IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

INSIDE RECOMMENDATION SYSTEM: SURVEY, RESEARCH AREA

Vishwas Patel*, Mosin Hasan
* Computer Engineering, Birla Vishwakarma Mahavidyalaya, Vallabh Vidya Nagar, Anand,Gujarat,India.

ABSTRACT
With the increase in E-commerce, Recommendation Systems are getting popular to provide recommendations of various items (movies, books, music) to users. To build the Recommendation System(RS), Collaborative Filtering (CF) techniques are proven efficient. The main two Collaborative Filtering techniques are User-Based and Item-Based, but from survey it can be said that item-based CF provides better recommendations. A novel approach, Ratio-Based CF provides recommendation depending upon the item has more accuracy amongst item-based CF technique. The problem with CF techniques is more execution time i.e. O(mn). To improve the execution time a parallel platform or technique can be adopted to reduce the time complexity of recommendation system. Hence, for better and faster recommendation parallel Ratio-Based Collaborative Filtering algorithm should be used.

KEYWORDS: Recommendation System, Collaborative Filtering, Item-Based, User-Based, Ratio-Based, MovieLens.

INTRODUCTION
In the era of internet, everything is going online whether it is information regarding some topics or internet banking or online shopping. Day-by-day the speed of internet as well as data on that is increasing enormously. E-commerce is the latest trend in today's internet market. People now a days prefer to buy the goods online as they can compare the price and feature of items on various E-commerce websites and can do smart shopping. Along with E-commerce, social networking sites are also getting popular as people can connect to their friends and family through chat, audio calling and video calling.

Due to this growth on internet where customer prefer to buy books, music or any item online Recommendation System(RS) came into the picture. With Recommendation System, on the basis of the person previous choice or purchase of customer, RS can predict item that customer is likely to purchase. Recommender systems or recommendation systems (sometimes replacing "system" with a synonym such as platform or engine) are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that a user would give to an item[13].

Collaborative Filtering (CF) technology is the most widely used in various Recommendation System [2][3][4][7][8][9][12]. Figure 1 shows a simple collaborative filtering process. The first system to use Collaborative Filtering was the Information Tapestry project at Xerox PARC [14]. This system allowed users to find books, music, movie or any item based on previous comments by other users.

RECOMMENDATION SYSTEM
Recommender Systems can be basically implemented one of two ways - through collaborative or content-based filtering [1][2]. Collaborative filtering approaches building a model from a user's past behaviour as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often combined and known as Hybrid Recommender Systems [2].

Recommender Systems are a useful alternative to search algorithms, since they help users to discover items that they might not have found by themselves. Interestingly enough, recommender systems are often

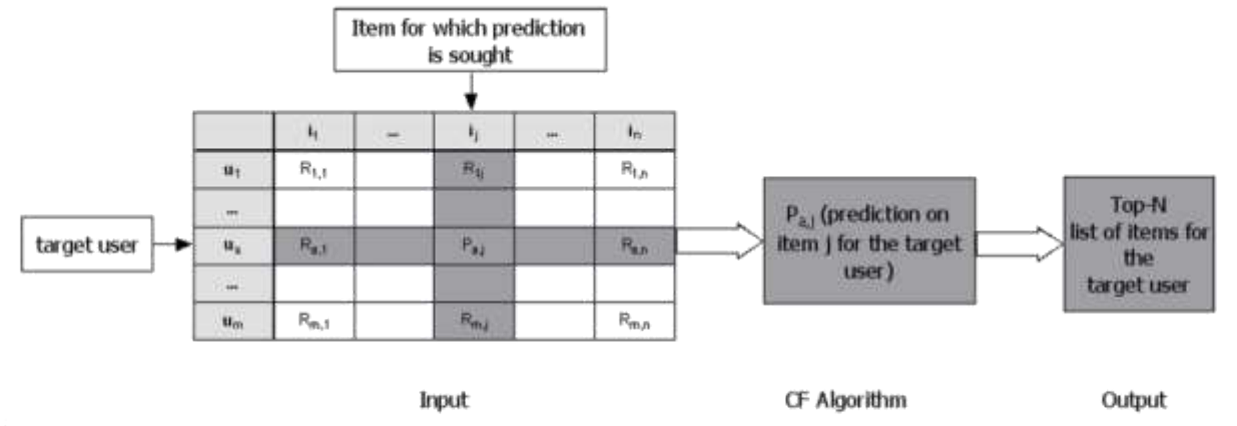

Figure 1 : A Simple Collaborative Filtering Process[7]

implemented using collaborative filtering [2][3][4][7][8][9][12].

COLLABORATIVE FILTERING

Collaborative Filtering (CF) is a technique which is popularly used by Amazon [2] and others. The user information such as ratings, clicks, purchases, or any other past activities helps recommendation system to provide predictions or recommendation for other users. CF is often used to predict items for users - such as books, music, and movies, but it is also useful in other applications where more number of entities need to collaborate to narrow down data.

CF technique can be categories into following [2][4]:

1. **User-based**: Similarity between users is calculated to predict the items. This practice is face scalability problem because of the dynamic nature of users.
2. **Item-based**: Calculates similarity between items and provides predicted item to users as recommendation. Items usually don't change much, so this often can be computed offline.
3. **Slope-One**: A very fast and simple item-based recommendation approach applicable when users have given ratings (and not just Boolean preferences).
4. **Model-based**: Provide prediction or recommendations based on developing a model of users and their ratings.

All CF approaches end up calculating a notion of similarity between users and their rated items. There are many ways to compute similarity, and most CF systems allow you to plug in different measures so that you can determine which one works best for your data [16].

Collaborative Filtering algorithm is a classic recommendation algorithm; it’s widely used in many commercial Recommender Systems. Collaborative Filtering algorithm is an algorithm based on the following concept. Individual person have similar preferences and interest. Their preferences and interests are stable, we can predict their choice according to their past preferences. Also the Collaborative Filtering algorithm is based on the comparison of one, individual person’s behaviour with other individual person’s behaviour, to find his nearest neighbours, and according to his neighbour’s interests or preferences to predict his interests or preferences.

The first step of Collaborative Filtering algorithm is to obtain the users history profile, which can be represented as a ratings matrix where each entry is the rate of a user given to an item. A ratings matrix consists of a table where each row represents a user, each column represents a specific product, and the number at the intersection of a row and a column represents the user’s rating value. The absence of a rating score at this intersection indicates that user has not yet rated the item. Owing to the existence problem of sparse scoring, use the list to replace the matrix. The second step is to calculate the similarity between users and find their nearest neighbours. There are many similarity measure methods. But Pearson Correlation is widely used and served as the benchmark for CF technique [9].

LITERATURE REVIEW

Recommendation System can be build using various CF techniques. Xiaoyuan Su and Taghi M. Khoshgoftaar [1] provide information regarding various CF algorithms with its advantages and disadvantages. Sachin Walunj [2] provides the basic idea about how to build the recommendation system and also he explained various challenges of CF algorithm like data sparsity, scalability, synonyms, grey sheep, shilling attacks, and diversity.

To deal with CF challenges and various researchers have worked on it to improve the performance of Recommendation System on basis of User-Based CF technique. DanEr CHEN[3] used BPNN algorithm to reduce the sparsity of the matrix, also he used cosine measure for calculating similarity and k-nearest neighbour method for predicting item. Vandana Patil [4] introduced a new concept to handle the fake user and to improve the accuracy by using Trust Based CF technique. Another novel method is introduced in which the uses are assigned specific category – Normal User, Most Popular User and Star User which improved the accuracy of CF algorithm by Qiang Liu[5].

Many researchers have also worked on Item-Based CF technique to improve the accuracy of Recommendation System. Badrul Sarwar [6] compared three different item-based CF techniques and compared best item-based CF technique with best user-based CF technique. A new Ratio-Based CF technique Yaqiu Liu[7] used to improve the accuracy to recommendations.

CF algorithm has the scalability problem. To overcome it various approaches used so that it can speed up the execution time. Y.L. Zhang [8] have implemented nearest neighbour method in Hadoop to deal with cold start problem and speedup the execution of recommendations. Zhi-Dan Zhao [9] used Hadoop MapReduce technique to reduce the computational time of CF algorithm. Pearson Correlation Coefficient and Log Likelihood Ratio Similarity CF technique was compared by Dr. G. R. Bamnote [10] using Apache Mahout Platform. Swati Pandey [11] compared item-based and user-based CF technique by separate and combine approach. Jing Jiang [12] Scaling-up Item-based the Collaborative Filtering Recommendation Algorithm based on Hadoop.

COMPARISON ANALYSIS

From literature survey, it has been said that Recommendation System can be build using majorly two CF techniques i.e. User-Based and Item-Based. User-Based CF techniques are k-nearest neighbour, Pearson correlation coefficient, Vector correlation coefficient and so on. Amongst all Pearson correlation coefficient is benchmark of User-Based CF technique [6][9].

While Item-Based CF techniques are cosine based similarity, correlation based similarity, adjusted cosine based similarity and so on. The accuracy of adjusted cosine based similarity is more compared to others [6]. But Yaqiu Liu [7] introduce new concept of Ratio-Based similarities (basically item-based) which is more accurate than adjusted cosine based similarity. Ratio-based similarities have more accuracy but the problem is scalability, it takes more execution time same as other CF techniques.

To overcome the problem of more execution time many researchers have used Hadoop and Mahout platform to reduce the execution time of CF techniques [8][9][10][11][12]. Using Hadoop and Mahout platform can scale up or speed up the execution of the CF techniques.

The most popular dataset used for Recommendation System is MovieLens [3]4][5][6][7][8][10][11]. MovieLens is research site run by GroupLens Research in the computer science department at the University of Minnesota[4][7]. MovieLens consists of 1682 movies (items), 943 users and 10000 rating of users (scale 1 – 5).

RESEARCH GAP

The potential points observed from the literature review are following:

- Item-Based Collaborative Filtering algorithm is more efficient than User-Based Collaborative Filtering [2][7].
- Ratio-Based Collaborative Filtering algorithm is best amongst Item-Based Collaborative Filtering [7].
- The complexity of Collaborative Filtering is O(mn) where m is number of users and n is number of items.
- Collaborative Filtering Algorithm face scalability problem [1][2].
- To overcome it parallel implementation of CF algorithm can be done.
- Hadoop is platform which provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs [8][9][12].
- Apache Mahout is a project of the Apache Software Foundation to produce free implementations of distributed or machine learning algorithms focused primarily in the areas of collaborative filtering, clustering and classification [10][11].

- CUDA, which stands for Compute Unified Device Architecture, is a parallel computing platform and application programming interface (API) model created by NVIDIA [17].
- Hadoop, Mahout and CUDA can be used for implementation of Ratio-Based Collaborative Filtering to reduce the time complexity, as these platform have ability of parallel computation.

CONCLUSION

Recommendation System is used to predict the items for the user depending on some measure. From the survey it has been found out that Item-Based Collaborative Filtering provides better results in recommendations than User-Based Collaborative Filtering. In Item-Based Collaborative filtering, adjusted cosine similarity has better accuracy than other item based similarity. Also the survey shows that Ratio-Based Collaborative Filtering provides better accuracy than Adjusted Cosine Similarity. But the problem with all Collaborative Filtering algorithms is that, they have more execution time. So in order to achieve higher accuracy and less execution time, Recommendation System can be built using Parallel Ratio-Based Collaborative Filtering.

REFERENCES

[1] Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." *Advances in artificial intelligence* 2009 (2009): 4.

[2] Walunj, Sachin Gulabrao, and Kishor Sadafale. "An online recommendation system for e-commerce based on apache mahout framework." *Proceedings of the 2013 annual conference on Computers and people research*. ACM, 2013.

[3] Chen, DanEr. "The collaborative filtering recommendation algorithm based on BP neural networks." *Intelligent Ubiquitous Computing and Education, 2009 International Symposium on*. IEEE, 2009.

[4] Patil, Vandana, and Lata Ragha. "Comparing performance of collaborative filtering algorithms." *Communication, Information & Computing Technology (ICCICT), 2012 International Conference on*. IEEE, 2012.

[5] Liu, Qiang, Bingfei Cheng, and Congfu Xu. "Collaborative Filtering Based on Star Users." *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*. IEEE, 2011.

[6] Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." *Proceedings of the 10th international conference on World Wide Web*. ACM, 2001.

[7] Liu, Yaqiu, Zhendi Wang, and Man Li. "Ratio-based collaborative filtering algorithms." *Systems and Control in Aerospace and Astronautics, 2008. ISSCAA 2008. 2nd International Symposium on*. IEEE, 2008.

[8] Ma, M. M., and S. P. Wang. "Research of User-Based Collaborative Filtering Recommendation Algorithm Based on Hadoop." *International Conference on Computer Information Systems and Industrial Applications*. Atlantis Press, 2015.

[9] Zhao, Zhi-Dan, and Ming-Sheng Shang. "User-based collaborative-filtering recommendation algorithms on hadoop." *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on*. IEEE, 2010.

[10] Bamnote, G. R., and S. S. Agrawal. "Evaluating and Implementing Collaborative Filtering Systems Using Apache Mahout." *Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on*. IEEE, 2015.

[11] Kumar, Thangavel Senthil, and Swati Pandey. "Customization of Recommendation System Using Collaborative Filtering Algorithm on Cloud Using Mahout." *Intelligent Distributed Computing*. Springer International Publishing, 2015. 1-10.

[12] Jiang, Jing, et al. "Scaling-up item-based collaborative filtering recommendation algorithm based on hadoop." *Services (SERVICES), 2011 IEEE World Congress on*. IEEE, 2011.

[13] https://en.wikipedia.org/wiki/Recommender_system

[14] Goldberg, David, et al. "Using collaborative filtering to weave an information tapestry." Communications of the ACM 35.12 (1992): 61-70.

[15] Miller, Bradley N., et al. "MovieLens unplugged: experiences with an occasionally connected recommender system." Proceedings of the 8th international conference on Intelligent user interfaces. ACM, 2003.

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

[17] https://en.wikipedia.org/wiki/CUDA