Study of Recommendation Engines for E-Commerce Websites

Varisht Ghedia[1], Gaurang Pansare[2], Dhaval Mehta[3] and Prof. Amruta Mhatre[4] (Co-Author)
Atharva College of Engineering (Malad)
Mumbai University, Mumbai, India

***Abstract*—Recommender systems used by majority e-commerce websites have many drawbacks and limitations and are inaccurate. Since they learn from the customer's browsing habits to come up with recommendations, they tend to recommend products more from categories that the user has visited and purchased from before. For example, if you look for any mobile on any e-commerce website, you keep getting recommendations for mobiles even after you've bought it through the website. This is due to their reliance on older feedback-based, clustering based or collaborative filtering based algorithms. Session based collaborative filtering tries to solve major drawbacks offered by these algorithms and obtains relevant recommendations for users.**

Keywords — Recommender system; Collaborative filtering; Item based filtering; Clustering models; Item to Item collaborative filtering; Implicit feedback; Explicit ratings

I. INTRODUCTION

A product recommender engine is widely used by many e-commerce websites. Recommender engines have been developed for recommending songs, movies, TV shows and a variety of other content which involves some degree of personalization. Its basic task is to identify the needs the of the users and recommend them on the basis of what other users looked at, were interested in, liked it and a host of other pre-determined parameters.

It is a serious business tool that is reshaping the world of e-commerce. Many large e-commerce websites use recommender systems to help their customers easily find products they might be interested in purchasing, leading to faster checkouts and more revenue for the store.

Basically a recommender system is an intelligent system which seeks to identify and serve the user's need based on user analysis and user information analysis. To perform such an analysis there exist collaborative as well as content oriented systems. This artificial intelligence digs through huge amounts of user data (visits, purchases, ratings, etc.), applying various algorithms [2][3] to come up with recommendations.

II. LITERATURE REVIEW

DEFINITIONS:

A. *Recommender System:*

A recommendation engine is a feature that filters items by predicting how a user might rate them. It solves the problem of connecting existing users with potential items of interest from the massive inventory (i.e. tens of thousands to millions) of products or content.

B. *Collaborative filtering (CF):*

This technique finds users with 'similar taste' and then uses that information to generate recommendations.

C. *Content based filtering:*

This technique works in two parts.

a. It builds a user profile for the users based on their feedback and other information.
b. It profiles the products based on its characteristics.

This information is used to match appropriate users to appropriate products.

D. *Cluster analysis (Clustering):*

The task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

E. *Feedback:*

Feedback is the measure of customer satisfaction which can be used as a basis for improvement. It can be of two types – Implicit or Explicit.

F. *Explicit feedback:*

Feedback acquired externally by asking the user to provide feedback.

G. *Implicit feedback:*

Feedback implied by studying user behavior instead of by actually asking the user's opinion.

III. STUDY OF EXISTING SYSTEMS

Recommender systems are used quite extensively these days. As a result of this, this topic has caught the attention of many researchers over the past couple of years. The first research paper on recommender systems [4] dates back to 1998. This paper focused on using implicit feedback in place of explicit ratings given by individual users. The 2003 industry report [7] on amazon.com recommendations by Greg Linden, Brent Smith and Jeremy York from Amazon.com focuses on the algorithms the shopping website employed in order to recommend products to their customers. This paper covers collaborative filtering and clustering models. Here they consider each customer as an N-dimensional vector of items. Here N denotes the number of distinct catalogue items. The array represents purchased or positively rated items as positive factors and negatively for negatively rated items. The algorithm sees to it that the vector for each customer is extremely sparse. In order to find the similarity between two users to provide them recommendations it uses a cosine product of the two vectors to get the similarity between the users. It also concludes that this process is computationally intensive and of an order of O(MN), where M is the number of customers and N is the number of product categories. This can be reduced to the order of O(M+N). However, this gives rise to a number of irregularities and inaccurate results as it assumes that the customers purchase sparse items or rate only a few selected categories. In a real world scenario this is not true. The paper also focuses on clustering models where the customers are divided into various segments and treats the task as a classification problem. It then proceeds to suggest the products to the users of the segments using their ratings and purchases. Once the segments are generated the algorithm computes the vectors that summarize each of the generated segments. Though this model is scalable, the task of clustering and categorizing the user base is computationally difficult when the user base is constantly changing and expanding. The paper also explains Item to Item collaborative filtering. Here the items from the user's shopping cart are used for recommendations. We focused here mainly on the collaborative filtering algorithm.

We also referred to the 2007 paper on "Recommender Systems Based on Consumer Product Reviews" [5]. This paper uses the reviews given by the user to generate a rating for the product. It also rates the quality of the review and tries to extract information from the review to find how likely someone else is going to buy the product. It uses opinion quality metrics to get these results. Our conclusion from this study was that customer reviews are not a full proof metric for recommendations. The recommendations can change drastically on the basis of the quality of the review, the writing skills of the reviewer, the information contained in the review and a many unseen factors.

We then proceeded to look at other approaches such as hybrid methods [6] which used both collaborative filtering, user based and item based filtering to give product recommendations. This faced the drawbacks similar to collaborative filtering and thus we looked for other better approaches towards collaborative filtering.

IV. SUMMARY OF ISSUES ENCOUNTERED

A. *Complexity:*

The big online stores like Amazon.com have huge number of products (probably in billions) and thus it is computationally very expensive to analyze all the products to generate recommendations. This has been a major drawback of collaborative filtering.

B. *Accuracy:*

One way to reduce the computational complexity is by dividing the user space into clusters and thus using clustering analysis. But due to this, the individuality in the recommendations is lost and the user gets recommendations that are generalized for the cluster and are not unique to the said user. Implicit feedback is inaccurate as it does not tell if the customer actually likes the product or not.

C. *Reliability of ratings:*

a. Explicit feedback relies on user input and hence is very unreliable since the criterion for judging varies from user to user and cannot be entirely trusted.
b. On the other hand, implicit feedback avoids the uncertainty of user inputs and human error by studying user behavior instead of taking explicit user feedback. But the quality of feedback depends on the parameters of user feedback considered and they may not always accurately describe the user's satisfaction level.

V. SOLUTION

We later referred to a recent paper (2015) "Incremental Session Based Collaborative Filtering with Forgetting Mechanisms" [1]. This paper identifies the drawbacks of the standard collaborative filtering model and tries to find solutions to them. The major drawback of the collaborative filtering approach was that since it used a lot of old data to provide recommendations, computation was on an exponential level as the data kept on increasing. This approach is not suitable for current generation applications where recommendations are required like, e-commerce, music streaming, video streaming, movie recommendations among various other data intensive applications. In all these applications, a lot of real time data is generated and is considered much more relevant to the application rather than the old data. The example of a music service perfectly illustrates this point. A person is more likely to hear songs on the basis of the songs he has heard in the recent time frame, rather than listening to the songs he listened to 1 or 2 years ago [8]. Songs, movies, products, fashion trends change at a really fast pace [9]. Thus in such a scenario it is much more beneficial to consider the new data instead of the old data.

The paper presents the idea of an incremental session based collaborative filtering technique with forgetting mechanisms. This approach uses sessions instead of the entire user history. This reduces the use of old and stagnant data and uses the data generated in the current and most recent sessions to generate user profiles. This gives the more accurate representation of the user and thus it is able to give much better user recommendations. Another highlight of this algorithm was the forgetting mechanism. The approach used here is that of using sliding windows. In sliding windows, a sliding window of a fixed size is maintained. We use a sliding window with the user's sessions with a first in first out (FIFO) data structure. Once a new incoming session arrives and the sliding window is reached the old session is discarded and the new session is incorporated into the sliding window [10]. In this scenario there exists a small possibility that the old important data may be lost with the deleted session. In order to tackle this problem we use the concept of fading factors. The fading factor approach is a gradual fading or full fading approach which slowly decreases importance of the older sessions by reducing their weight.

VI. CONCLUSION

Thus different techniques have been used to develop recommender systems. An analysis of the different recommender systems has been done as well. We can conclude that a lot of research has been done on collaborative filtering based recommender systems; however these systems are not fit for commercial applications or large scale applications such as e-commerce, music streaming, video streaming and others. Collaborative filtering systems with forgetting mechanisms and an incremental session based approach can address a number of drawbacks of these systems. However, for optimal performance, they need to be modified to some degree for use in recommendation engines of e-commerce meta-search engines.

REFERENCES

[1] Suksawatchon Ureerat, Sumet Darapisut, and Jakkarin Suksawatchon. "Incremental session based collaborative filtering with forgetting mechanisms." 2015 International Computer Science and Engineering Conference (ICSEC). IEEE, 2015.

[2] Sarwar, Badrul, et al. "Analysis of recommendation algorithms for e-commerce." Proceedings of the 2nd ACM conference on Electronic commerce. ACM, 2000.

[3] Mustafa, Ghulam, and Ingo Frommholz. "Performance comparison of top N recommendation algorithms." 2015 Fourth International Conference on Future Generation Communication Technology (FGCT). IEEE, 2015.

[4] Oard, Douglas W., and Jinmook Kim. "Implicit feedback for recommender systems." Proceedings of the AAAI workshop on recommender systems. 1998.

[5] Aciar, Silvana, et al. "Recommender system based on consumer product reviews." Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence. IEEE Computer Society, 2006.

[6] Romadhony, Ade, Said Al Faraby, and Bambang Pudjoatmodjo. "Online shopping recommender system using hybrid method." Information and Communication Technology (ICoICT), 2013 International Conference of. IEEE, 2013.

[7] Linden, Greg, Brent Smith, and Jeremy York. "Amazon.com recommendations: Item-to-item collaborative filtering." IEEE Internet computing 7.1 (2003): 76-80

[8] Park, Chan Ho, and Minsuk Kahng. "Temporal dynamics in music listening behavior: A case study of online music service." Computer and Information Science (ICIS), 2010 IEEE/ACIS 9th International Conference on. IEEE, 2010.

[9] Park, Sung Eun, Sangkeun Lee, and Sang-goo Lee. "Session-based collaborative filtering for predicting the next song." Computers, Networks, Systems and Industrial Engineering (CNSI), 2011 First ACIS/JNU International Conference on. IEEE, 2011.

[10] Vinagre, João, and Alípio Mário Jorge. "Forgetting mechanisms for scalable collaborative filtering." Journal of the Brazilian Computer Society 18.4 (2012): 271-282.