A Recommendation Engine Model for Giant Social Media Platforms using a Probabilistic Approach

Aadil Alshammari
Information Systems Department, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
aadil.al-shammari@nbu.edu.sa

Mohammed Alshammari
Computer Science Department, Faculty of Computing and Information Technology, Northern Border University, Saudi Arabia
mohammed.alshammari@nbu.edu.sa (corresponding author)

Received: 25 August 2023 | Revised: 14 September 2023 | Accepted: 16 September 2023

 DOI: https://doi.org/10.48084/etasr.6325

ABSTRACT

Existing recommender system algorithms often find it difficult to interpret and, as a result, to extract meaningful recommendations from social media. Because of this, there is a growing demand for more powerful algorithms that are able to extract information from low-dimensional spaces. One such approach would be the cutting-edge matrix factorization technique. Facebook is one of the most widely used social networking platforms. It has more than one billion monthly active users who engage with each other on the platform by sharing status updates, images, events, and other types of content. Facebook's mission includes fostering stronger connections between individuals, and to that end, the platform employs techniques from recommender systems in an effort to better comprehend the actions and patterns of its users, after which it suggests forming new connections with other users. However, relatively little study has been done in this area to investigate the low-dimensional spaces included within the black box system by employing methods such as matrix factorization. Using a probabilistic matrix factorization approach, the interactions that users have with the posts of other users, such as liking, commenting, and other similar activities, were utilized in an effort to generate a list of potential friends that the user who is the focus of this work may not yet be familiar with. The proposed model performed better in terms of suggestion accuracy in comparison to the original matrix factorization, which resulted in the creation of a recommendation list that contained more correct information.

Keywords-artificial intelligence; machine learning; recommender systems; probabilistic models; social media

I. INTRODUCTION

In today's internet, the flow of information is constantly expanding, reaching a point where it makes the information in question notably scarce. As a result, it can be difficult for recommender systems to handle the information in an effective and correct manner. On social media sites, individuals connect in a variety of ways, each of which is unique to the user. Some consumers give the recommender system a great deal of detailed and specific information regarding their preferences and wants, which makes the task at hand relatively simple. On the other side, there are some users who rarely directly submit their preferences and choices to the system, making the task of the recommender system in trying to comprehend those users more difficult. Examples of preferences and choices include demographic information and interests. Recommender systems, in general, consist of many types [1], but the most popular are content-based recommender systems and collaborative filtering recommender systems [2-4]. Content-based recommender systems construct their recommendation models based purely on information that is contained within the content itself, such as product descriptions, car colors, movie actors, book authors, users' genders, ages, etc., among other things [5-7]. This necessitates the availability of information in order for the recommender system to carry out its duties in an appropriate manner. The other widely used technique for recommender systems is called collaborative filtering. This technique requires the participation of either users or things in order to compile a list of recommendations. Using the behavior of other users who are similar to the target user, this method analyzes user behavior in an effort to better understand it and then makes product recommendations to that user. Based on this idea, this technique requires users to explicitly show their preferences by doing things like rating more movies, purchasing more products, watching extra clips, etc. This is done so that the recommender system can find similar users or similar items and attempt to build a recommendation list for that user or item based on that similarity, which is all done from the perspective of collaboration. The well-known "cold start" problem can be solved by using the approaches of content-based recommender systems. However, in terms of purely collaborative filtering, this problem has not been resolved [8]. Recommender systems methodologies help in other aspects of life such as agriculture production and other industries [9].

In addition to helping users connect with new people and develop new relationships, one of the purposes of social media platforms is to provide a forum in which users can freely express their thoughts. For this reason, these platforms make use of recommender system techniques to assist in understanding the behavior of their consumers and to recommend new partnerships in an effort to keep users actively engaged on their platforms in order to achieve business goals. As mentioned above, social media platforms have not yet explored the use of probabilistic-wise low-dimensional latent spaces. Due to this, in this paper, we attempt to take advantage of this technique and recommend a more accurate relationship list for the target user, even with a small number of interactions using a probabilistic approach.

In this research, we used a real Facebook dataset to investigate a technique called Probabilistic Matrix Factorization (PMF) [10], with the goal of developing a recommendation engine that is more accurate.

The research questions that we are attempting to address are: Can we crawl the Facebook platform and extract associations between users based on the interactions that they have had with one another using a probabilistic model even when the data are sparse? Is it feasible to determine whether or not two users who have not yet become friends will get along based on how they engage with the friends they share in common? The purpose of this study is to investigate these topics by doing experiments using a real dataset from Facebook.

II. LITERATURE REVIEW

A. Probabilistic Machine Learning

In the field of machine learning, PMF has been integrated into a large number of research projects and further improved through those initiatives. Authors in [11] investigated PMF in order to develop a recommendation model that overcomes the cold start problem and reduced the mistake rate by employing the concept of user trust. According to the authors, if two users X and Y are friends, then they trust each other to a certain degree, depending on the total interaction that each user has had on that platform. When compared to the baseline models, the model demonstrated significantly better performance [11]. Authors in [12] studied how PMF may increase the effectiveness of recommendation engines while keeping in mind that the absence of data is not a random occurrence. They asserted that the majority of the existing algorithms for matrix factorization take into account the possibility that missing data are randomly dispersed across the dataset that was evaluated. However, if a user only interacts with things that interest him or her and ignores everything else, then the missing data are not random. This indicates that the performance of existing models, which assume that missing data are always arranged in a random order, will be skewed. They assert that the proposed strategy is superior to correspondence models in terms of performance. Authors in [13] broaden the usage of PMF to not only recommend a new item to the target user, but also provide a method that analyzes numerous datasets, as well as models, and predicts the performance of such datasets in such models. In addition to this, they used PMF to recommend a new item to the target user. The final model displayed encouraging performance indicators. According to [14, 15], the preferences of users have a vital influence on the improvement of the performance of recommendation engines. The authors demonstrated that the accuracy of a suggestion list may be improved by employing PMF in conjunction with the incorporation of preferred preferences during the process of developing the model. The findings that their model produced were superior to the baselines. According to [16], supplementary information and graphs are another type of supporting content that can assist PMF models in improving their overall performance. According to [17], the addition of side information for both users and items in the PMF models by utilizing the power of convolutional neural networks not only improves the model performance, but also resolves well-known issues within the recommender systems research field such as cold start and sparse data. The authors conducted their analysis of the suggested model utilizing three different datasets, namely MovieLens 100K, MovieLens 1M, and HetRec 2011. The findings revealed that, in comparison to the baselines, the outputs provided by the proposed model were significantly more promising.

B. Probabilistic Matrix Factorization in Social Media

PMF has been implemented in a variety of fields, including medical, social media, image annotation, and other applications [18]. The new model has been proposed in [19] uses PMF to make predictions about the likelihood of a long-non-coding RNA (lncRNA) disease relationship for certain individuals. Constructing three distinct association networks connected to lncRNA, then preprocessing those networks with the KNN method, and finally obtaining a weighted version of the lncRNA are the objectives of this project. After that, PMF generates a likelihood that the lncRNA in question is associated with the target patient. They investigated the model with a variety of datasets pertaining to cancer-related diseases, such as breast, lung, and colorectal cancer, as a means of conducting additional testing on the model. When compared to various baselines, their model produced some encouraging findings. The accuracy of the PMF was also employed by the authors in [20] to calculate the likelihood of an association between certain medications and diseases like diabetes, cardiovascular disease, and neoplasms. They began by constructing a matrix that represented drug-to-disease linkages via the target-pathway-gene midway. This allowed them to make an educated guess as to the likelihood that particular medications are responsible for particular diseases. The new model performed better than the previous model when precision and area under curve metrics were taken into consideration. Using an extended version of PMF designed specifically for the purpose of addressing the image-to-word correlation problem, authors in [21] suggested a solution to the problem. They built a latent space matrix for determining the links between images and words. When compared to existing state-of-the-art models utilizing the Flicker and Corel datasets, the developed model correctly predicted more accurate words annotated to images.

Authors in [22] drew attention to the problem of data scarcity in today's internet and explained how the performance of recommender systems can be influenced in terms of prediction accuracy. They also claimed that the present social recommendation algorithms rely solely on a two-way trust relationship between users. Nevertheless, in their method, they included information about topical attention and confidence in the PMF model and the neural network. This resulted in the extraction of several features that assisted the recommender system in functioning more effectively. They tested their proposed model by comparing its accuracy to that of two different real-world datasets, and the results demonstrated that their model's performance was superior to that of some of the most recent models. Authors in [23] discussed the sparsity problem that occurs in recommendation engines in the social realm. They underlined that the accuracy of predictions and the scalability of current collaborative recommender systems are two additional difficulties that collaborative recommender systems struggle with. In an effort to find a solution to these problems, they suggested a probabilistic model that, in addition to rating history, makes use of social information regarding the users of the system. The goal of this model was to assist systems in better understanding their users and, as a result, in providing more accurate recommendations. When testing the model, they compared it to methods that were considered to be state-of-the art and discovered that their model performed significantly better, especially when dealing with sparse data. It is possible to gain an understanding of the behavior of users on a particular platform by analyzing activities that take place on social networks (such as Twitter and Facebook), such as retweets [24]. In order to construct a prediction model with PMF, authors in [24] created a model that not only takes into account social influence and the semantics of the text, but also activities like retweeting. Their proposed method proved to be superior to existing models that are considered to be state-of-the-art.

When it comes to combating the sparsity problem that adversely impacts the efficiency of recommendation engines, one body of research emphasizes the significance of the ancillary information that social platforms make available, such as tags, ratings, and the reciprocal relationships that exist between users. In [26], the regularization terms that were described earlier were incorporated into the PMF algorithm so that it could take into account the side information. During the evaluation, the authors trained and evaluated the suggested model by making use of the Last.fm dataset. They used the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics to determine how accurate the recommendations were. The findings indicate that their proposed model is superior to the baseline models in terms of performance. Authors in [27] created a framework to analyze tweets in an effort to prevent the spread of SARS-CoV-2 Omicron variant, highlighting the significance of social media in machine learning development. A black box recommendation model was suggested in [28] utilizing the MF approach [29] and applying it to a real Facebook dataset. In the current research we extend that by introducing a probability methodology that makes use of the PMF technique [10], and we do so while utilizing the same data.

III. RESEARCH METHODOLOGY

In this research, we investigate the efficacy of the PMF algorithm in predicting possible interactions between users on Facebook [10]. Users on Facebook connect with one another by giving likes to posts, commenting on photographs, and other similar activities. These interactions can be extended in latent spaces for both users and their friends with hidden features, which ultimately leads to the prediction of how each user will interact with all other users and, as a result, the discovery of reasonable matches. An illustration of the proposed model can be seen in Figure 1.

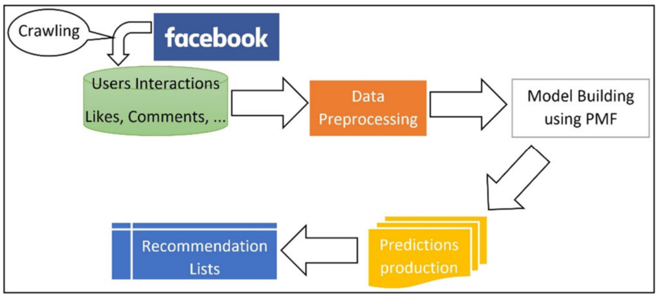

Fig. 1. An illustration of the proposed model beginning with crawling the real data from Facebook and ending with the production of recommendation lists.

The PMF makes use of the Gaussian distribution or the normal distribution in order to observe expected friendship interactions. The following is a representation of the conditional distribution based on the observed interactions:

$$p(R|U,F,\sigma^2) = \prod_{i=1}^{u}\prod_{j=1}^{f} I_{i,j}\,[N(R_{i,j}|g(U_iF_j^T),\sigma^2)] \quad (1)$$

The value of $R_{i,j}$ conveys the frequency with which user i communicates with friend j. U and F refer to the user and friend latent spaces, while u and f are, respectively, the user set and the friend set. The probability density of the Gaussian distribution is denoted by the expression $N\left(R_{i,j}\middle|g\left(U_iF_j^T\right),\sigma^2\right)$. The variance is denoted by σ^2, and the terminating factor is denoted by $I_{i,j}$. The terminating factor has a value of 1 if user i interacts with friend j, and a value of 0 if they do not. $t(x) = \frac{x-1}{N-1}$ is the formula that is used to normalize the interactions such that they sum up to 1, where N is the maximum number of times that two users in our dataset have interacted with one another, which is 84. In addition, the following are the two-zero-mean spherical Gaussian priors that apply to the vectors of both users and friends:

$$p(U|\sigma_U^2) = \prod_{i=1}^{u}[N(U_i|(U_i|0,\sigma_U^2 I)] \quad (2)$$

$$p(F|\sigma_F^2) = \prod_{j=1}^{f}[N(F_i|(F_i|0,\sigma_F^2 I)] \quad (3)$$

The log of the afterward distribution over users and friends can be obtained by:

$$p(U,F|R,\sigma R2,\sigma U2,\sigma F2) \propto p(R|U,F,\sigma R2)p(U|\sigma U2)p(F|\sigma F2) \quad (4)$$

The idea behind the PMF technique is that as the training process progresses, the data distribution of users' friendships and interactions becomes more accurate. The future interactions predicted between users in each training cycle are fed into the subsequent training cycle until both the previous and subsequent predictions converge, and similarity becomes high, at which point the training process should be terminated. As a result, PMF will generate a prediction for all users' mutual interactions even though expressed interactions are few and sparse. We tested this notion using a real dataset crawled from the Facebook platform, which has not been investigated previously in this context. In the following section, we will present our experimental findings. The code is available on GitHub [25].

IV. RESULT ANALYSIS

A. Dataset

We required a real dataset from an asynchronous social network to validate the accuracy of our methodology and evaluate its results. Facebook is the largest social network that provides bidirectional friendship connections, which is the focus of our study. We developed a web crawler to retrieve publicly accessible Facebook users' profiles. The crawler retrieves the users' data and their friends' profiles and proceeds to retrieve profiles of friends of friends. We collected the list of genuine friends with bi-directional relationships, but we did not collect follow connections or other one-directional relationships. To acquire an effective dataset, we ran our web crawler on various user IDs (seeds) from various regions of the United States and the United Kingdom. After one month, we stopped our crawler and as a result we have fetched a total of 16500 user profiles. If any portion of a user's data is not publicly accessible, an empty list will be displayed. For instance, if a user sets his/her friends list to private, his/her friends list will appear as an empty list in our dataset. In addition, for privacy and to ensure anonymity, user IDs are substituted with randomly generated numeric IDs. Every user profile we retrieved from our dataset contains the user ID, friends list, gender, place of origin, self-reported interests, and interactions on posts which are comments and/or likes. For each retrieved post, we collected the post ID and the user IDs of those who liked and/or commented on the respective post. Our collected dataset consists of more than 15 million interactions.

For the purpose of this study, we made use of friends' lists and interactions, and we want to make use of the remaining data for future studies. Facebook employs a number of distinct friendship recommendation algorithms, all of which are geared toward fostering new connections. One of them is called the friends-of-friends (*fof*) algorithm, and its purpose is to locate and promote friendship relationships that are easily accepted. This is because the algorithm searches for and suggests users who are already known to or are already familiar with the platform. This is extremely important for friending algorithms since people have a strong tendency to avoid developing relationship connections with unknown individuals. Our model employs the *fof* algorithm to detect interactive friends, whereas Facebook's algorithm for finding friends of friends finds friends who are already known to the user.

Figure 2 illustrates how the proposed model determines the existence of interacting linkages. In this scenario, Users A, R, S, E, and M are friends of both the target user (x) and the *fof*. If user x interacts with friends with whom user *fof* also interacts, then it is likely that users x and *fof* will interact with one another if the target interacts with friends with whom user *fof* also interacts. This strategy is predicated on the idea that if two people interact with content that was posted by the same set of mutual friends, then it is highly probable that those users will interact with the content that was posted by the other person. The Facebook Friends of Friends method operates under the presumption that if two users have friends in common, then it is highly likely that they are acquainted with one another.

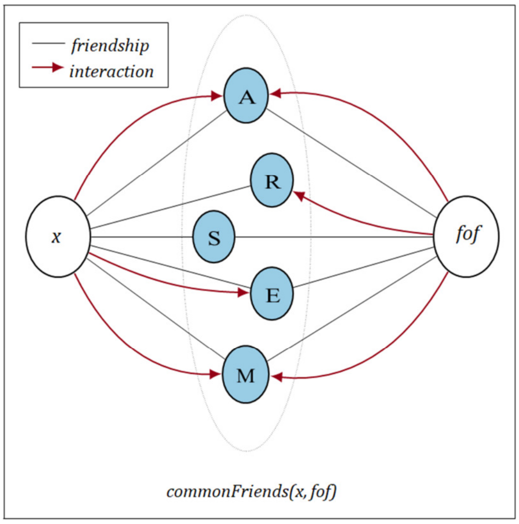

Fig. 2. Interactive friends-of-friends sub-graph example.

B. Experimental Setting

The recommendations for interactive buddies are going to be generated by the proposed model. The acquired dataset from Facebook included the posts that users made, and for each post, we collected a list of all the user IDs of people who liked the relevant posts and interacted with them. In addition to that, we additionally retrieved a list of the user IDs of everyone who had commented on the relevant post. On Facebook, the most prevalent forms of interaction are liking and commenting posts. It's arguable that a user's likes and comments are more accurate reflections of their character than what they say about themselves on their Facebook pages which include a variety of interests and activities. According to the findings of our investigation, therefore, likes and comments are indicators of interacting friends. Figure 3 shows the ways in which the data for our model have been prepared:

User ID	Friend ID	# of Interactions

Fig. 3. Data labels.

To determine the total number of interactions that took place between two users (the user and the friend), we added up all the likes and comments that were passed back and forth between them. When preprocessing the dataset, users who had fewer than three explicit interactions were removed from consideration. There have been a total of 386480 interactions, and a total of 6426 individuals that have engaged in conversation with 8638 friends. Cross-validation was utilized in order to fine-tune the model's hyperparameters. In addition, the number of interactions was brought down to a single level. The unknown interactions that occurred between users during the course of the experiment were shuffled around and given a random order 10 times across the 10 separate iterations of the experiment that we conducted. When reporting the results of the error rate calculation, the mean of all 10 runs was used. In terms of the train-test split, we segmented our dataset so that 90% of it would be used for training, while the remaining 10% would be concealed for testing.

C. Evaluation Metrics

RMSE is the measure that we utilized in order to evaluate how accurate the proposed strategy was. The error rate of the model is computed by adding up all of the discrepancies that exist between the interactions that are expected to take place between Facebook users and the actual interactions that are used in the test set. The values were squared to remove any negative integers and rooted to retrieve the actual information for each user.

$$RMSE = \sqrt{\frac{1}{|T|}\sum_{(i,j)\in T}\left(r'_{ij} - r_{ij}\right)^2} \quad (1)$$

where T denotes the total number of users, while i and j stand in for the user and the friend, respectively. r'_{ij} and r_{ij} denote the anticipated and actual interactions between user i and friend j, respectively.

Figure 4 demonstrates that the performance of the proposed model is superior to that of the baseline MF [26, 27]. As can be seen in Figure 5, the training error rate begins to decrease, as anticipated, after the first epoch. During our experiment, we determined that the optimal number of epochs to use was 10, given that the error rate remained constant beyond that point.

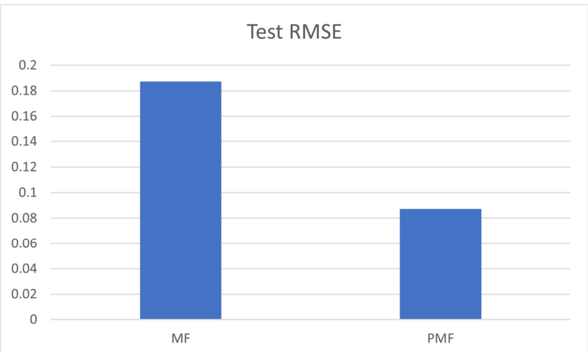

Fig. 4. A comparison of the test RMSE performance between the proposed PMF and baseline MF.

D. Instances of Recommendation Results

We give six specific instances of random target users who have a high number of active users as friends, a mid-range number of active users as friends, and a low number of active users as friends in order to keep things brief. Tables I-VI display the performance of the model in predicting interactive friends for the target users. In Tables I and II, the model reports that the target users 1037 and 289 each have 197 and 181 interactive friends, respectively. It also predicts that the recommended friends will interact with the target users between 8.7 and 10.9 times. The highest expected number of interactions between target users and their recommended interactive friends is reduced when target users engage less frequently with their own friends. According to the data shown in Tables III and IV, the target users have 33 and 44 interactive friends, respectively, and the model expectation for the number of interactions is between 3.25 and 5.19 times.

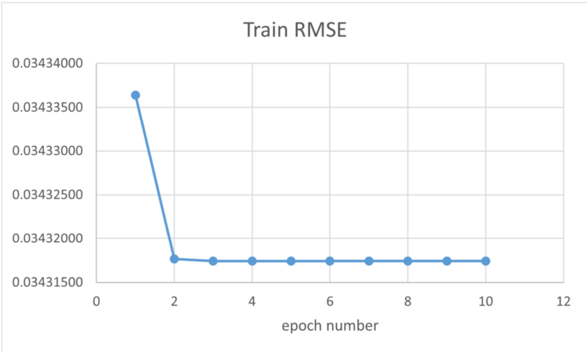

Fig. 5. PMF performance in the training process using RMSE.

TABLE I. EXAMPLE 1 OF HIGHLY ACTIVE TARGET USER

User ID	# of interactive friends	Friend ID	# of predicted interactions
1037	197	16644	10.9
		104649	9.9
		37163	9.8
		50525	9.7
		979710	9.2
		248304	9.0
		235852	8.9
		84052	8.8
		711283	8.5
		682540	8.1

TABLE II. EXAMPLE 2 OF HIGHLY ACTIVE TARGET USER

User ID	# of interactive friends	Friend ID	# of predicted interactions
289	181 friends	2016	11.3
		41538	10.6
		378726	9.6
		328142	9.4
		720814	9.1
		896432	8.7
		246978	8.7
		19361	8.7
		5361	8.6
		83995	8.6

TABLE III. EXAMPLE 1 OF MID-ACTIVE TARGET USER

User ID	# of interactive friends	Friend ID	# of predicted interactions
15354	33 friends	16520	5.1
		13460	4.8
		6363	4.8
		381593	4.8
		242746	4.7
		5878	4.6
		83907	4.6
		4828	4.6
		1380564	4.6
		499184	4.5

TABLE IV. EXAMPLE 2 OF MID-ACTIVE TARGET USER

User ID	# of interactive friends	Friend ID	# of predicted interactions
411568	44 friends	66577	3.9
		5770	3.5
		254107	3.5
		41538	3.4
		13208	3.4
		295620	3.4
		370432	3.3
		65770	3.3
		59526	3.3
		28966	3.2

TABLE V. EXAMPLE 1 OF LOW-ACTIVE TARGET USER

User ID	# of interactive friends	Friend ID	# of predicted interactions
178741	7 friends	247287	2.3
		238092	1.8
		95127	1.7
		484116	1.7
		201396	1.6
		896574	1.6
		24869	1.6
		56805	1.6
		43996	1.6
		257436	1.6

TABLE VI. EXAMPLE 2 OF LOW-ACTIVE TARGET USERS

User ID	# of interactive friends	Friend ID	# of predicted interactions
94606	2 friends	335330	1.3
		56105	1.3
		199802	1.3
		7561	1.2
		94935	1.2
		1587350	1.1
		46648	1.1
		284747	1.0
		395289	1.0
		373383	1.0

Tables V and VI provide examples of a low-target user who has only 2 friends who connect with each other, and the model predicted only 1 to 1.4 times as many interactions between the target user and his or her suggested friends. In conclusion, the performance of the proposed model when applied to the dataset obtained from Facebook depended on the engagement of the user. If the target user is already friends with a large number of highly interactive people, the model will suggest additional highly interactive people who are not currently friends with the target user. The same principle applies to users who are members of interaction groups in the middle and lower classes.

V. CONCLUSION

In conclusion, our goal was to find a solution to the sparsity problem, which occurs when there are insufficient explicit facts and makes it difficult to recommend new friends with a level of accuracy that is considered acceptable. In this paper, we focused our efforts on the goal of building a recommendation model using a probabilistic approach to the matrix factorization method.

This method leverages the known side information in our Facebook dataset, which is the explicit interactions between users, such as likes to posts, comments, and so on, and predicts the unknown interactions between non-friend users, therefore suggesting new friendships that are expected to be accepted. Our dataset is an actual one that was crawled from Facebook, preprocessed, and then added to the model. It was demonstrated that the proposed model performed better than the baseline model and was successful in predicting more accurate suggestions. In addition, our model was successful in predicting strong interactions between the target user and a true hidden friend. In the future, we are interested in including additional supplementary data, such as a user's preferred movies, books, music, etc., in order to improve the precision of the model and to produce explanations for the recommendations. This, in turn, will undoubtedly result in an improvement in the level of satisfaction experienced by users who utilize the system.

REFERENCES

[1] R. Burke, "Hybrid Web Recommender Systems," in *The Adaptive Web*, vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. New York, NY, USA: Springer, 2007, pp. 377–408.

[2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *ACM conference on Computer supported cooperative work*, North Carolina, USA, Oct. 1994, pp. 175–186, https://doi.org/10.1145/192844.192905.

[3] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, Jan. 2003, https://doi.org/10.1109/MIC.2003.1167344.

[4] F. Ricci, "Recommender Systems: Models and Techniques," in *Encyclopedia of Social Network Analysis and Mining*, New York, NY, USA: Springer, 2014, pp. 1511–1522.

[5] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in *Eighteenth National Conference on Artificial Intelligence*, Alberta, Canada, Aug. 2002, pp. 187–192.

[6] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The adaptive web: methods and strategies of web personalization*, Berlin, Germany: Springer, 2007, pp. 325–341.

[7] V. Rohilla, M. Kaur, and S. Chakraborty, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9186–9191, Oct. 2022, https://doi.org/10.48084/etasr.5126.

[8] A. M. Rashid, G. Karypis, and J. Riedl, "Learning preferences of new users in recommender systems: an information theoretic approach," *ACM SIGKDD Explorations Newsletter*, vol. 10, no. 2, pp. 90–100, Sep. 2008, https://doi.org/10.1145/1540276.1540302.

[9] S. R. Gopi and M. Karthikeyan, "Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame

Optimization with Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11360–11365, Aug. 2023, https://doi.org/10.48084/etasr.6092.

[10] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization," in *20th International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, Dec. 2007, pp. 1257–1264.

[11] Y. Dong, S. Fang, K. Jiang, F. Chen, and G. Yin, "Probabilistic Matrix Factorization Recommendation Algorithm with User Trust Similarity," *MATEC Web of Conferences*, vol. 208, 2018, Art. no. 05004, https://doi.org/10.1051/matecconf/201820805004.

[12] J. M. Hernandez-Lobato, N. Houlsby, and Z. Ghahramani, "Probabilistic Matrix Factorization with Non-random Missing Data," in *31st International Conference on Machine Learning*, Beijing, China, Jun. 2014, pp. 1512–1520.

[13] N. Fusi, R. Sheth, and M. Elibol, "Probabilistic Matrix Factorization for Automated Machine Learning," in *32nd Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 2018, vol. 31, pp. 1–10.

[14] J. Liu, C. Wu, Y. Xiong, and W. Liu, "List-wise probabilistic matrix factorization for recommendation," *Information Sciences*, vol. 278, pp. 434–447, Sep. 2014, https://doi.org/10.1016/j.ins.2014.03.063.

[15] M. Alshammari, O. Nasraoui, and S. Sanders, "Mining Semantic Knowledge Graphs to Add Explainability to Black Box Recommender Systems," *IEEE Access*, vol. 7, pp. 110563–110579, 2019, https://doi.org/10.1109/ACCESS.2019.2934633.

[16] T. Zhou, H. Shan, A. Banerjee, and G. Sapiro, "Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information," in *12th SIAM International Conference on Data Mining*, Anaheim, CA, USA, Apr. 2012, pp. 403–414, https://doi.org/10.1137/1.9781611972825.35.

[17] K. Li, X. Zhou, F. Lin, W. Zeng, and G. Alterovitz, "Deep Probabilistic Matrix Factorization Framework for Online Collaborative Filtering," *IEEE Access*, vol. 7, pp. 56117–56128, 2019, https://doi.org/10.1109/ACCESS.2019.2900698.

[18] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. Cambridge, MA, United States: MIT Press, 2022.

[19] Z. Xuan, J. Li, J. Yu, X. Feng, B. Zhao, and L. Wang, "A Probabilistic Matrix Factorization Method for Identifying lncRNA-Disease Associations," *Genes*, vol. 10, no. 2, Feb. 2019, Art. no. 126, https://doi.org/10.3390/genes10020126.

[20] J. Yang, Z. Li, X. Fan, and Y. Cheng, "Drug–Disease Association and Drug-Repositioning Predictions in Complex Diseases Using Causal Inference–Probabilistic Matrix Factorization," *Journal of Chemical Information and Modeling*, vol. 54, no. 9, pp. 2562–2569, Sep. 2014, https://doi.org/10.1021/ci500340n.

[21] Z. Li, J. Liu, X. Zhu, T. Liu, and H. Lu, "Image annotation using multi-correlation probabilistic matrix factorization," in *18th ACM international Conference on Multimedia*, Florence, Italy, Oct. 2010, pp. 1187–1190, https://doi.org/10.1145/1873951.1874183.

[22] W. Zhang, F. Liu, D. Xu, and L. Jiang, "Recommendation system in social networks with topical attention and probabilistic matrix factorization," *PLOS ONE*, vol. 14, no. 10, Sep. 2019, Art. no. e0223967, https://doi.org/10.1371/journal.pone.0223967.

[23] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: social recommendation using probabilistic matrix factorization," in *17th ACM conference on Information and knowledge management*, Napa, CA, USA, Oct. 2008, pp. 931–940, https://doi.org/10.1145/1458082.1458205.

[24] B. Jiang, Z. Lu, N. Li, J. Wu, and Z. Jiang, "Retweet Prediction Using Social-Aware Probabilistic Matrix Factorization," in *18th International Conference on Computational Science*, Wuxi, China, Jun. 2018, pp. 316–327, https://doi.org/10.1007/978-3-319-93698-7_24.

[25] D. M. S. Alshammari, "FacebookPMF." May 24, 2023, Accessed: Sep. 21, 2023. [Online]. Available: https://github.com/drmohammedsanad/FacebookPMF.

[26] Z. Liu and H. Zhong, "Study on Tag, Trust and Probability Matrix Factorization Based Social Network Recommendation," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 5, pp. 2082–2102, 2018, https://doi.org/10.3837/tiis.2018.05.010.

[27] M. Mahyoob, J. Algaraady, M. Alrahiali, and A. Alblwi, "Sentiment Analysis of Public Tweets Towards the Emergence of SARS-CoV-2 Omicron Variant: A Social Media Analytics Framework," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8525–8531, Jun. 2022, https://doi.org/10.48084/etasr.4865.

[28] M. Alshammari and A. Alshammari, "Friend Recommendation Engine for Facebook Users via Collaborative Filtering," *International Journal of Computers Communications & Control*, vol. 18, no. 2, Apr. 2023, Art. no. 4998, https://doi.org/10.15837/ijccc.2023.2.4998.

[29] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol. 42, no. 8, pp. 30–37, Dec. 2009, https://doi.org/10.1109/MC.2009.263.

AUTHORS PROFILE

Aadil Alshammari received his Ph.D. degree in computer science from the New Mexico Institute of Mining and Technology, USA. He is a Faculty Member of the College of Computing and Information Technology, Northern Border University, Saudi Arabia. He received the M.S. degree in information systems from the University of Surrey, UK. His research interests include recommender algorithms, social networks, machine learning and data analytics.

Mohammed Alshammari received a Ph.D. in Computer Science from the Computer Engineering and Computer Science Department at the University of Louisville, Kentucky, USA, in 2019. He was a lecturer in the Department of Computer Science at the Faculty of Computing and Information Technology at the Northern Border University in Saudi Arabia from 2012 to 2019, and an assistant professor from 2019 until now. His areas of interest include artificial intelligence, machine learning, semantic web, recommender systems, and explainability.