

# Memory-based Collaborative Filtering: Impacting of Common Items on the Quality of Recommendation

Hael Al-bashiri<sup>1</sup>, Hasan Kahtan<sup>2</sup>  
Awanis Romli<sup>4</sup>  
Faculty of Computing, Univirsity  
Malaysia Pahang  
Gambang, Kuantan, Malaysia

Mansoor Abdullateef Abdulgaber<sup>3</sup>  
Faculty of Computer and Cyber  
Sciences  
University of Prince Mugrin, Madinah  
Saudi Arabia

Mohammad Adam Ibrahim  
Fakhreldin<sup>5</sup>  
Faculty of Computer Science and  
Information Systems, Jazan University,  
Jazan, Saudi Arabia

**Abstract**—In this study, the impact of the common items between a pair of users on the accuracy of memory-based collaborative filtering (CF) is investigated. Although CF systems are a widely used recommender system, data sparsity remains an issue. As a result, the similarity weight between a pair of users with few ratings is almost a fake relationship. In this work, the similarity weight of the traditional similarity methods is determined using exponential functions with various thresholds. These thresholds are used to specify the size of the common items amongst the users. Exponential functions can devalue the similarity weight between a pair of users who has few common items and increase the similarity weight for users who have sufficient co-rated items. Therefore, the pair of users with sufficient co-rated items obtains a stronger relationship than those with few common items. Thus, the significance of this paper is to succinctly test the impacting of common items on the quality of recommendation that creates an understanding for the researchers by discussing the findings presented in this study. The MovieLens datasets are used as benchmark datasets to measure the effect of the ratio of common items on the accuracy. The result verifies the considerable impact exerted by the factor of common items.

**Keywords**—Collaborative filtering; memory-based; similarity method; data sparsity

## I. INTRODUCTION

Currently, Internet users have faced information overload issues. Therefore, these users resort to traditional recommendation methods to make their decisions, such as asking friends, scanning newspapers and following advertisements. These basic solutions might help alleviate information overload [1]. However, the amount of information on the Internet increases tremendously every day, thus complicating the decision-making process of Internet users. To assist the users in dealing with information overload, researchers have developed a recommender system (RS) that can provide a list of preferable items among the huge amount of items available by predicting users' preferred items [2-7]. State-of-the-art RSs can be grouped into content-based, collaborative filtering (CF) or combined (i.e. hybrid) approaches [4, 8-11].

CF, which is one of the most successful recommendation methods, uses the feedback provided by the users to generate recommendations [2, 12-15]. CF can be classified as model- or memory-based [1, 16]. In the former, a part of the dataset is used to build a model that can predict the preferred items. By contrast, the latter does not require a model. Instead, user feedback (e.g. ratings) is used to compute the similarity amongst the users directly. The computation of the similarity can be conducted in the space of users (user-based) or items (item-based) [1, 8, 17]. The idea behind CF is that the users will have similar favourites in the future if they have shared similar preferences in the past [18, 19]. The key step in CF is finding the right neighbours. Therefore, selecting the appropriate similarity is fundamental to the system's performance. Several methods have been introduced in CF systems, including Pearson's correlation coefficient (PCC), cosine similarity and other derivative methods [20, 21]. However, the key challenge faced in a CF system is providing high-quality recommendations to users who do not have enough information preferences. Most users do not evaluate a sufficient number of items in the database, thereby making the user-item rating matrix sparse [22-27]. As a result, calculating the similarity amongst these users may lead to locating unsuccessful neighbours and consequently low performance.

Therefore, the primary goal of this study is to investigate the impact of co-rated item size based on sigmoid function on enhancing the accuracy of the recommendation. This evaluation will rely on the various sizes of common items, which will be represented by threshold values, to measure the impact on solving the issue of data sparsity. PCC and cosine similarity will be used and modified by adopting a sigmoid function under several thresholds. The correlation weight between the pair of users should be devalued if the number of the co-rated items is smaller than the threshold. The experiments will be conducted using MovieLens benchmark datasets.

The remainder of this paper is organized as follows. Section 2 presents the literature review. Then, the developed similarity method and its phases are discussed in Section 3. In Section 4, the evaluation process and the experimental results are presented. The conclusions are provided in Section 5.

## II. RELATED WORK

Given the great impact of the quality of similarity measure on the accuracy of recommendation, several similarity measures have been developed based on the ratio of common items. In this section, certain methods that consider the proportion of co-rated items in improving the accuracy of recommendation will be presented, as well as how these factors influence the quality of recommendation.

Resnick, et al. [2] applied PCC similarity measures to compute the linear correlation between two objects. The outcome was between 1 and -1, where 1 denoted the total positive correlation, 0 meant no correlation and -1 denoted complete negative correlation. However, the PCC similarity method exerted a remarkable influence on the sparsity of data. Therefore, the correlation calculation amongst users who had few co-rated items was difficult and might lead to high/low similarity and weak recommendations. To scale the similarity properly when the number of common items was not enough, [28] introduced a significant factor  $(\min[I_{x,y}, \gamma]/\gamma)$  to devalue the similarity weights. They applied these weights for the users who had fewer common co-rated items than the threshold. In addition, they used threshold  $\gamma$  to determine the minimum number of co-rated items. In their experiments, when  $\gamma \geq 25$ , the accuracy of the predicted ratings improved; the optimal result was when  $\gamma = 50$ . However, the proportion of the common ratings was not considered. Thus, the sparsity issue still influenced the determination of neighbours who had common items that were bigger than the threshold.

In [24], the authors introduced a heuristic measure called Proximity-Impact-Popularity (PIP), which consisted three essential factors that played important roles in identifying the relationship amongst users. PIP utilised the ratings given by a set of users to enhance the accuracy of RS under cold-star conditions. However, PIP ignored the proportion of the common ratings and performed excessive similarity computations [29]. Author in [30] proposed a weighted item-based similarity measure based on the sigmoid function. This function devalued the similarity weight if few co-ratings were available. If the size of the set of common items was sufficiently large, the weight value was 1. Otherwise, the weight value was 0.6. Another improved similarity measure was introduced by adding the similarity impact factor  $\varepsilon$  to the traditional similarity measure to alleviate the effect of data sparsity [31]. This factor represented the proportion of the common items rated by a pair of users. If the pair of users did not have co-rated items, then the in-between similarity was 0, which was considered a weakness in this method. Author in [32] formulated a weight distance model to compute the association between two users based on the ratio of the common ratings and the relationship between the target and the co-rated items. The proportion of the common ratings was computed using the Jaccard index, whereas the relationship between the target and the co-rated items were computed using PCC. A singularity-based similarity measure was introduced in [33]. This method hypothesised that a stronger relationship existed between two users if they rated the items that had been rated by few users than if they rated the items that had been rated by numerous users. The singularity values of each user replaced the similarity value to improve the PCC

method. The Jaccard measure was then modified based on singularity to consider the proportion of the common items.

A new weight similarity model called NWSM was proposed by Zang, et al. [34]. This model took into account the proportion of the common rating, user rating preference and the different contributions of other users to the target. To improve the accuracy of the recommendation, the final similarity formula was obtained by integrating three factors: (1) the PCC method with influence weight (i.e. neighbourhood's rating information), (2) the Jaccard measure to compute the proportion of co-ratings and (3) the mean variance of the rating to calculate the differences in the preference of each user. Zhang, et al. [35] presented a new effective CF method to decrease the impact of data sparsity based on user preference clustering. The users were first grouped into clusters according to preferences before the neighbour for the active user was selected from these clusters. Zhang and Yuan [26] presented an improved similarity method to overcome the problem of data sparsity by analysing the shortcomings of a traditional memory-based CF similarity method. In the enhanced similarity method, the relationship between the users' common rating items and all items rated by the target user was considered. Author in [29] proposed a new linear combination similarity method called weight-based modified heuristic similarity measure to solve the problem of data sparsity.

The above-mentioned studies highlighted the significance of the ratio of common item in the process of developing a similarity measure that can improve the recommendation accuracy. However, studies on the determination of the optimal size of co-rated items amongst users are lacking. Therefore, the sigmoid function will be used in this study to test the impact of the item size on the accuracy of recommendation using co-rated items with various sizes.

## III. METHODOLOGY

This section details the step in obtaining the ratio of the common items. Firstly, the Bray-Curtis (BC) distance measurement is used to compute the distance between two different sites based on the counts at each site [36]. Normalisation is performed using absolute difference divided by the summation. The output value of the BC ranges from 0 and 1, where a value of 0 indicates a complete matching of the two data records in the n-dimensional space, and 1 means that the records are different. The BC distance performed better than the 10 and 9 distance measures used in [37, 38], respectively. The general formula for the BC distance can be expressed as

$$Bc(u, v) = 1 / (1 + \frac{\sum_{g=1}^k |w_{u,g} - w_{v,g}|}{\sum_{g=1}^k w_{u,g} + \sum_{g=1}^k w_{v,g}}) \quad (1)$$

where  $k$  represents the number of item categories in the database, and the  $w_{u,g}$  and  $w_{v,g}$  represent the ratio rating of type  $g$  for users  $u$  and  $v$ , respectively. The computation of ratio rating values are provided in the previous work [21].

Then, the BC similarity is multiplied by the sigmoid function ( $Sf$ ) to devalue the similarity in cases of few co-ratings. The  $\theta$  in the equation below is used to determine the

minimum size of co-rated items. If the size of the set of common items is sufficiently large, then the sigmoid value will be higher than 0.9. Otherwise, the value will be less than 0.9. For example, if  $\theta = 1$  and the number of the common ratings of the pair of users is equal to 0, the sigmoid value will be 0.5. If the size of co-rated items is greater than 3, the sigmoid value will be greater than 0.95. The sigmoid function can be computed as

$$Sf(u, v) = Sf(v, u) = \frac{1}{1 + \text{Exp}\left(-\frac{|I_{u,v}|}{\theta}\right)}, \quad (2)$$

where  $|I_{u,v}|$  represents the number of items rated by users  $u$  and  $v$ . The Blazing Signature Filter (BSF) method can then be defined as

$$BSF(u, v) = \frac{1}{1 + \frac{\sum_{g=1}^k |w_{v,g} - w_{u,g}|}{\sum_{g=1}^k w_{v,g} + \sum_{g=1}^k w_{u,g}}} * sf(u, v). \quad (3)$$

Several denominator values of the sigmoid function are tested to determine the acceptable number of common items.

After the similarity is computed, the most similar users in the database will be located as neighbours. Then, the adjusted weighted method is used to compute the predictions score for the user  $u$  on each neighbours' item as can be defined in Eq. 4.

$$P_{u,i} = \bar{r}_u + \frac{\sum_{v \in k} \text{sim}^{BSF}(u,v)}{\sum_{v \in N} |\text{sim}(u,v)|} \quad (4)$$

Where  $P_{u,i}$  is the prediction value for  $u$  about a specific item  $i$ , and  $N$  is the nearest neighbour of user  $u$ .

In the next phase, M-top items will be provided to the target user as a set of recommendation.

Finally, the performance accuracy of the BSF similarity method is evaluated using MovieLens 100K dataset with holdout splitting methods using selected metrics (MAE, Recall, Precision, and F-measure). All these steps are presented in Fig. 1.

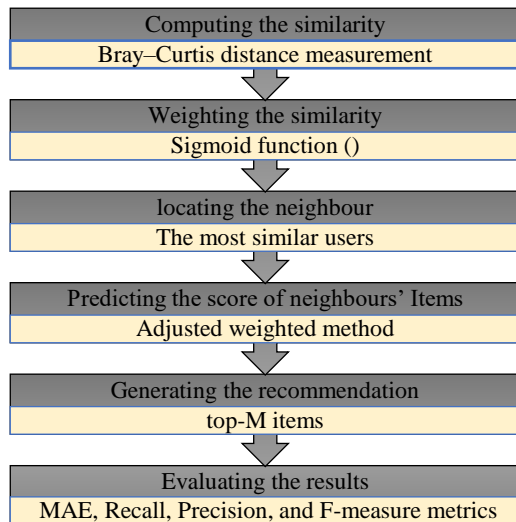


Fig. 1. Methodology.

#### IV. EXPERIMENT AND RESULT

To include the proportion of the common ratings in the calculation of the similarity between a pair of users, the necessary size of the co-rated items to increase or decrease the weight of similarity between the pair of users should be determined. The use of the sigmoid function depends on the denominator value. Therefore, several experiments have been conducted to determine the appropriate denominator value that can improve the similarity measure and produce acceptable results. Table I presents the description of various initial testing denominator values and their corresponding effect on the sigmoid value.

Fig. 2 illustrates the mean absolute error (MAE) rate for the CF-BSF using MovieLens 100K. The numbers of neighbourhoods were 30, 50, 70, 100 and 150. A slight improvement can be observed in the MAE value when the number of neighbours increases. Similarly, the MAE value increases when the denominator increases. In summary, the MAE values are acceptable when the size of the neighbours is 150 in all denominator values. Moreover, the lowest MAE rate is observed when the denominator is equal to 13. When the denominator is nine, the MAE rate is approximately near the optimal value.

TABLE I. DESCRIPTION OF VARIOUS DENOMINATOR VALUES

$\theta$	Description
5	The sigmoid value will be greater than 0.9 if the number of co-rated items is more than 10. Otherwise, the value will be less than 0.9. However, if the number of the common ratings of the pair of users is equal to 0, the sigmoid value is equal to 0.5.
7	The sigmoid value will be greater than 0.9 if the number of co-rated items is more than 15. Otherwise, the value will be less than 0.9.
9	The sigmoid value will be greater than 0.9 if the number of co-rated items is more than 20. Otherwise, the value will be less than 0.9.
11	The sigmoid value will be greater than 0.9 if the number of co-rated items is more than 25. Otherwise, the value will be less than 0.9.
13	The sigmoid value will be greater than 0.9 if the number of co-rated items is more than 30. Otherwise, the value will be less than 0.9.

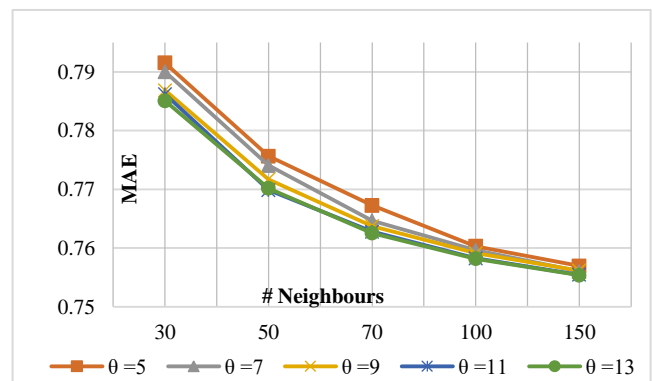


Fig. 2. MAE of CF-BSF vs.  $\theta$ .

Fig. 3 presents the comparison of the recall rates for CF-BSF using the initial denominator values. The subfigures A, B, C, D and E represent the respective recall rates for neighbour sizes 10, 20, 30, 40 and 50. The horizontal axis represents the sizes of the recommendations. Based on the

illustrations, the recall rates in most cases are the highest when the denominator value is 9. Similarly, the recall percentages in some cases demonstrate a good rate when the denominator is less than 9. The lowest recall values are observed when the denominator exceeds 9.

Fig. 4 shows the precision rate of CF-BSF under five different denominator values: 5, 7, 9, 11 and 13. The subfigures represent the respective precision rates of neighbour sizes 10, 20, 30, 40 and 50. When the denominator

value is 9, the rate of precision is the highest in most cases. The opposite is true for the denominator values equal to 11 and 13. Moreover, the precision rate does not exceed the precision rate of the denominator values less than 9. The highest precision rate is observed when the number of recommended items is small. The precision rate slightly decreases when the number of recommended items increases. In summary, the precision rate has exhibited variations, wherein the precision is higher than any other values when the denominator is equal to 9.

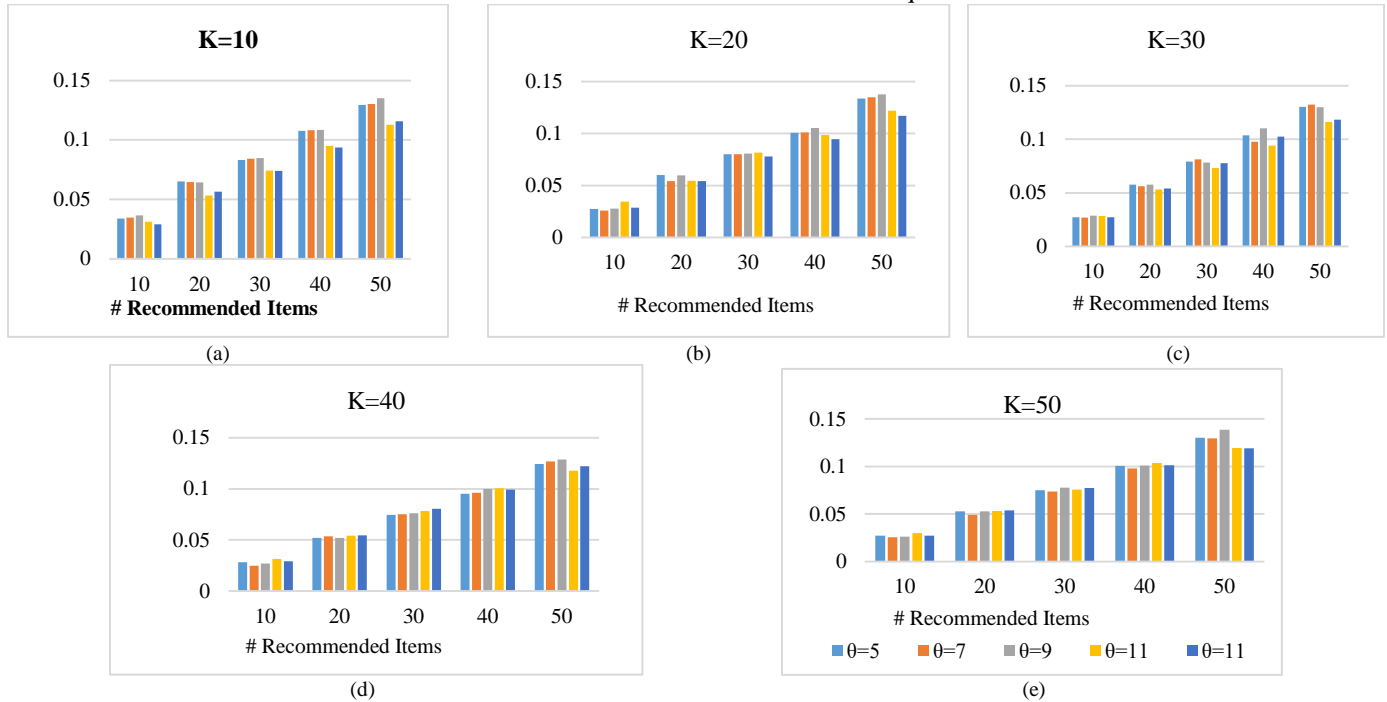


Fig. 3. Recall of CF-BSF vs.  $\theta$ .

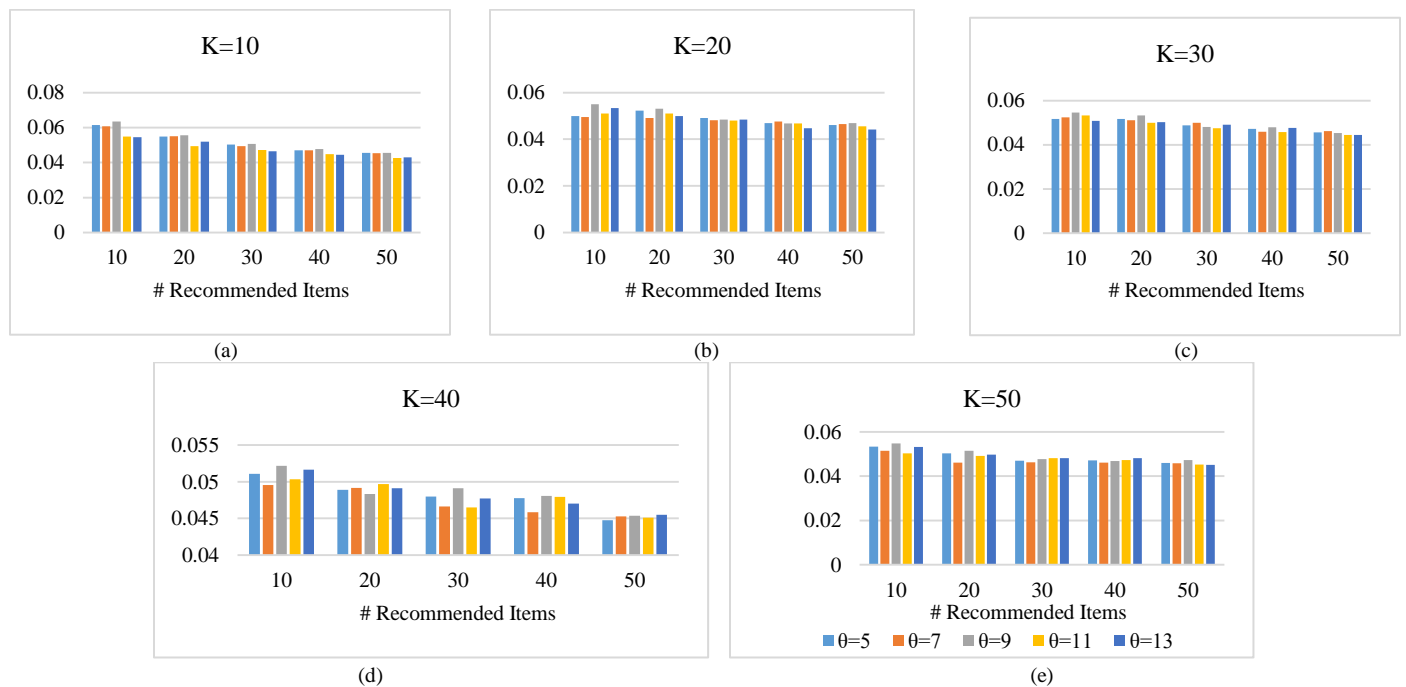


Fig. 4. Precision of CF-BSF vs.  $\theta$ .

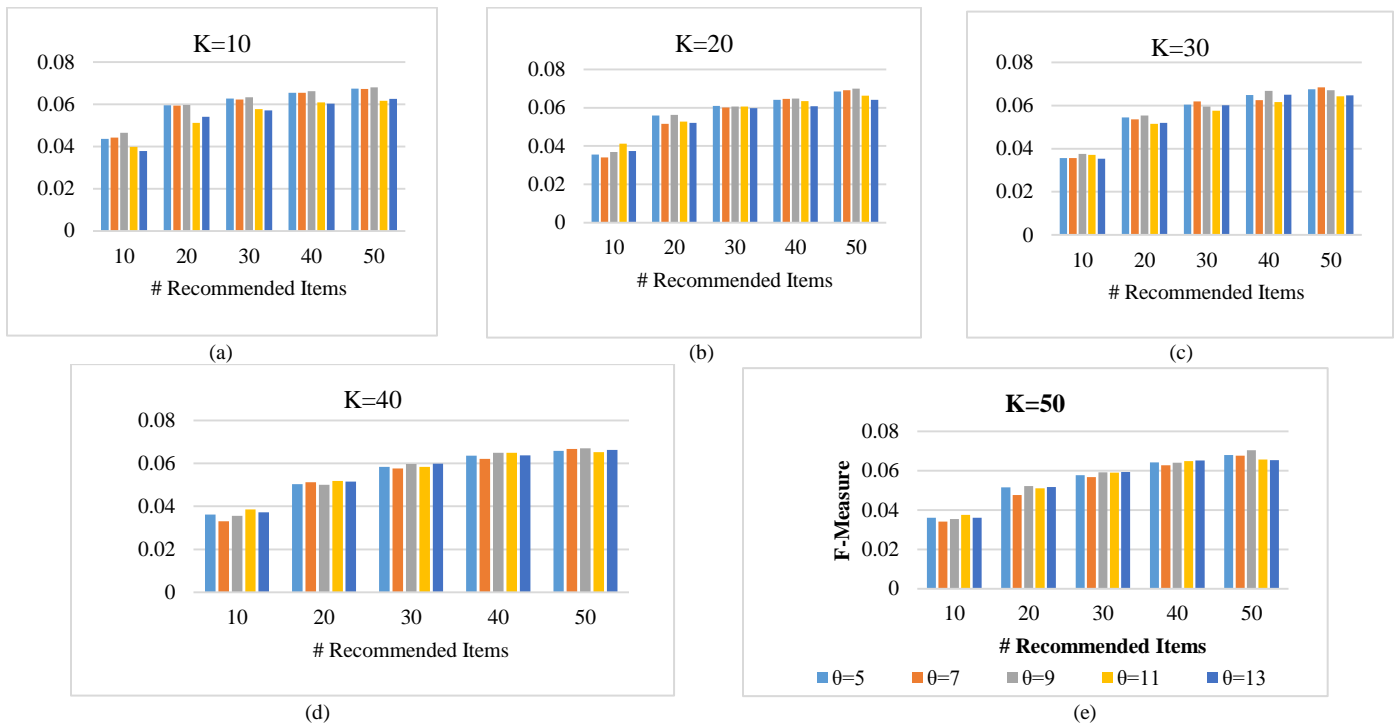


Fig. 5. F-Measure of CF-BSF vs.  $\theta$ .

Fig. 5 depicts the F-measure rate of CF-BSF under the five initial denominator values. The result shows that the highest rate is observed when the denominator value is equal to 9, and the lowest rate appears when the denominator is equal to 11 and 13. For denominator values less than 9, the F-measure rate has approximately the same rate as when the denominator value is 9. The highest percentage of F-measure is observed when the number of recommended items is 50. The F-measure rate was slightly enhanced when the number of recommended items increases. Similar to the results of the precision rates, the F-measure rates exhibit variation under the five denominator values. Nevertheless, the F-measure rate almost reaches the highest value when the denominator is equal to 9.

## V. DISCUSSION AND FINDING

Fig. 1 to 4 indicates that the denominator value has affected prediction and performance accuracies. In addition, the size of neighbours and the number of recommended items, as significant variables, have an impact on both metrics. As previously mentioned, the aim of these experiments is to determine the appropriate value of the denominator. This value will be used as a primary input for the sigmoid function to identify the right number of the common rating items amongst the users. If the number of co-rated items between a pair of users is insufficient, the similarity weight will be devalued using the sigmoid function.

Although the MAE rate is superior to the other rates for denominators higher than 9, the recall, precision and F-measure rates do not produce acceptable values. By contrast, the MAE rate is the lowest when the value of the denominator is less than 9; the recall, precision and F-measure percentages

are not higher than the percentages when the denominator is equal to 9. In conclusion, the most appropriate value for the denominator is 9. However, the result is dependent on the type and size of the dataset. Thus, using datasets with different sizes and types might produce different results.

## VI. CONCLUSION

In this work, the effect of the size of co-rated items among users on recommendation accuracy was tested using exponential function. Sparsity problems had a negative impact on the accuracy of memory-based CF. The similarity weight between a pair of users with a small number of ratings, regardless if the similarity was high or low, produced an unrealistic relationship amongst users compared with a real relationship with either a high or low similarity. Therefore, the exponential function using varying threshold values was used to determine the optimal number of the common items amongst the users to revalue the similarity weight of the traditional similarity methods. The exponential function could devalue the similarity weight between a pair of users who had few co-rated items. Consequently, the similarity weight between users who had sufficient co-rated items would be increased. Moreover, the MovieLens 100K was used as the benchmark dataset to quantify the impact of the ratio of co-rated items on memory-based CF accuracy. The results showed that the factor of common items exhibited a significant impact on the accuracy. Several evaluation metrics were utilised for the investigation, namely, MAE, recall, precision and F-measure. A similarity method can be developed in the future to improve the accuracy of memory-based CF by considering the co-rated item size and the different rating degrees of users and different datasets.

## ACKNOWLEDGMENT

This research is supported by the Department of Research and Innovation of University Malaysia Pahang under RDU190365 grant.

## REFERENCES

- [1] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, 2009.
- [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, Chapel Hill, North Carolina, USA, 1994, pp. 175-186.
- [3] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth",", in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Denver, Colorado, USA 1995, pp. 210-217.
- [4] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, pp. 56-58, 1997.
- [5] O. S. Revankar and Y. Haribhakta, "Survey On Collaborative Filtering Technique In Recommendation System," *IJAIEM*, vol. 4, pp. 85-91, 2015.
- [6] N. Salehudin, H. Kahtan, M. A. Abdulgaber, and H. Al-Bashiri, "A Proposed Course Recommender Model based on Collaborative Filtering for Course Registration," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 10, pp. 162-168, 2019.
- [7] H. Al-bashiri, M. A. Abdulgaber, A. Romli, and H. Kahtan, "An improved memory-based collaborative filtering method based on the TOPSIS technique," *PLoS one*, vol. 13, p. e0204434, 2018.
- [8] F. Ricci, L. Rokach, and B. Shapira, *Introduction to recommender systems handbook*: Springer, 2011.
- [9] L. Sharma and A. Gera, "A survey of recommendation system: Research challenges," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 4, pp. 1989-1992, 2013.
- [10] J. A. Konstan, J. Riedl, A. Borchers, and J. L. Herlocker, "Recommender systems: A grouplens perspective," in *Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08)*, 1998, pp. 60-64.
- [11] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, pp. 1-49, 2012.
- [12] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The adaptive web*, ed: Springer, 2007, pp. 291-324.
- [13] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative filtering recommender systems," *Foundations and Trends in Human-Computer Interaction*, vol. 4, pp. 81-173, 2011.
- [14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, Hong Kong, Hong Kong, 2001, pp. 285-295.
- [15] J. L. Herlocker, J. A. Konstan, K. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *Acm Transactions on Information Systems*, vol. 22, pp. 5-53, Jan 2004.
- [16] R. Zhang, Q.-d. Liu, and J.-X. Wei, "Collaborative Filtering for Recommender Systems," in *Advanced Cloud and Big Data (CBD)*, 2014 Second International Conference on, 2014, pp. 301-308.
- [17] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, pp. 734-749, 2005.
- [18] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender systems: an introduction*: Cambridge University Press, 2010.
- [19] H. Al-bashiri, M. A. Abdulgaber, A. Romli, and F. Hujainah, "Collaborative Filtering Similarity Measures: Revisiting," in *Proceedings of the International Conference on Advances in Image Processing*, 2017, pp. 195-200.
- [20] Q. Cheng, X. Wang, D. Yin, Y. Niu, X. Xiang, J. Yang, et al., "The new similarity measure based on user preference models for collaborative filtering," in *Information and Automation*, 2015 IEEE International Conference on, 2015, pp. 577-582.
- [21] H. Al-Bashiri, M. A. Abdulgaber, A. Romli, and N. Salehudin, "A Developed Collaborative Filtering Similarity Method to Improve the Accuracy of Recommendations under Data Sparsity," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 9, pp. 135-142, 2018.
- [22] H. Koochi and K. Kiani, "A new method to find neighbor users that improves the performance of Collaborative Filtering," *Expert Systems with Applications*, vol. 83, pp. 30-39, 2017.
- [23] B. K. Patra, R. Launonen, V. Ollikainen, and S. Nandi, "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data," *Knowledge-Based Systems*, vol. 82, pp. 163-177, Jul 2015.
- [24] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," *Information Sciences*, vol. 178, pp. 37-51, Jan 2 2008.
- [25] O. S. Revankar and Y. Haribhakta, "Survey On Collaborative Filtering Technique In Recommendation System," *IJAIEM*, ISSN 2319-4847, vol. 4, 2015.
- [26] B. Zhang and B. Yuan, "Improved collaborative filtering recommendation algorithm of similarity measure," in *AIP Conference Proceedings*, 2017, pp. 020167-1-020167-6.
- [27] H. Al-Bashiri, M. A. Abdulgaber, A. Romli, and F. Hujainah, "Collaborative Filtering Recommender System: Overview and Challenges," *Advanced Science Letters*, vol. 23, pp. 9045-9049, 2017.
- [28] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 1999, pp. 230-237.
- [29] S. Kg and G. S. Sadasivam, "Modified Heuristic Similarity Measure for Personalization using Collaborative Filtering Technique," *Appl. Math*, vol. 11, pp. 317-325, 2017.
- [30] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 397-406.
- [31] J. Mao, Z. Cui, P. Zhao, and X. Li, "An improved similarity measure method in Collaborative Filtering Recommendation Algorithm," in *Cloud Computing and Big Data (CloudCom-Asia)*, 2013 International Conference on, 2013, pp. 297-303.
- [32] B.-H. Huang and B.-R. Dai, "A Weighted Distance Similarity Model to Improve the Accuracy of Collaborative Recommender System," in *Mobile Data Management (MDM)*, 2015 16th IEEE International Conference on, 2015, pp. 104-109.
- [33] L. Shunpan, M. Lin, and Fuyongyuan, "A singularity-based user similarity measure for recommender systems," *IJICIC*, vol. 11, pp. 1629-1638, 5 October 2015.
- [34] X. Zang, T. Liu, S. Qiao, W. Gao, J. Wang, X. Sun, et al., "A New Weighted Similarity Method Based on Neighborhood User Contributions for Collaborative Filtering," in *Data Science in Cyberspace (DSC)*, IEEE International Conference on, 2016, pp. 376-381.
- [35] J. Zhang, Y. J. Lin, M. L. Lin, and J. H. Liu, "An effective collaborative filtering algorithm based on user preference clustering," *Applied Intelligence*, vol. 45, pp. 230-240, Sep 2016.
- [36] J. R. Bray and J. T. Curtis, "An ordination of the upland forest communities of southern Wisconsin," *Ecological monographs*, vol. 27, pp. 325-349, 1957.
- [37] M. Kokare, B. Chatterji, and P. Biswas, "Comparison of similarity metrics for texture image retrieval," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, 2003, pp. 571-575.
- [38] S. Viriyavisuthisakul, P. Sanguansat, P. Charnkeitkong, and C. Haruechaiyasak, "A comparison of similarity measures for online social media Thai text classification," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2015 12th International Conference on, 2015, pp. 1-6.