Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 4, April 2020, pg.83 – 90

A MOVIE RECOMENDER SYSTEM BY COMBING BOTH CONTENT BASED AND COLLABORATIVE FILTERING ALGORITHMS

MUHAMMAD KAMRAN[1]; SYED SALMAN ALI SHAH[2]; MIRZA NAVEED JAHANGEER BAIG[3]; RIAZ HUSSAIN KHAN[4]

[1]Department of Mathematics & Computer Science, Virtual University of Pakistan, Pakistan
[2]Department of Computer Science, Wuhan University, China
[3]Department of Mathematics & Computer science Virtual university of Pakistan, Pakistan
[4]Department of Mathematics and Statistics, Institute of Southern Punjab, Multan, Pakistan
[1] kamranbloch512@gmail.com; [2] salmanali0313@gmail.com; [3] mirzanaveedbaigms@gmail.com; [4] riazsargani3@gmail.com

Abstract— The computerized world we are living in has a ton of information and data that is utilized by an assortment of clients, for example, videos, books and articles. Different users like different content. Getting what each user likes can be irritating. Each online services provider always aims in having many clients. Recommender systems importance's arises in such situations. The recommender framework proposes to a client some substance for example e.g. movies and books depending on what the user likes. In this research, a new movie recommender system is proposed, that will be able to improve the existing recommender systems. With this new recommender framework, the client will receive an improved forecast contrasted with different frameworks that as of now exist for example content-based filtering and collaborative-filtering. In order to overcome the disadvantages of the both methods (collaborative filtering algorithm (unsupervised learning) and content-based filtering algorithm (supervised learning)), the new system combines both methods. This will bring up a more stable system compared to the existing ones.

Keywords: movies, content-based filtering, collaborative-filtering, combined-algorithm, supervised-learning, unsupervised-learning, Pearson correlation.

I. INTRODUCTION

In today's world, many movies are released each month. This quantity is so much that a person may not be able to keep track of the movies he likes. Users gets confused because a lot of movies or other items exist on internet. To illustrate this, a user may wish to watch a movie, there are thousands of movies online and not all of them might be watched by a given user, based on preferences. This selection problem brings a need to have a system that suggests a movie or any other item to a user automatically as this will save user time.

Recommendation systems are data filtering systems that try to estimate what different users would prefer using the preference of the users, Neysiani et al. (2019). Recommendation system have offer a lot of help in recommending some content to users. Over the past few years, many recommendation systems have arisen that recommends different content to different people. Such content includes movies, books and music. The search engines and websites that sell products also use the recommendation systems to recommends products to their users. More often, the recommendations systems gather the user's preferences data and later uses the data to suggest to the user.

The recommendations systems can also be used to give improvements, according to the behaviour of most people, Asha & Rajkumar (2019, September). An example can be in Alibaba. If an observation can be made in Alibaba that a large number of people who buyers Samsung phone also buys a Samsung smartwatch, a recommendation to buy a Samsung smartwatch will be given to a user who buys a Samsung smartphone. With the many applications in the present days therefore, it is not possible to live without the recommendation systems. There are different types of recommending systems that exist. Some combine different techniques to be more accurate. In the recommendation systems, accuracy is the main objective when suggesting movies or any other item to a user. For that reason, this paper will combine two most popular techniques in order to give a user more accurate prediction.

Many recommendation systems are not able to suggest appropriately to a user, Vartak et al. (2017). This brings up the need to develop a more suitable recommending system that will overcome the shortcomings that are experienced in the existing systems. Different users have different preferences, and this is one of the problems the recommendation systems need to solve. The proposed recommendation system in this paper will filter the information about a movie and predicts users the best movies. The system will utilize two algorithms, i.e., the collaborative filtering and content-based algorithms. The content filter will suggest movies to users based on the information collected about a user. The collaborative filtering method will be used to put users into groups according to their similarities. It will make a recommendation to a user based on the information about the group

The paper contains five themed parts which are literature review, research problem, methodology, testing and conclusion. The literature review section explore the earlier researches which are related to movie recommendation. Research problem is a section that will show a gap that exist which will be addressed in this paper. Methodology is the most crucial part of this paper. It will show how the two filtering algorithm will be merged to come up with a more accurate algorithm. The testing part is a section that will compare the results of the existing standalone recommender system with the proposed algorithm. Finally, the conclusion part which concludes the whole paper and gives how this may be improve d in the future.

II. LITERATURE REVIEW

Recommendation systems

Many movie recommendation systems have been created with the aim of improving movie suggestions to users. One of the systems is the MOVREC, Kumar, et al. (2015). This movie recommendation system gives recommendations to users using following the collaborative filtering algorithm. The system uses information that is provided by a user. The system provides a recommendation to a user by analyzing the user information. The movie with the highest rating is recommended to the user, followed by others with lower ratings. In this system, a user can also select the attributes which he wishes the systems to use to give a recommendation, Kumar et al. (2015). This system utilizing the collaborative approach only is faced with problems such as early rater problems; therefore, a gap exists.

Another research that is, Kuzelewska (2014), that proposed a recommendation system that would help users to get interesting products. This system used data from other users i.e., ratings of movies. The system used clustering to get similar preferences of different users, Kuzelewska, (2014). The implementation of the system was in Apache Mahout Environment. This recommendation system was tested on a movie database. The selected similarity measures were depending on Euclidean distance, loglikelihood, and the correlation coefficient, Kuzelewska (2014). Likewise, the system experienced shortcomings that can be solved by our new proposed system.

Another system is the MovieLens, Harper & Konstan, (2015). This recommendation system uses collaborative filtering to recommend movies. Many recommending systems use the single similarity scheme; however, this system uses both the item and the user similarities. This system makes use of matrix factorization. When a new user signs up in the application, the user is supposed to give ratings to a given number of movies; the movies are of a different type. This helps the system to learn user preferences, thus avoiding the cold start problem, Harper & Konstan, (2015). This system has demerits since the reliability of the rating given during the sign-up process is questionable. This problem will be solved in the proposed system.

Costin-Gabriel et al. (2015) did a research and come up with a movie recommender, a framework which utilizes the data known about the client to give film suggestions. This framework try to take care of the issue of interesting suggestions which comes about because of disregarding the information distinct to the client. The psychological profile of the client, their viewing history and the information including film scores from different sites is gathered. They are based on total closeness calculations. The framework is a half content based and half collaborative filtering.

Additionally, a research by Ponnam et al. (2016, February), propose to develop an item based collaborative filtering algorithm. The systems developed in this paper used the ratings given to same items to give recommendation. This system aimed at replacing the content base filtering algorithm which has many disadvantages, compared to the collaborative filtering algorithm. This paper using only the collaborative filtering technique, was not able to more accurate. This paper will address this by using both content based and collaborative filtering algorithm.

The proposed system will combine both the content-based algorithm and the collaborative filtering algorithm. By so doing, this system will have the advantages of both the content-based and collaborative algorithm. The system will also be able to overcome the demerits of each algorithm hence bringing up a more stable and reliable system.

Content-based filtering (background information)

This is one of the basic methodology of predicting items or substance to the client. The thought here is that if a client demonstrates he loves an item by clicking, or by giving high appraising, or via searching or perusing it implies he has the high chance that they would he likes the item. Content-based filtering, additionally alluded to as cognitive filtering, suggests things dependent on an evaluation between the substance of the things and a client profile. A couple of issues must be seen while executing a content based filtering framework. In the first place, terms can either be consigned automatically or physically .when terms are appointed out automatically a procedure must be picked that can pick these terms from things. Second, the terms must be addressed with the ultimate objective that both the customer profile and the things can be pondered in a significant way. Third, a learning calculation, must be picked that can get acquainted with the customer profile reliant on watched things and can make recommendations subject to this customer profile. Independent content based isn't common yet, it is used with various sorts of recommendations structures to shape join algorithms, to frame progressively precise and stable systems.

In movies recommendation, Content-based filtering is a supervised learning algorithm, Roy et al. (2019, October). This method of filtering gives recommendations to a movie by comparing the movie profile and the customer profile. It recommends movies to users depending on what the users like.

An example of a movie (m) genre can be said to be:

m= (romance, action, scifi)

Supervised learning can be applied to learn the genre of a movie from the marketing sites. A genre of an action movie, for example, can be defined as

w= (0, 1, 0)

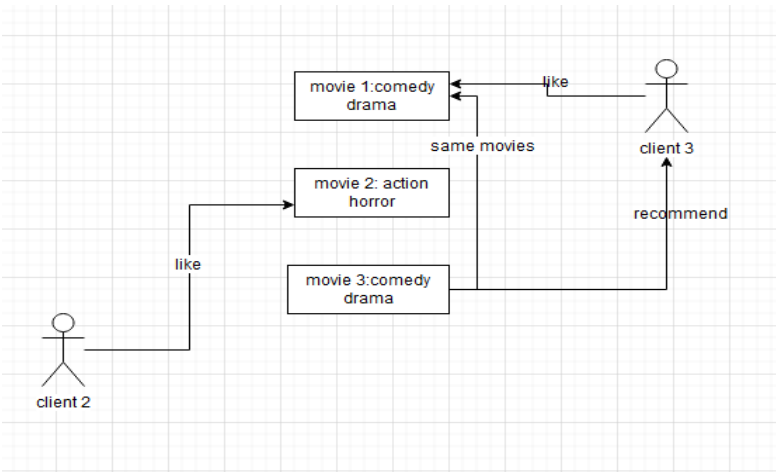

Fig 1: content-based filtering

The algorithm also learns how a user can like a given movie based on the movie's user likes. The preference m1 of a given user who likes the genres action and romance but does not like scifi can be defined as,

m1= (0.8, 08, 0),

based on how much the user likes the genres. The similarity between the two movies can be calculated using the cosine similarity. The cosine similarity is used because it is fast and does not depend on the magnitude, Luo et al. (2018, October). The similarity of vector v and u, is the ratio of the dot product and product of their magnitudes.

$$similarity = cos(\theta) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}}$$

Similarity.

This will get the movies that are similar to a given movie. This method of recommendation has the following advantages:

- Unrated items can be rated with this method
- This method does not need information of other users to give prediction.

The method has the following demerits:

- Does not work for a new user since it requires users to have rated some movies to get their preferences
- Cannot recommend unexpected movies

Collaborative filtering (background information)

Collaborative filtering is procedure behind many predictor systems. suggestions outline works study information about customers with similar tastes to assess the probability that a target individual will value something, for instance, a video, a book or any other item. Collaborative filtering utilizes calculations to channel information from client ratings to make customized suggestions for clients with comparative inclinations. Collaborative filtering is likewise used to choose substance and publicizing for people via web-based networking media. Three sorts of Collaborative filtering generally utilized in prediction frameworks are neighbor-based, item to-item and classification-based. In neighbor-based method, clients are chosen for their closeness to the client who is active. This closeness is controlled by coordinating clients who have posted comparative ratings. In view of the past likeness, it is assumed that future preferences will likewise be comparable. From the mean rating of the neighbors, suggestions are made for the client who is active. The item to-item filtering system utilizes a matrix to decide the similarity of sets of things.

The classification based collaborative filtering framework suggests things dependent on how comparable clients loved that characterization or type. It is expected that clients that appreciate or hate comparative encounters inside a characterization will likewise appreciate others inside that grouping. Some collaborative filtering frameworks are memory-based, such as neighboring-and item to-item models, which think about similitudes of clients or things. Others are model-based, utilizing AI to think about divergent things. Combined frameworks incorporate highlights of both memory-based and model-based sifting.

Recommendation frameworks are utilized to give suggestions to a wide services. In any case, they can experience various troubles. The sparsity is one of the fundamental obstacles to collaborative filtering especially in frameworks with numerous things. New things likewise will in general be hard to give suggestions to. In new recommendation frameworks, it is difficult to give great recommendations before enough clients have entered ratings.

In movie recommendations, collaborative gives movie recommendations based on the ratings of other people. It is an unsupervised learning algorithm. This algorithm gives recommendations clients depending on similarities between clients or the films, Sadanand et al. (2018). This similarity is computed by the Pearson correlation coefficient as shown below.

$$sim(u,v) = \frac{\sum_{i \in com(u,v)} \left(r_{u,i} - \bar{r_u}\right)\left(r_{v,i} - \bar{r_v}\right)}{\sqrt{\sum_{i \in com(u,v)} \left(r_{u,i} - \bar{r_u}\right)^2} \sqrt{\sum_{i \in com(u,v)} \left(r_{v,i} - \bar{r_v}\right)^2}}$$

ru, i is the rating of the ith item of client u,
ru is the mean rating of client u,
$com\ (u,)$ is the items rated by both client u and client v

An example is that one user can ask another user to recommend movies if they have similar taste as illustrated in figure 2. According to Ponnam et al. (2016, February), on their paper about creating a movie system using item based collaborative algorithm, they indicated that collaborative techniques are more accurate compared to content based techniques. The paper also indicated that rather than the low accuracy, content based filtering method are faced with more challenges big error prediction and scalability preferring collaborative filtering to content based filtering.

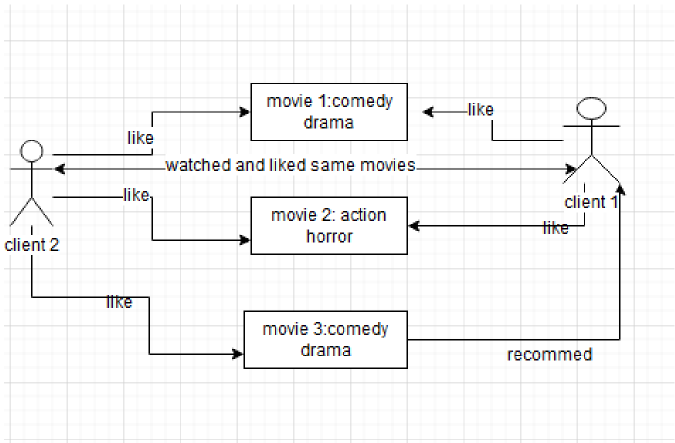

Fig 1: collaborative filtering algorithm

This method also has some demerits, which are:

- Cannot recommend fresh movies
- The algorithm encounters sparsity problem, and this occurs when the response data is too sparse to get the neighbors of a given movie.

Research Problem

Although deep research has been done earlier concerning the recommendation systems, problem exist with the stand-alone content-based filtering and the collaborative filtering algorithm. The main problem of this standalone systems is cold start. Cold start is a problem that exist with new users, i.e. they cannot recommend any information to a new user. Due to such disadvantages of standalone content-based filtering and collaborative filtering, a problem exists which will be addressed in this paper by combining both recommended methods. Although collaborative and content-based filtering algorithms can be implemented in many different ways, the problem still persist if they are not combining. This research will combine both methods in order to get the benefits of both filtering algorithms in the proposed system. This will result in the building of a hybrid system that is more stable than the existing ones. The resultant algorithm will be tested and compared with standalone algorithms to prove its efficiency.

III. METHODOLOGY

The proposed systems combine the two discussed filtering techniques i.e. the collaborative and the content-based. This new algorithm will utilize the user ratings. The user-rating rating matrix will be filled by the content-based filtering since it is sparse. There will be a need to download movie content. This will be done using the web crawler and stored. The downloaded content will be the user rating matrix. After this, the content-based will be used in training individual user rating vectors, using the downloaded content. The user rating vector will be transformed into an artificial rating matrix. The artificial rating matrix will be composed of the real and the output rating matrix after training. This new artificial user rating matrix will be filtered using the collaborative algorithm and recommendation given to a user. The figure below shows how the new algorithm will work.

The new algorithm explained in steps:

1. By using the content-based, the algorithm calculates the artificial user rating vector v, for user u in the datastore.

$v_{u,i} = r_{u,i}$: is user u rated item i

$v_{u,i} = r_{u,i}$: otherwise

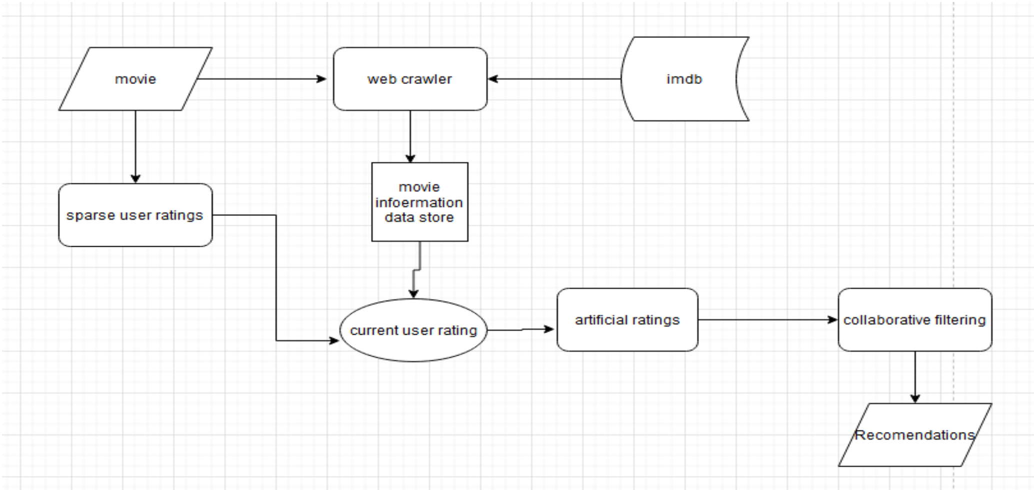

Fig 2: the information flow of the new algorithm

2. The algorithm weights all the users based on similarity with the user who is active. This is calculated using the Pearson correlation, Zhou et al. (2016), of their rating vectors, as shown below.

$$P_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a)^2 \times \sum_{i=1}^{m}(r_{u,i} - \bar{r}_u)^2}}$$

In this formula, ra, i is appraising given to item I by client a.
$\bar{ra}$ is the average rating given to item I by client a
m is the quantity of items.

The algorithm selects a given number of users with the closest similarity with the user who is active. These are the neighbors of the active users

The algorithm then computes the prediction. This is done by using the formulae below.
Prediction is calculated as follows:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^{n}(r_{u,i} - \bar{r}_u) \times P_{a,u}}{\sum_{u=1}^{n} P_{a,u}}$$

pa, is the expectation for the active client for item I,
a, is the similitude between clients a and u;
n is the quantity of clients in the area

IV. Testing and Results

So as to know whether our new framework works better than the individual collaborative and content based, I evaluated the two strategies and the integrated framework utilizing a similar set of information. The content and collaborative frameworks were evaluated first. The users who have rated at least 50 movies were used and result presented in a bar graph. The results of each framework were recorded for the purpose of comparison. Collaborative filtering system was 0.69 accurate. Collaborative filtering was computed by the Pearson correlation the content-based frame work was also evaluated. This algorithm produced a precision value of 0.68 accuracy and a 0.42 recall value. This result for each method alone was quite good, but combining the two methods was better.

The combined algorithm produced an accuracy value of 0.8. Comparing this value to that of the individual algorithms, it means that this combined algorithm increased accuracy by 16%. This new value indicates that the new framework can defeat the inadequacy that every individual framework has, in this line giving the film suggestions more precisely. The interrogated calculation additionally improved the time it takes to predict a film. The independent recommenders generally take more time while suggesting videos. The groundwork results have shown that the running time of the new framework is better. The figure below shows the results of the precision of every framework.

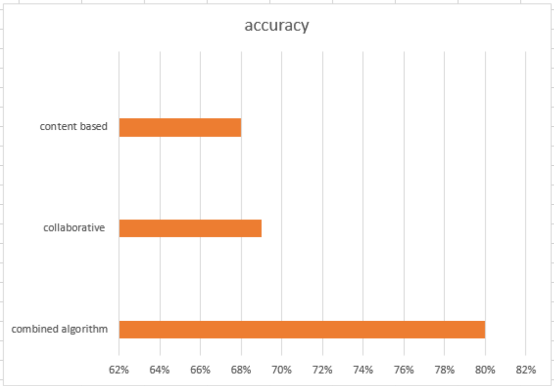

Fig 3: Results of the 3 methods represented

V. Conclusions

In conclusion, there are many existing recommender system which uses one algorithm while recommending thus are faced with disadvantages. To avoid such shortcoming, this paper has researched and developed a new movie recommender algorithm that combine the two most common algorithm in recommender systems. The collaborative algorithm and the content based algorithm are the most common algorithms in this field. The resultant framework appreciates the benefits and keep away from the bad marks of each independent calculation. The resultant algorithm also give predictions to a user about a movie efficiently compared to the stand alone algorithm and this is evidenced by the experiment that was done. The two frameworks which are the collaborative filtering that places clients into bunches as per their similitude and content-based framework predicting films to clients dependent on the data gathered about a client are interrogated. The consequences of the joined framework are vastly improved contrasted with those of a different technique. The new framework had a precision of 80% instead of that of 69 % of the different frameworks. This new framework has addressed problems in each separate framework.

Even if the new algorithm has addressed the problem with the stand alone recommender system, there is still room for improvement. In the future, this system can be improved further by combining clustering techniques and similarities which is will make the recommendations even more accurate. Combining this two will reduce the mean absolute error. In the future, the system could also be made in such a way that it will recommend other items, not just movies e.g., songs and books. This algorithm can also be improved further so that it reduces even the time it takes to recommend a movie.

References

[1] Asha, K. N., and R. Rajkumar. "A Comprehensive Survey on Web Recommendations Systems with Special Focus on Filtering Techniques and Usage of Machine Learning." In *International Conference On Computational Vision and Bio Inspired Computing*, pp. 1009-1022. Springer, Cham, 2019.

[2] Chiru, Costin-Gabriel, Cătălina Preda, Vladimir-Nicolae Dinu, and Matei Macri. "Movie Recommender system using the user's psychological profile." In *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 93-99. IEEE, 2015.

[3] Harper, F. Maxwell, and Joseph A. Konstan. "The movielens datasets: History and context." *Acm transactions on interactive intelligent systems (tiis)* 5, no. 4 (2015): 1-19.

[4] Kumar, Manoj, D. K. Yadav, Ankur Singh, and Vijay Kr Gupta. "A movie recommender system: Movrec." *International Journal of Computer Applications* 124, no. 3 (2015).

[5] Kuzelewska, Urszula. "Clustering algorithms in hybrid recommender system on movielens data." *Studies in logic, grammar and rhetoric* 37, no. 1 (2014): 125-139.

[6] Luo, Chunjie, Jianfeng Zhan, Xiaohe Xue, Lei Wang, Rui Ren, and Qiang Yang. "Cosine normalization: Using cosine similarity instead of dot product in neural networks." In *International Conference on Artificial Neural Networks*, pp. 382-391. Springer, Cham, 2018.

[7] Neysiani, Behzad Soleimani, Nasim Soltani, Reza Mofidi, and Mohammad Hossein Nadimi-Shahraki. "Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm." *Int. J. Inf Technol. Comput. Sci.* 2 (2019): 48-55.

[8] Ponnam, Lakshmi Tharun, Sreenivasa Deepak Punyasamudram, Siva Nagaraju Nallagulla, and Srikanth Yellamati. "Movie recommender system using item based collaborative filtering technique." In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pp. 1-5. IEEE, 2016.

[9] Roy, Sushmita, Mahendra Sharma, and Santosh Kumar Singh. "Movie Recommendation System Using Semi-Supervised Learning." In *2019 Global Conference for Advancement in Technology (GCAT)*, pp. 1-5. IEEE, 2019.

[10] Sadanand, Howal, Desai Vrushali, Nerlekar Rohan, Mote Avadhut, Vanjari Rushikesh, and Rananaware Harshada. "Movie recommender engine using collaborative filtering." In *Smart Computing and Informatics*, pp. 599-608. Springer, Singapore, 2018.

[11] Vartak, Manasi, Silu Huang, Tarique Siddiqui, Samuel Madden, and Aditya Parameswaran. "Towards visualization recommendation systems." *ACM SIGMOD Record* 45, no. 4 (2017): 34-39.

[12] Zhou, Haomiao, Zhihong Deng, Yuanqing Xia, and Mengyin Fu. "A new sampling method in particle filter based on Pearson correlation coefficient." *Neurocomputing* 216 (2016): 208-215.

