Recommender System Using Item-Based Collaborative Filtering

Cherry Soe, Thet Thet Khaing
Computer University (Hinthada)
sagawar.htd@gmail.com

Abstract

The tremendous growth in the amount of available information and the number of visitors to Web sites in recent years poses some key challenges for recommender systems. Recommender systems form a specific type of information filtering (IF) technique that attempts to present information items (movies, music, books, news, images, web pages, etc.,) that are likely of interest to the user. They produce high quality recommendation and perform many recommendations per second for millions of users and items and achieving high coverage in the face of data sparsity. Most systems are implemented using the k-nearest neighbor collaborative filtering but have some weakness in searching on the Web. To address these issues, item-based collaborative filtering techniques have been explored. Firstly, item-based techniques analyze the user-item matrix to identify relationship between different items and then use these relationships to compute indirectly the user's profile to some reference characteristics, and seek to predict the 'rating' that a user would give to an item they had not yet considered.

Keyword: Collaborative Filtering, Recommender, e-Commerce, Data-Mining, knowledge Discovery.

1. Introduction

The amount of information in the world is increasing far more quickly than our ability to process it. All of us have known the feeling of being overwhelmed by the number of new books, journal articles and conference proceeding coming out each year. One of the most promising such technologies is Collaborative Filtering (CF).

Collaborative filtering works by building a database of preference for items by users. A new user, Neo, use the database to discover neighbors, which are other users who have historically and similar taste to Neo. Items that the neighbors like are then recommended to Neo, as he will probably like them. Collaborative filtering has been very successful in both research and practice and in both information filtering applied and E-commerce application. Collaborative filtering (CF) query consists of an array of (item, rating) pairs of a single user. The response to that query is an array of predicted (item, rating) pairs for those items the user has not yet rated.

2. Related Work of Collaborative Filtering

Model based (Item-based) collaborative filtering algorithms provide item recommendation by first developing a model of user ratings. The model building process is performed by different machine learning algorithms such as Bayesian network, clustering, rule-based approaches and etc. The Bayesian networks create a probabilistic model based on a training set with a decision tree at each node and edges representing user information. The model can be built off-line over a matter of hours or days. The resulting model is very small, very fast, and essential as accurate as nearest neighbor methods.

Clustering techniques work by identifying groups of users who appear to have similar preferences. Once the clusters are created, predictions for an individual can be made by averaging the opinions of the other users in that cluster. The rule-based approach applies association rule discovery algorithms to find association between co-purchased items and then generates item recommendation based on the strength of the association between items.

In comparison to memory-based schemes, model-based Collaborative filtering algorithms are typically faster at query time though they might have expensive learning or updating phases. Model based schemes can be preferable to memory based schemes when query speed is crucial. In this paper, we will apply the Slope One algorithm [9] for predicting the items. Our predictions are of the form $f(x) = x + b$, hence the name "Slope One" where b is a constant and x is a variable representing rating values.

3. Collaborative Filtering

Collaborative filtering [10] is the process of predicting ratings based on a database of ratings from various users. It is widely applicable to e-Commerce, e-Learning, and so on. Collaborative filtering is the method of making automatic

predictions (filtering) about the interests of a user by collecting taste information from many users (collaborative).The underlying assumption of CF approach is that these who agreed in the past tend to agree again in the future. For example, a collaborative filtering or recommendation system for music tastes could make given a partial list of that user's tastes (likes or dislikes). Note that these predictions are specific to the user, but use information gleaned from many users.

However, there are the important research questions in overcoming two fundamental challenges for collaborative filtering. The first challenge is to improve the scalability of the collaborative filtering algorithms. These algorithms are able to search tens of thousands of potential neighbors in real-time, but the demands of modern systems are to search tens of millions of potential neighbors. Further, existing algorithms have performance problems with individual users for whom the site has large amount of information. The second challenge is to improve the quality of the recommendations for the users. Users need recommendations that can trust to help them find items they will like. Users will refuse to use recommender systems that are not consistently accurate for them.

Collaborative filtering is a technique used by recommender systems to combine different users' opinions and tastes in order to achieve personalized recommendations. There are at least two classes of collaborative filtering: user-based techniques are derived from similarity measures between users and item-based technique compare the ratings given by different users. Item-based filtering algorithm is divided into three categories called (1) Item-based collaborative filtering of rated resources and over-fitting (2) Item-based collaborative filtering of purchase statistics and (3) Slope one collaborative filtering for rated resources.

3.1 Slope one collaborative filtering for rated resources

The Slope One [9] family of easily implemented Item-based Rating-Based collaborative filtering algorithms was proposed. Essentially, instead of using linear regression from one item's ratings to another item's ratings $(f(x) = ax + b)$, it uses a simple form of regression with a single free parameter $(f(x) = x + b)$, where b is a constant and x is a variable representing rating values. The free parameter is then simply the average difference between the two items' ratings. It was shown to be much more accurate than linear regression in some instances, and it takes half the storage or less.

Slope One Equation

Let rating [i] [j] be rating of item for customer , where i= customer, j= item
Let h be number of historical row
Let p be number of items
Let n be the item that has rating for all customers in number of historical rows (h+1)
Let m be the item that has no rating for one customer in number of historical rows (h+1)

$$A = \sum_{item=1}^{j-1} h_{item} * \left(B + \frac{\left(C + \sum_{k=1}^{hitem} (D - E)\right)}{2} \right)$$

$$S = \sum_{item=j+1}^{p} h_{item} * \left(B + \frac{\left(C + \sum_{k=1}^{hitem} (D - E)\right)}{2} \right)$$

$$B = rating\,[i][item], C = \text{rating}[h_{item} + 1]$$

$$D = \text{rating}[k][n], E = \text{rating}[k][m]$$

$$\text{rating [i] [j]} = \frac{A+S}{\sum_{item=1}^{j-1} h_{item} + \sum_{item=j+1}^{p} h_{item}}$$

For example,
Sample rating database

Customer	Item1	Item2	Item3
John	5	3	2
Mark	3	4	Didn't rate it
Lucy	Didn't rate it	2	5

In this case, the average different in ratings between item 2 and 1 is (2+ (-1) /2) = 0.5. Similarly, the average difference between item 3 and 1 is 3. Hence, to predict the rating of Lucy for item 1 using her rating for item 2, this comes to 2+0.5 =2.5. Similarity, to predict her rating for item 1 using her rating of item 3, this comes to 5+3=8. In the above example, the following rating for Lucy on item 1 would be predicted

$$\frac{2*2.5+1*8}{2+1} = \frac{13}{3} = 4.33$$

Slope One is a family of algorithms used for Collaborative filtering introduced in Slope One Predictors for Online Rating-Based Collaborative Filtering.

4. Implementation for the System

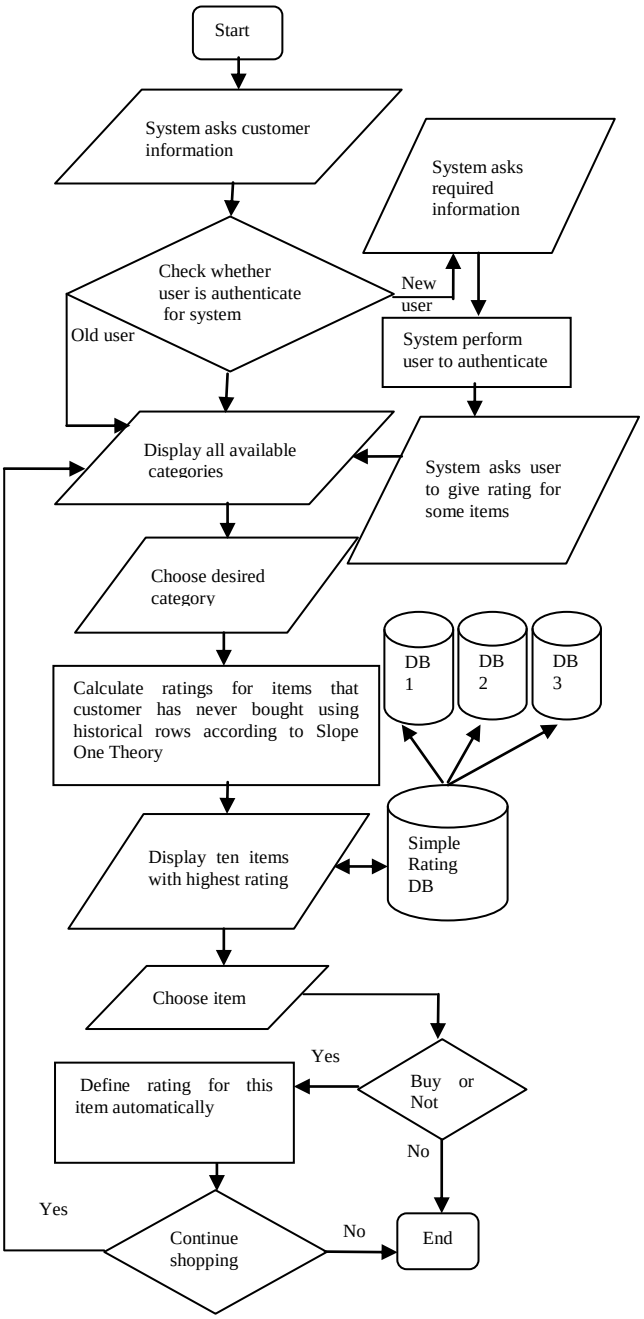

Figure 1: Flow of system implementation

In our system, there are three main steps to give recommendation for users. They are (i) get user rate on item (ii) calculate rating using collaborative filtering (iii) show information rated value with item and category. In this system,

Step1: Start
Step2: The system asks customer information (eg; NRC NO)
Step3: If user is authenticated
3.1: Old user (go to Step4)
3.2: New user (3.2.1: system asks required user information
3.2.2: system perform user to authenticate
3.2.3: system asks user to give rating for some items in all categories and go to Step4)
Step4: Display all available categories
Step5: Choose desired category
Step6: Calculate rating for items that customer has never bought using historical rows according to Slope One Theory
Step7: Display ten items with highest rating
Step8: User choose item
Step9: Does user buy or not?
9.1: Yes (Define rating for this item automatically and go to Step10)
9.2: No (go to Step11)
Step10: Does user continue the shopping?
10.1: Yes (go to step4 for chosen category)
10.2: No (go to step11)
Step11: End

5. Simulation Result for the System

The entity relationship diagram describes the relation of the tables in this system.

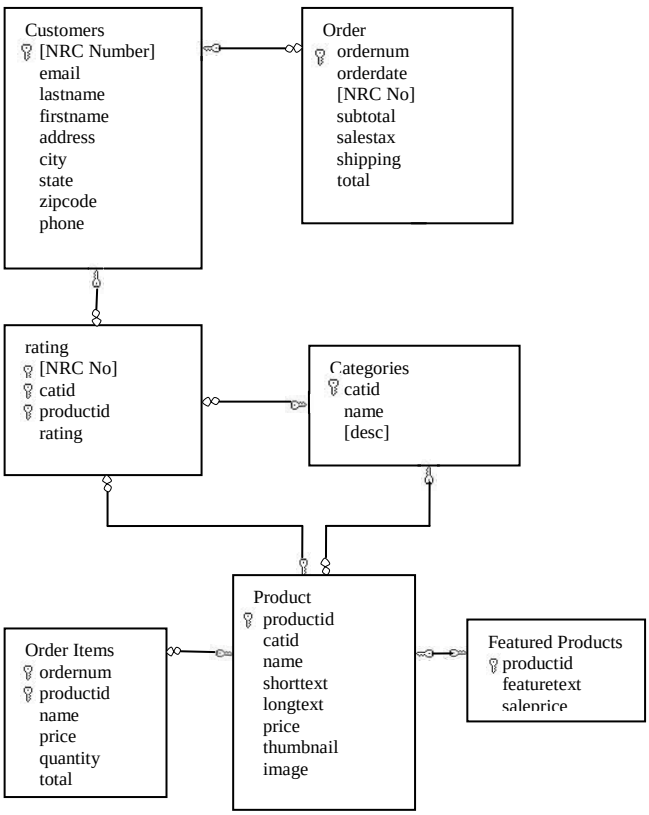

Figure 2: Entity relationship diagram

Three categories (DVDs & Videos, Music or Novels) of testing data and each category have two hundreds data are used.

The user can click DVDs & Videos, Music or Novels, then system display the items.

6. Conclusion

Recommender systems [8] are powerful new technologies for extracting additional value for a business from its user database. These systems help users to find items they want to rate. Recommender systems give benefits to users by enabling them to find item they like. Recommender systems are rapidly becoming a crucial tool for E-commerce on the Web. Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. In traditional collaborative filtering systems, the amount of work increases with the number of participants in the system. The recommender system that uses the item-based collaborative filtering techniques can produce the high quality recommendations.

7. References

[1] Ailin Deng, Yangyoung Zhu, and Baile Shi, "A Collaborative Filtering Recommendation Algorithms Based on Item Rating Prediction", Journal of Software, vol.13, pp.1621-1628, 2003.

[2] Aggarwal, C.C., Wolf, J.L., Wu, K., and Yu, P.S (1999). Horting Hatches an Egg: Anew Graph-theoretic Approach to Collaborative Filtering. In Proceedings of the ACM KDD'99 Conference. San Diego, CA. pp 201-212.

[3] Billsus, D., and Pazzani, M.J (1998). "Learning Collaborative Information Filters". In Proceddings of ICML' 98.pp.46-53.

[4] B.M.Sarwar, G.karpis, J.A.konstan, and J.Riedl."Item-based collaborative filtering recommender algoirithms". In WWW 10, 2001.

[5] Badrul Sarwar "Item-Based Collaborative Filterimg Recommendation Algorithms" GroupLens Research Group/Army HPC Research Center,Department of Computer Science and Enigneering University of Minnesta, Minneapolis, MN 55455.

[6] Basu, C., Hirsh, H., and Cohen,W.(1998). Recommendation as Classification Using Social and Content-based Information in Recommendation. In Recommender System Workshop'98. Pp.11-15.

[7] Breese, J.S., Heckerman, D., and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Processings of the 14th Conference on Uncertanity in Artifical Intelligence, pp.43-52.

[8] Chan Hein Zaw,"Recommender system based on rating items by using TOP-N recommendation algorithms". (April, 2009).

[9] Daniel Lemire*, Anna Macla chlan *,"Slope One Predictors for Onlinr Rating-Based Collaborative Filtering",February 7,2005.

[10] D.Billsus and M.Pazzani.learning collaborative information filtering. In AAAI Workshop on Recommender Systems, 1998.