International Journal of Information Retrieval Research
Volume 12 • Issue 2

A Hybrid Approach of Prediction Using Rating and Review Data

Aseem Srivastava, KNIT Sultanpur,India

Rafeeq Ahmed, Department of Computer Science & Engineering, KL University, Vaddeswaram, Andhra Pradesh, India

Pradeep Kumar Singh, GLA University, India

Mohammed Shuaib, Department of Computer Science, College of Computer Science & IT, Jazan University, Jazan, Saudi Arabia*

https://orcid.org/0000-0001-6657-2587

Tanweer Alam, Islamic University of Madinah, Saudi Arabia

https://orcid.org/0000-0003-2731-4627

ABSTRACT

A collaborative filtering technique has proven to be the preferable approach for personalized recommendations. Traditionally, collaborative filtering recommends target items to those users who have similar tastes. The performance of collaborative filtering degrades significantly when a considerable number of users do not provide ratings on recommended products. In such a scenario, the dataset utilized in recommendation becomes highly sparse, and ratings become very few or none co-rated. To mitigate the problem, as mentioned earlier, and to improve the performance of collaborative filtering, the authors propose an approach that adopts users' textual reviews and ratings in the rating prediction. The dataset used is Amazon Fine Food Reviews containing rating and text review with 568,454 reviews from October 1999 to October 2012. The proposed model is tested on the collected dataset. The experimental results provide the proper evidence that the proposed model outperforms traditional algorithms of collaborative filtering techniques.

KEYWORDS

Collaborative Filtering, Content-Based Filtering, Prediction Approaches, Recommendation System, Similarity Metrics, Text Mining

INTRODUCTION

The need for the Recommender System (RS) has been increasing with the number of data companies generating these days. It's worth mentioning how these recommendation systems use user-specific and item-specific details to understand recommendations better. The key target remains traditional, that is to recommend personalized recommendations of a product to similar users or by finding similar products for the same user. To complete this objective, we use the filtering process. Filtering algorithms in a recommendation system are studied in two forms, one being content-based-filtering (CBF) and the second being, collaborative filtering (CF).

CBF focused on features such as profiles and items. With such specifications and the rating feedback of users for each set of things in the filtering process. However, this filtering process offers

DOI: 10.4018/IJIRR.299942

*Corresponding Author

many advantages to the users but also has some limitations. These limitations are limited content analysis and over-specialization etc. On the other hand, the general philosophy is that similar users have similar tastes or related items that follow the same rating pattern.

To target the specified problem statement, there are a plethora of workflows available. Still, we introduce you to a novel rating prediction based recommendation system that surpasses either of the two CF-based algorithms that could be bifurcated as model-based CF or memory-based CF. In general, it is observed that model-based CF may provide better accuracy than memory-based, CF, but it has some limitations such as inflexibility and quality of predictions. Memory-based CF adopts a similarity computation method, which can be further bifurcated into user-based and item-based CF. Apart from this, we used user-similarity in predicting the rating in our proposed user-based CF approach, and at the same time, item similarity is used in item-based CF.

The intensive research on recommendation system components becomes very tedious when users generally do not provide the items' rating. Since we are dealing with the Amazon products daily, the number of users buying the same thing is comparatively very high than the number of ratings received in the user's feedback. Also, it remains null in many cases, which means the product has been rated even once. The above phenomenon increases the sparsity in the similarity matrix. Out of all the outstanding work yet proposed by many researchers to mitigate high sparsity weakness in CF-based RS. Most of all, utilize only the users' textual review Chen et al. (2015) Wu (2020) and ratings to mitigate the existing dataset's high sparsity level. As we propagated our research in the proposed model's direction, we observed that our proposed CF-based RS had significantly improved the performance. Our contribution in this paper has been summarized below:

- Apply traditional CF-based methodology, i.e. matrix factorization using SVD to predict the ratings.
- Apply proposed recommendation methodology, i.e. recommendation based on the users' reviews and rating as provided in the feedback process.
- Comparison of traditional CF algorithm and proposed novel recommendation algorithm.

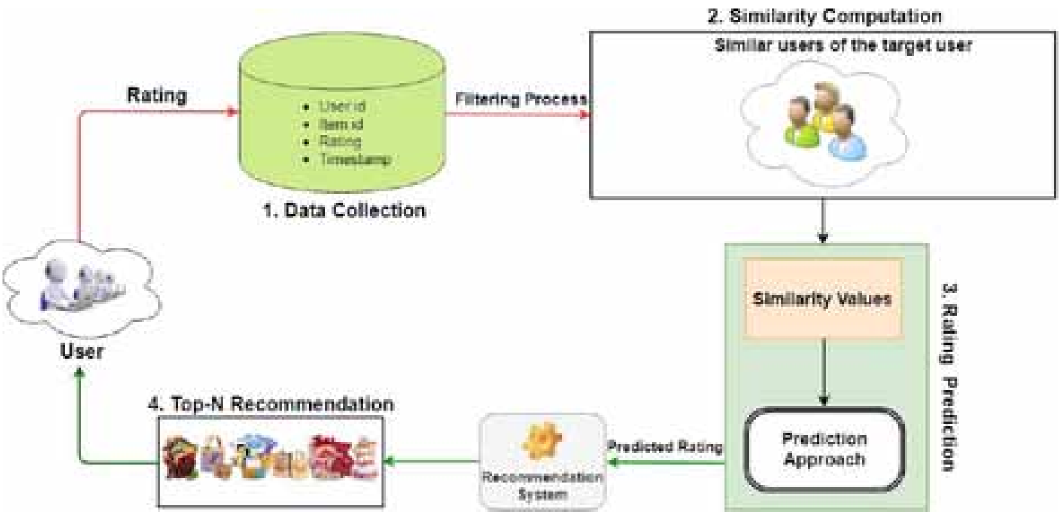
BACKGROUND

Findings users' interest is always a challenging task due to the dynamic nature of a user. CF has become a solution to deal with this challenge described above since the early '90s. User-based CF is the most widely used technique in memory-based CF in the past few years. To achieve CF's general philosophy, a user-based CF utilizes a framework, as shown in figure 1. The detailed description of the framework of the user-based CF is as follows:

Data Collection

Let U= $\{\alpha_1, \alpha_2, \ldots . \alpha_m\}$ and I=$\{\beta_1, \beta_2. \ldots .. \beta_n\}$ are the sets of m users and n items. A user-item rating matrix can be formed when a user gave a rating to a particular item. Implicit and explicit are the two methods utilized in data collection of CF Núñez-Valdez et al. (2018). In implicit data collection; systems gather data from users' activities such as searching behaviour, browsing history, time spent on a particular item, cursor movement, etc. On the other hand, in an explicit method, the system collects data directly from the user in reviews or ratings Berbatova (2019).

Figure 1. Framework of the user-based Collaborative Filtering



Finding k Similar User

Computing k most similar use of the target user in a sparse dataset is a challenging process. Various similarity measures (SMEs) have been utilized to improve the performance of CF. Table 1 shows the notations used in the computation of similarity measure (SM), and table 2 represents a list of existing SMs in the literature of CF.

Table 1. Notations used in Similarity Measure

Notation	Definition
$\Psi\ (\alpha_1, \alpha_2)$	Correlation of users α_1 and α_2
$\eta_{\alpha\beta}$, $r_{\alpha\beta}$	User’s (α) ranking on object β
$\overline{\eta_{\beta}}$	Mean Object (β) Ranking
$\lvert I_{\alpha_1,\alpha_2} \rvert$	Max number of scores (ratings) for two α_1 and α_2 users
$\lvert U_{\beta_1,\beta_2} \rvert$	Total number of users who rate both the items β_1, and $^2{}_2$.
$\overline{\eta_{\alpha}}$	Average rating of user α.
$K_{\alpha', \beta}$	Rank of user α based on ratings of item β.
$\overline{k_{\alpha}}$	Average rank of user α based on ratings of items.

COLLABORATIVE FILTERING ALGORITHM

Similarity Approaches

The following similarity approaches have been generally used in many fields like text mining, users similarity or item similarity in recommendation systems Zhu (2020a) Zhu (2020b) Yang et al. (2017) Ah- mad et al. (2017).

Table 2. Similarity formula for items

Approach	Similarity Equation
Cosine Similarity	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}}\right)\left(\eta_{\pm,^{2}{}_{2}}\right)}{\sqrt[2]{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}}\right)^2}\sqrt[2]{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{2}}\right)^2}}$
Adjusted Cosine Similarity	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \overline{\eta_{\pm}}\right)\left(\eta_{\pm,^{2}{}_{2}} - \overline{\eta_{\pm}}\right)}{\sqrt[2]{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \overline{\eta_{\pm}}\right)^2}\sqrt[2]{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{2}} - \overline{\eta_{\pm}}\right)^2}}$
Euclidean Distance	$\Psi\left(\beta_1,\beta_2\right) = \sqrt[2]{\frac{\sum_{\alpha\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \eta_{\pm,^{2}{}_{2}}\right)^2}{\left\vert U_{\beta_1,\beta_2}\right\vert}}$
Manhattan Distance	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{\alpha\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \eta_{\pm,^{2}{}_{2}}\right)^2}{\left\vert U_{\beta_1,\beta_2}\right\vert}$
Pearson Correlation	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \overline{\eta_{\beta_1}}\right)\left(\eta_{\pm,^{2}{}_{2}} - \overline{\eta_{\beta_2}}\right)}{\sqrt[2]{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \overline{\eta_{\beta_1}}\right)^2}\sqrt[2]{\sum_{u,v\epsilon U}\left(\eta_{\pm,^{2}{}_{2}} - \overline{\eta_{\beta_2}}\right)^2}}$
Spearman Correlation	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{u\epsilon U}\left(k_{\pm,^{2}{}_{1}} - \bar{k}_{2_{1}}\right)\left(k_{\pm,^{2}{}_{2}} - \bar{k}_{2_{2}}\right)}{\sqrt[2]{\sum_{\pm\epsilon U}\left(k_{\pm,^{2}{}_{1}} - \bar{k}_{2_{1}}\right)^2}\sqrt[2]{\sum_{\pm\epsilon U}\left(k_{\pm,^{2}{}_{2}} - \bar{k}_{2_{2}}\right)^2}}$
Bhattacharya Coefficient	$\Psi\left(\beta_1,\beta_2\right) = Jacc\left(\beta_1,\beta_2\right) + \sum_{\acute{a}_1\epsilon U_i}\sum_{\acute{a}_2\epsilon U_j} BC\left(\acute{a}_1,\acute{a}_2\right) loc\left(r_{i_u},\ r_{i_v}\right)$
Mean Squared Distance	$\Psi\left(\beta_1,\beta_2\right) = \frac{\sum_{\pm\epsilon U}\left(\eta_{\pm,^{2}{}_{1}} - \eta_{\pm,^{2}{}_{2}}\right)}{\left\vert U_{\beta_1,\beta_2}\right\vert}$

Rating Prediction

Prediction approaches Zhang et al. (2020) has been given below. These methods are:

Table 3. Prediction of item for user

Mean Centering (MC):	$\hat{r}_{\alpha\beta_1} = \overline{r}_{\beta_1} + \frac{\sum_{\beta_2 \epsilon N_u(\beta_1)} \Psi\left(\beta_1,\beta_2\right)\left(r_{\alpha\beta_2} - r_{\beta_2}\right)}{\sum_{\beta_2 \epsilon N_u(\beta_1)} \left\|\Psi\left(\beta_1,\beta_2\right)\right\|}$
Weighted Average (WA):	$\hat{r}_{\alpha\beta_1} = \frac{\sum_{\beta_2 \epsilon N_u(\beta_1)} \Psi\left(\beta_1,\beta_2\right)\left(r_{\alpha\beta_2}\right)}{\sum_{\beta_2 \epsilon N_u(\beta_1)} \left\|\Psi\left(\beta_1,\beta_2\right)\right\|}$
Z-Score (ZS):	$\hat{r}_{\alpha\beta_1} = \overline{r}_{\alpha} + \sigma_{\beta_1} \frac{\sum_{\beta_2 \epsilon N_u(\beta_1)} \Psi\left(\beta_1,\beta_2\right)\left(r_{\alpha\beta_2} - \overline{r_{\beta_2}}\right) / \sigma_{\beta_2}}{\sum_{\beta_2 \epsilon N_u(\beta_1)} \left\|\Psi\left(\beta_1,\beta_2\right)\right\|}$

Featured N-best recommendation

The performance of RS is measured by the accuracy of this recommended list of items. Some recent work on rating-based CF is as follows:

- In the computation of similarity using JS, Only the ratio of co-rated items has been considered instead of the absolute value of rating. Therefore, Mubbashir Ayub et al. have proposed a modified Jaccard similarity to mitigate the existing problem of JS Ayub et al. (2018). As a result, the modified equation of JS uses the average rating of users. AL-Bakri et al. have utilized the Pearson Correlation, Jaccard distance measure, and inverse user frequency in their proposed model Al-Bakri and Hashim (2018). The authors previously mentioned that even when user ratings are very few and zero co-rated, all SMs excluding BC couldn't evaluate SM. Consequently, in a very sparse data gathering, the above-proposed solution has some drawbacks.
- Capturing users' interest is always a challenging task because the interests of the users change over time. Haoyuan Feng et al. used a time-weighted association rule mining rule to incorporate a sequential overlapping detection mechanism in this line Feng et al. (2015). The advantage of the proposed method is that it can enhance the diversity in the recommendation. Daniel Kluver et al. have provided reviews of the CF algorithms Kluver et al. (2018). Furthermore, the Evaluation of these CF-algorithms has also been discussed in this paper.

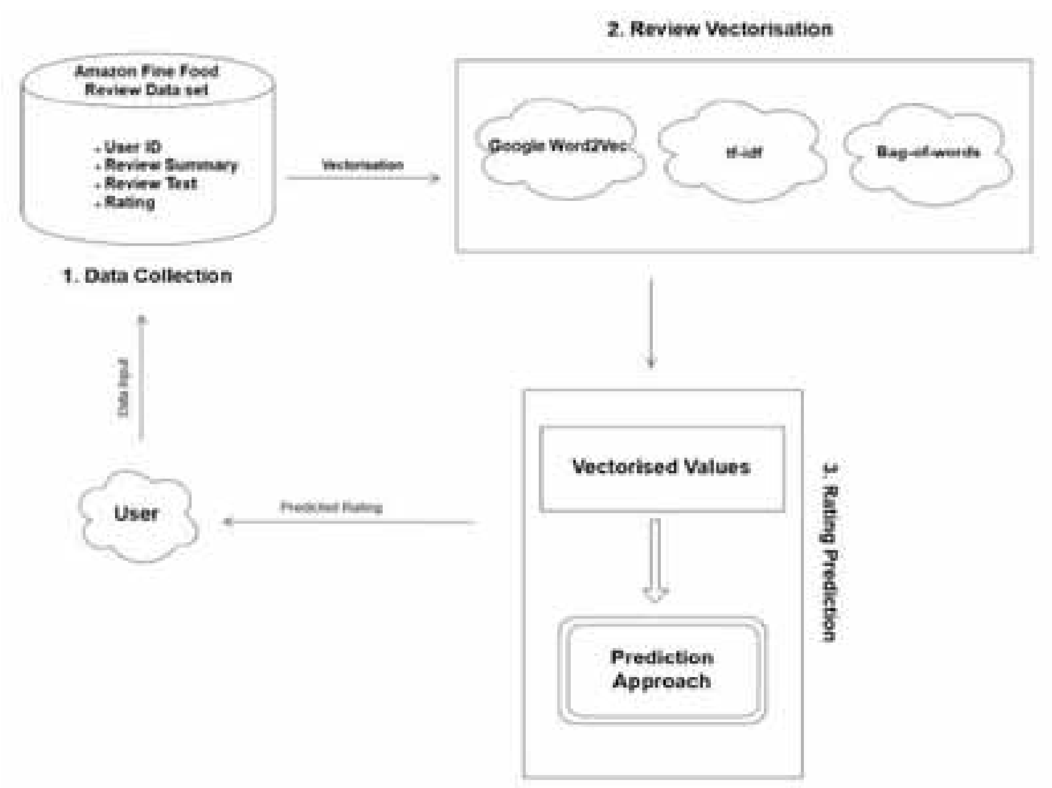
Prediction techniques may very well play an essential role in CF-based RS precision. Hence, Hael Al- Bashiri et al. have introduced a TOPSIS method as a prediction approach Al-Bakri and Hashim (2018). The experimental results of the TOPSIS method have provided evidence that it outperforms the other CF-algorithms. Sparsity is a vital problem of CF-based RS. As stated earlier, Patra et al. proposed a new similarity measure using BC to deal with Patra et al. (2015). It can compute the similarity between users quickly where their ratings are few or none co-rated. In this direction, to further enhance the accuracy of CF-based RS, P Jain et al. have applied the contextual information of users and items with BC as an SM Dixit and Jain (2019). W.yang et al. have illustrated the bidirectional encoder representation transformers model to extract the users' rating styles Yang et al. (2019). Furthermore, the proposed model has used in the elimination of rating style differences of users and the computation of similarity measure Shuaib et al. (2020b)– Siddiqui et al. (2020). It can be seen that only a user-item ranking dataset is used by all the recent works to enhance the recommendation accuracy of CF-based RS.

IMPLEMENTATION

The general framework has been shown in Fig 2. Data collection, review vectorization Ahmed and Ahmad (2019) and rating predictions are the proposed framework's key components. Document preprocessing has been done to make the data ready to be used in the algorithm Ahmed et al. (2018). The necessary steps applied in the data are:-

- Notation like "won't" is converted to "Would not", "Couldn't" corrected to "Could not."
- Stemming is done to get the root words
- Stopwords like is, am, are, etc. are removed
- Words with number or symbols in them are removed
- All reviews lowercased
- URLs removed

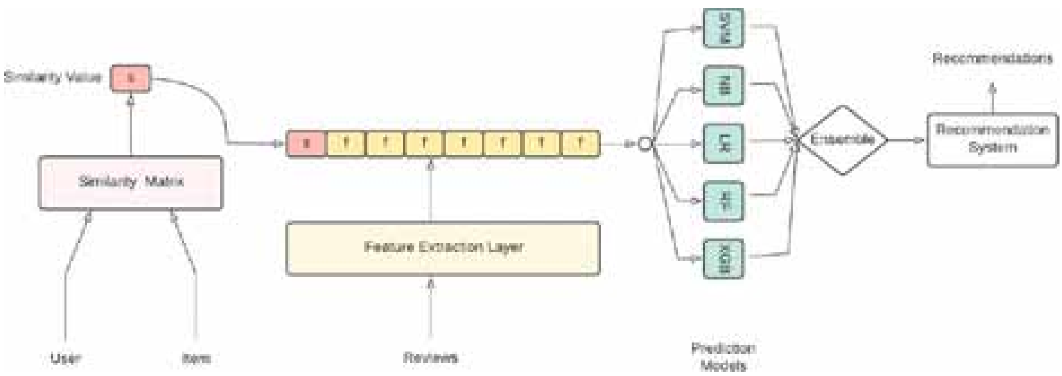
Figure 2. Framework of text mining based recommendation



EXPERIMENTAL ANALYSIS

The similarity matrix formation in the initial steps is concluded concerning the similarity formula mentioned in figure 2. Feature Extraction Layer is configured with tfidf/word2vec, which produced embedding further concatenated with the similarity value. The prediction model is a layer of multiple individual models on which ensembling has been performed. The recommendation system receives ensemble predictions to generate recommendations.

Figure 3. Proposed architecture of the model is shown above



The dataset used is Amazon fine Food Reviews containing rating and text review. The data collected in the experiment phase is from Stanford University, available on Kaggle. The data set collected 568454 reviews from October 1999 to October 2012. The number of users is 256059and number of products being rated and reviewed are 74258. The schema of the dataset is shown in figure 3

Figure 4. Schema of collected dataset

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...

The data's essential features are Id, ProductId, User Id, Profile Name, Score (rating between 1 and 5), Timestamp, Summary of review and Text (text of the study). The tabular presentation of data shows Score as a rating provided by the user and Summary and Text of reviews offered by corresponding users. To use review-text to predict the rating of the products, reviews needed to be preprocessed and cleaned. The above data set contains scores as each product's rating for the user's corresponding review for a particular product is given. Fig 5 and 6 denote the pictorial characteristics of the collected data. The above data set contains scores as each product's rating for the user's corresponding review for a particular product is given. The same data is split into train, cross-validation and test dataset using time-based splitting. First, 80% of time-sorted information is considered train data, whereas the last 20% of data is kept for testing purposes. The data collected in the experiment phase is from

Stanford University Zhang et al. (2020). The schema of the dataset is shown below: Furthermore, the same data is split into train, cross-validation and test dataset using time-based splitting. First, 80 of time-sorted information is considered a strain data, whereas the last 20 data are kept for testing purposes. A user-item similarity matrix is generated with userID, productID and rating as parameters to perform collaborative filtering. To perform Matrix Factorization using SVD to predict rating in the sparse matrix of user-item relation matrix, a sparse matrix is first transformed to fit in the surprise library. A baseline model is trained and tested to predict a standard weak MAE and RMSE score so that, rating accuracy can be predicted with a lower MAE and RMSE score than the baseline model.

Figure 5. Rating distribution in the dataset

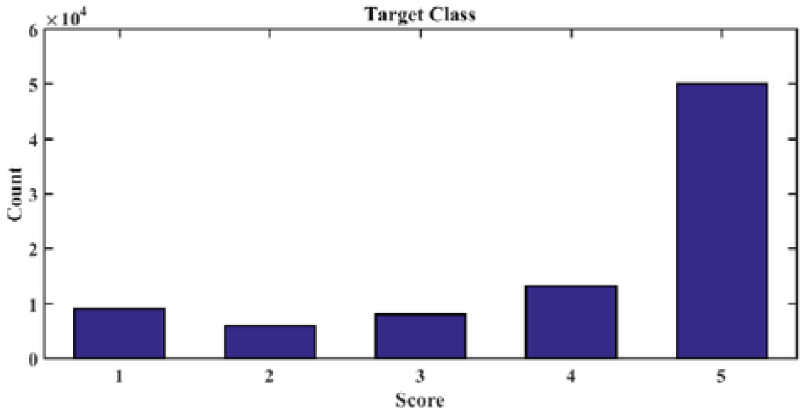


Figure 6. Unique words of collected dataset

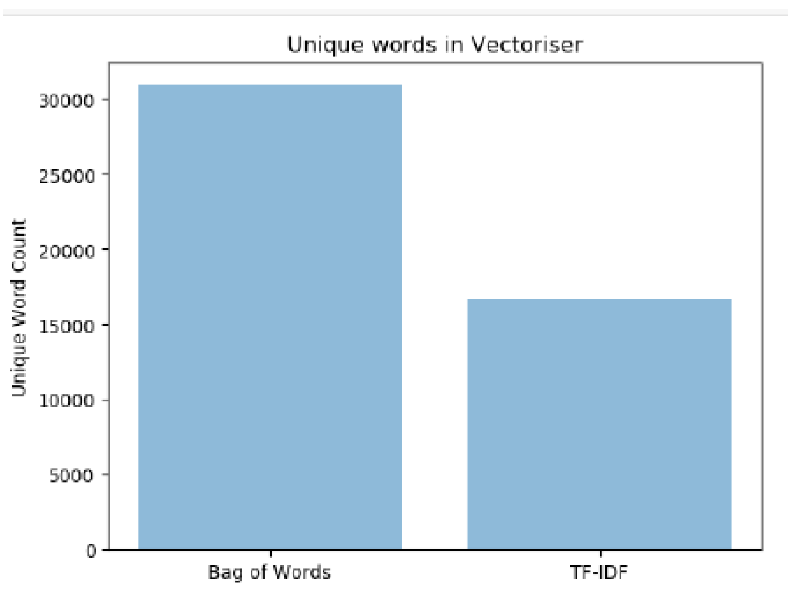
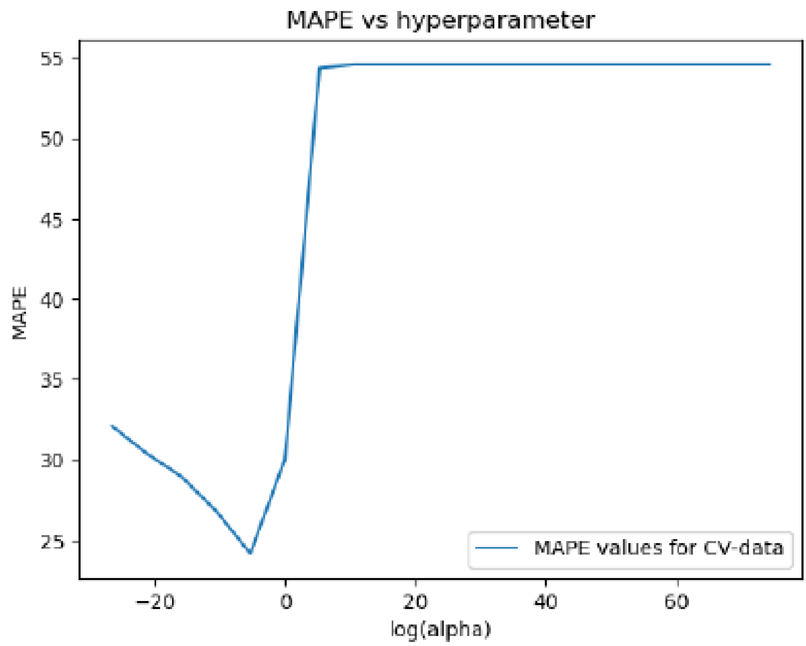


Figure 7. Comparison of MAPE vs Hyper parameter in used dataset



Furthermore, to cover the test matrix's sparsity employing SVD, the Dimensionality Reduction method is used. The following results were obtained using SVD to predict rating in the test matrix sparsity, as shown in Figures 8 and 9. Dataset is sorted by time and split into train and test set as in ration 80% and 20% respectively. Then reviews are vectored using Bag of Words and TF-IDF vectorizer. The equations of MAE and RMSE are as follows:

$$MAE = \frac{\sum_{i=1}^{N} \left| p_{\beta i} - \widehat{q_{\beta i}} \right|}{N}$$

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{N} \left(p_{\beta i} - q_{\beta i} \right)^{2}}{N}}$$

Here, $p_{\beta i}$ and $q_{\beta i}$ define the expected and actual rating of item β β, respectively. The overall number of forecasts is explained by N.

From fig 8 to 9, we can easily observe that the proposed approach attains low MAE and RMSE values than prevalent CF algorithm.

Figure 8. User-Item count in the train and test matrix

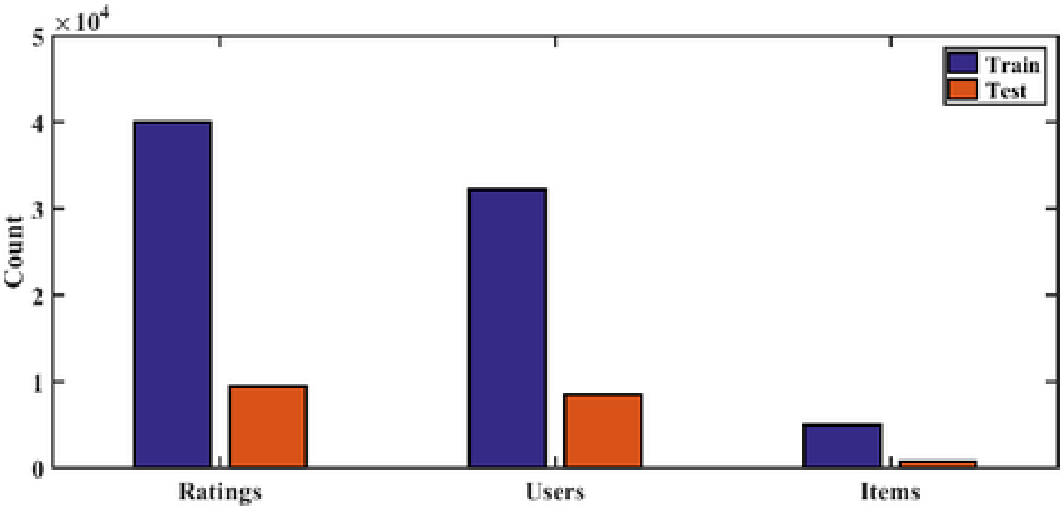
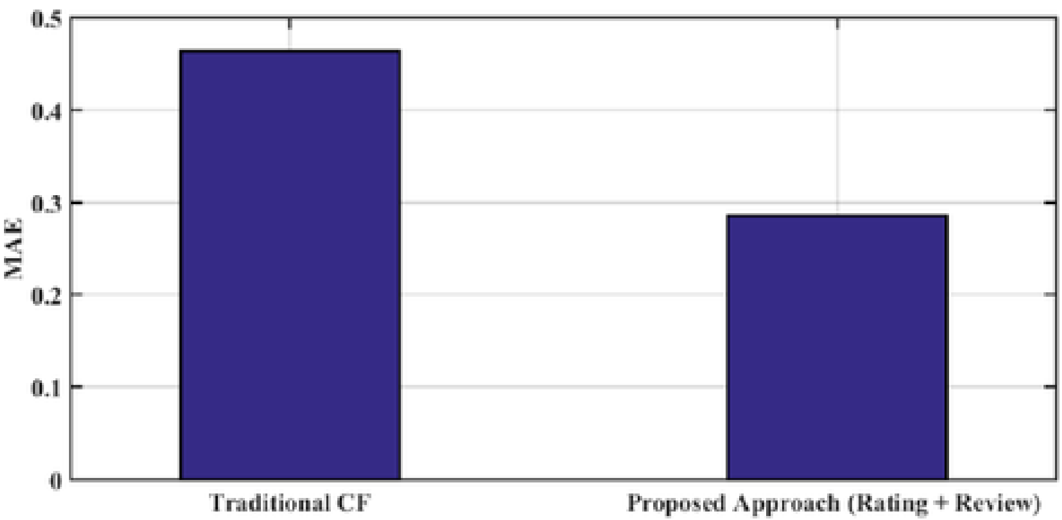


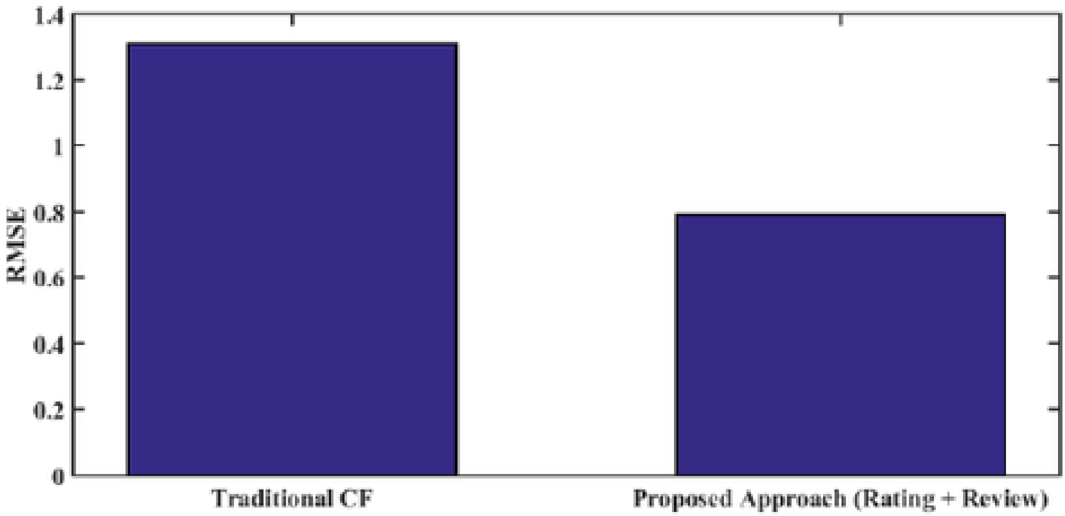
Figure 9. Comparison based on MAE



CONCLUSION

While building a recommendation engine to recommend a product that fits into users' interest, using traditional approaches like matrix factorization technique, including SVD to recommend the product works well on data with considerable sparsity. Still, it is observed that the traditional techniques fail with a very sparse dataset where user-item similarity cannot show relevant results and cannot predict similarity to obtain good accuracy. Therefore, an alternative method is used to obtain better results to recommend products to the user at better accuracy where reviews given to each product are used and its corresponding rating available in the dataset.

Figure 10. Comparison based on RMSE



FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

Ahmad, T., Ahmad, R., Masud, S., & Nilofer, F. (2017). Framework to extract context vectors from unstructured data using big data analytics. *2016 9th International Conference on Contemporary Computing, IC3 2016*, 1–6.

Ahmed, R., & Ahmad, T. (2019). Fuzzy Cross Domain Concept Mining. *International Journal of Innovative Technology and Exploring Engineering*, *8*(10S), 326–333. doi:10.35940/ijitee.J1059.08810S19

Ahmed, R., Singh, P., & Ahmad, T. (2018). *Novel Semantic Relatedness Computation for Multi- Domain Unstructured Data*. EAI Endorsed Transactions on Energy Web. doi:10.4108/eai.13-7-2018.165503

Al-Bakri, N. F., & Hashim, S. H. (2018). A Modified Similarity Measure for Improving Accuracy of User-Based Collaborative Filtering. *IRAQI J. Sci*, *59*(2B).

Alam, S., Shuaib, M., & Samad, A. (2019). A Collaborative Study of Intrusion Detection and Pre- vention Techniques in Cloud Computing. *Lecture Notes in Networks and Systems*, *55*, 231–240. doi:10.1007/978-981-13-2324-9_23

Alam, S., Siddiqui, S. T., Ahmad, A., Ahmad, R., & Shuaib, M. (2020). Internet of Things (IoT) Enabling Technologies, Requirements, and Security Challenges. *Lecture Notes in Networks and Systems*, *94*, 119–126. doi:10.1007/978-981-15-0694-9_12

Ayub, M., Ghazanfar, M. A., Maqsood, M., & Saleem, A. (2018). A Jaccard base similarity measure to improve performance of CF based recommender systems. *2018 International Conference on Information Networking (ICOIN)*, 1–6. doi:10.1109/ICOIN.2018.8343073

Berbatova, M. (2019). Overview on NLP Techniques for Content-Based Recommender Systems for Books. *Proceedings of the Student Research Workshop Associated with RANLP 2019*, 55–61. doi:10.26615/issn.2603-2821.2019_009

Chen, L., Chen, G., & Wang, F. (2015). Recommender systems based on user reviews: The state of the art. *User Modeling and User-Adapted Interaction*, *25*(2), 99–154. doi:10.1007/s11257-015-9155-5

Dixit, V. S., & Jain, P. (2019). Proposed similarity measure using Bhattacharyya coefficient for context aware recommender system. *Journal of Intelligent & Fuzzy Systems*, *36*(4), 3105–3117. doi:10.3233/JIFS-18341

Feng, H., Tian, J., Wang, H. J., & Li, M. (2015). Personalized recommendations based on time- weighted overlapping community detection. *Information & Management*, *52*(7), 789–800. doi:10.1016/j.im.2015.02.004

Kluver, D., Ekstrand, M. D., & Konstan, J. A. (2018). Rating-Based Collaborative Filtering: Algo- rithms and Evaluation. Lecture Notes in Computer Science, 10100, 344–390.

Núñez-Valdez, E. R., Quintana, D., Crespo, R. G., Isasi, P., & Herrera-Viedma, E. (2018). A recommender system based on implicit feedback for selective dissemination of ebooks. *Information Sciences*, *467*, 87–98. doi:10.1016/j.ins.2018.07.068

Patra, B. K., Launonen, R., Ollikainen, V., & Nandi, S. (2015). A new similarity measure using Bhat- tacharyya coefficient for collaborative filtering in sparse data. *Knowledge-Based Systems*, *82*, 163–177. doi:10.1016/j.knosys.2015.03.001

Samad, A., Shuaib, M., & Beg, M. R. (2017). Monitoring of Military Base Station using Flood- ing and ACO Technique: An Efficient Approach. *International Journal of Computer Network and Information Security*, *9*(12), 36–44. doi:10.5815/ijcnis.2017.12.05

Shuaib, M., Alam, S., Mohd, S., & Ahmad, S. (2020a). Blockchain-Based Initiatives in Social Security Sector. *EAI 2nd International Conference on ICT for Digital, Smart, and Sustainable Development (ICIDSSD)*, 8–8.

Shuaib, M., Daud, S. M., Alam, S., & Khan, W. Z. (2020b). Blockchain-based framework for secure and reliable land registry system. *TELKOMNIKA*, *18*(5), 2560–2560. doi:10.12928/telkomnika.v18i5.15787

Siddiqui, S. T., Alam, S., Ahmad, R., & Shuaib, M. (2020). Security Threats, Attacks, and Possible Countermeasures in Internet of Things. *Lecture Notes in Networks and Systems*, *94*, 35–46. doi:10.1007/978-981-15-0694-9_5

Wu, C., Wang, C., Zhou, Y., Wu, D., Chen, M., Wang, J. H., & Qin, J. (2020). Exploiting user reviews for automatic movie tagging. *Multimedia Tools and Applications*, *79*(17-18), 11399–11419. doi:10.1007/s11042-019-08513-0

Yang, B., Lei, Y., Liu, J., & Li, W. (2017). Social Collaborative Filtering by Trust. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(8), 1633–1647. doi:10.1109/TPAMI.2016.2605085 PMID:27608451

Yang, W., Guo, S. H., & Zhang, C. J. (2019). A novel rating style mining method to improve collab- orative filtering algorithm. *Journal of Physics: Conference Series*, *1187*(5), 52101–52101. doi:10.1088/1742-6596/1187/5/052101

Zhang, Z., Dong, M., Ota, K., & Kudo, Y. (2020). Alleviating New User Cold-Start in User-Based Collaborative Filtering via Bipartite Network. *IEEE Transactions on Computational Social Systems*, *7*(3), 672–685. doi:10.1109/TCSS.2020.2971942

Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., & Cai, D. (2020a). Addressing the Item Cold-Start Problem by Attribute-Driven Active Learning. *IEEE Transactions on Knowledge and Data Engineering*, *32*(4), 631–644. doi:10.1109/TKDE.2019.2891530

Aseem Srivastava is a Research Scholar at IIIT Delhi. He did his undergrad research under Dr. Rafeeq Ahmed in the domain of Natural Language Processing. He currently works on problems related to Natural Language Processing and Machine Learning. In particular, his research interest includes NLP to improvise Goal-Oriented Dialogue System.

Rafeeq Ahmed is currently working as an Associate Professor in CSE Department, KL University, Guntur, Andhra Pradesh, India. He has done Ph.D. from Jamia Millia Islamia, M.Tech (Software Engineering), and B.Tech (Computer Engineering) from Aligarh Muslim University. He has been given a Gold medal in M.Tech. He has also been awarded Maulana Azad National Fellowship (MANF) by UGC. He has teaching experience of more than 10 years. He has worked in the organizing committee in the International conferences ICACSE 2019 and ICACSE 2021. He has published International Journals and Conferences papers in SCI/SCOPUS indexed journals in text mining, Big Data, Recommendation systems, IoT, and many others. He has also got the best paper award in International Conference SIGMA-2018 held at NSIT, New Delhi.

Pradeep Kumar Singh is an Assistant Professor of Computer Science Department in Kamla Nehru Institute of Technology Sultanpur since September 2018 and also pursuing PhD from National Institute of Technology Durgapur. He has worked in the organizing committee in the conference of ICBIM 2016, and ICACSE 2019. In the past, he has completed his M.Tech degree from National Institute of Technology Durgapur in 2014. He has authored and co-authored about 9 international journal papers or conference proceedings. Currently, his research areas are data mining, machine learning, and data analysis.

Mohammed Shuaib is currently working as Lecturer in the Department of Computer Science, College of CS & IT Jazan University. He received the B-Tech. And M-Tech degrees in computer Engineering majoring in Software Engineering from the Aligarh Muslim University (AMU), Aligarh, India in 2010 and 2012 respectively. He has also served for three years as an Assistant Professor in the Department of Computer Engineering., Integral University, Lucknow, India. He is currently a PhD candidate in Advanced Informatics Department from Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia (UTM) Kuala Lumpur, Malaysia. His research area is focusing on Blockchain, Security, IoT, Cloud Computing and Identity Management.

Tanweer Alam is with the department of computer science, faculty of computer and information systems, Islamic University of Madinah, Saudi Arabia since 2013. He is awarded by Ph.D. (Computer Science and Engineering), M.Phil. (Computer Science), MTech (Information Technology), MCA (Computer Applications) and M.Sc. (mathematics). His area of research including Mobile Ad Hoc Network (MANET), Internet of Things, Block chain, Cloud Computing and wireless networking. He is a single author of twelve computer science textbooks. He is the member of various associations such as International Association Computer Science and Information Technology (IACSIT), International Association of Engineers, Internet Society (ISOC), Computer Science Teachers Association (CSTA), Indian Society of Technical Education (ISTE), etc.