same.

B. Rank-weighted Indexes

Since users have limited patience on inspecting individual objects in the recommended lists, user satisfaction is best measured by taking into account the position of each relevant object and assign weights to them accordingly. Here we introduce three representative indexes that follow this approach.

Half-life Utility:- The half-life utility metric attempts to evaluate the utility of a recommendation list to a user. It is based on the assumption that the likelihood that a user examines a recommended object decays exponentially with the object's ranking. The expected utility of recommendations given to user i hence becomes

$$HL_i = \sum_{\alpha=1}^{N} \frac{\max(r_{i\alpha} - d, 0)}{2^{(o_{i\alpha}-1)/(h-1)}},$$

where objects are sorted by their recommendation score r͂$_{i\alpha}$ in descending order, $o_{i\alpha}$ represents the predicted ranking of object α in the recommendation list of user i, d is the default rating (for example, the average rating), and the "half-life" h is the rank of the object on the list for which there is a 50% chance that the user will eventually examine it. This utility can be further normalized by the maximum utility (which is achieved when the user's all known ratings appear at the top of the recommendation list). When HL_i is averaged over all users, we obtain an overall utility of the whole system.

V. WEB SERVICES

Amazon Web Services:-

The system makes use of the Amazon Rest (representational state transfer) web service ecs4.0 to fetch metadata about the book. Yahoo Web Search Services

Allows the user to tap into the Yahoo! Search technologies from other applications. Related Suggestion/ Term extraction returns suggested queries to extend the power of a submitted query, providing variations on a theme to help you dig deeper. I tried to make use of yahoo web service in order to get related/main keywords, so that these keywords could be used to search the free tags entered by users. This would help to improve the results of context based engine and in turn would help to provide better recommendations.

VI. CONCLUSION

Different collaborative filtering techniques have been proposed to decrease the processing time and the data latency. The results from different recommendation systems indicate that collaborative filtering techniques afford the systems enough ability to provide recommendations to users. The Internet has become a major resource in modern business, thus electronic shopping has gained significance not only from the entrepreneur's but also from the customer's point of view. The system can be highly improved by making use of caching mechanisms, user clustering which will definitely boost the speed of the system, using yahoo term extraction web service to parse and get important keywords from the feedback provided by the user for an item and utilizing these keywords in context based engine.

REFERENCES

(Periodical style)
Christakou, C., & Stafylopatis, A. (2005). A hybrid movie recommender system based on neural networks. Paper presented at the 5th International Conference on Intelligent Systems Design and Applications, 500-505. doi:10.1109/ISDA.2005.9

Droux, S. (2008). Vogoo - recommendation engine and collaborative filtering. Retrieved March 30, 2008, from http://www.vogooapi.com/

E. Peis, J. M. Morales-del-Castillo (2008) Semantic Recommender Systems. Analysis of the state of the topic. Web: www.hipertext.net, ISSN 1695-5498.

Gediminas Adomavicius, Alexander Tuzhilin (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the- Art and Possible Extensions. IEEE, Transactions on Knowledge and Data Engineering, 734-749.

GroupLens research. (2008). Retrieved March 29, 2008, from http://www.grouplens.org/

Klir, G., & Folger, T. (1988). Equivalence and similarity relations. Fuzzy sets, uncertainty, and information (1st ed., pp. 82-85). Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Pereira, R., Ricarte, I., & Gomide, F. (2006). Fuzzy relational ontological model in information search systems. In E. Sanchez, & F. Harmelen (Eds.), Fuzzy logic and the semantic web (1st ed.,). New York,

NY, USA: Elsevier Science Inc. Retrieved from http://books.google.com/books?hl=en&lr=&id=Cidej 8b4ESIC&oi=fnd&pg=
PA395&dq=%22Pereira%22+%22Fuzzy+Relational +Ontological+Model+in+In
formation+Search+...%22+&ots=mt7470caQ3&sig= Yx8CZ49DEqnzS7ICaI5D_p DDPko#PPA395,M1

Sean, O. (2008). Taste collaborative filtering for java. Retrieved March 26, 2008, from http://taste.sourceforge.net/

Stuart E. Middleton, Nigel R. Shadbolt (2004). Ontological user profiling in recommender systems. ACM Transactions on Information Systems, 54-88.

Amzn Investor Relations. (2004). Amazon.com Apparel & Accessories Store Reaches One Millionth Customer Account. Amazon.com 2004 Press Releases. Retrieved 17 August 2004, from http://www.amazon.com.

Armstrong, R., Freitag, D., Joachims, T., & Mitchell, T. (1995). Webwatcher: A learning apprentice for the world wide web. AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, 6-12.

Balabanovic, M., & Shoham, Y. (1997). Fab: Content-based collaborative recommendation. Communications of the ACM , 40(3), 88-89.

Breese, J., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithm for collaborative filtering. Microsoft Technical Report MSR-TR-98-12.