Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASSESSING THE INFLUENCE OF MEMORY-BASED COLLABORATIVE FILTERING METHODS ON CONTEXTUAL SEGMENTS IN MULTICRITERIA RECOMMENDER SYSTEMS

[1]CHINTA VENKATA MURALI KRISHNA, [2]DAVULURI SUNITHA, [3]BOPANNA VENU GOPAL, [4]A SARVANI, [5]VELAGAPUDI SREENIVAS

[1]Associate Professor, Department of Computer Science Engineering, NRI INSTITUTE OF TECHNOLOGY, INDIA

[2] Professor, Department of Computer Science Engineering, NRI INSTITUTE OF TECHNOLOGY, India

[3] Associate Professor,Department of Computer Science Engineering, NRI INSTITUTE OF TECHNOLOGY, India

[4]Sr Assistant Professor, Department of Information Technology, LAKIREDDY BALI REDDY COLLEGE OF ENGINEERING, India

[3] Associate Professor ,Department of Computer Science Engineering, Dhanekula Institute of Engineering and Technology, India

E-mail: [1]muralikrishna_chinta2007@yahoo.co.in, [2]hod.cse@nriit.edu.in, [3]srees.boppana@gmail.com, [4]sarvani.anandarao@gmail.com, [5]velagapudisreenivas@gmail.com

ABSTRACT

Recommender Systems has grown significantly over the last two decades. Memory-based Collaborative Filtering is part of RS and is a powerful technology that has been applied in several well-established commercial applications.However, memory-based collaborative filtering fails to capture the dynamic user opinions in a detailed perceptive since it uses a two-dimensional rating approach.However, multicriteria RS dominates memory-based collaborative filtering with the inclusion of multiple contexts. In addition, significant research has been done to predict user gratification. However, recent multicriteria recommender systemsfail to avoid the significant issues of the curse of dimensionalitydue to the lower number of ratings among multiple dimensions, leading to poor predictions. This paper proposes a new prediction recommender model on multicriteria recommender systems to predict user gratification with the memory-based user and item collaborative filtering approaches used to impute the missing contextsin multicriteria RS. In addition, various regression models were applied to overall and predicted overall ratings. The results indicate that item-item collaborative filtering with Ordinary Least Squares(OLS) regression in multicriteria RS exhibits low Root Mean Squared Error(RMSE), indicating the accurate predictions of user gratification.

Keywords: *Cost Functions, Recommender Systems, Memory-Based Collaborative Filtering, Muti-Criteria Recommender Systems.*

1. INTRODUCTION

Information management has to deal with many problems caused by the vast and growing amount of information available today. Recommender systems (RS) are a popular technology that has been very beneficial in dealing with the massive growth of information [1]. News, items, music, and postscan all recommend using the algorithm's matching features and personalized profiling. Websites like Amazon, Netflix, and eBay incorporate RS to recommend products or items they might like based on their views.

Collaborative filtering is a traditional approach that provides the highest-quality recommendations with vast user information[2]. Ratings are employed in collaborative filtering recommender systems to generate suggestions. These preferences may be explicit, such as when a user chooses an item, or implicit, such as when the system determines a user's choice for a product by observing their behavior, such as by tracking purchases or mouse clicks on a computer screen[3].

In collaborative filtering, ratings are the basis for similarity measures used to find people with similar interests. Ratings are quantitative values representing how much a user likes or dislikes a particular item. According to [4], there are two ways to use collaborative filtering: memory-based and model-based. The memory-based method operates by computing the user similarities, selecting the most similar users based on the active users' neighbors, computing the similarity scores to generate predictions, and providing the top N recommendations based on the predicted value. In contrast, the model-based method uses a constructed model todescribe user behavior andforecastratings.

Recommender systems frequently use memory-based collaborative filtering algorithms due to their usability and high-quality forecasts [5]. User preferences predict with a pre-built model in Model-based approaches. In Memory-based strategies, the correlation between users or items can obtain by accessing the whole database of user-ratedthings. Memory-based recommendation algorithms can generally subdivide into user and item-based approaches [6].

Two fundamental operations require in user-based algorithms: computing similarity and prediction. During the computation process, the system attempts to identify user relationships; the target user's neighbors have the highest correlations. Then, a list of candidate items, consisting of all rated things the user still needs to purchase or obtain, is compiled. In the second step, the system predicts a user rating for each candidate item and recommends the things with the highest predicted ratings. Consequently, evaluating and ranking candidate items is crucial to the algorithm's accuracy[7].

Numerous studies have shown that user-based approaches are inapplicable when the number of users and items in a system is substantial—particularly the requirement to scan many neighbors.

As a result, a second method, known as an item-based nearest neighbor recommendation, was proposed to identify similarities between items and their nearest neighbors. An online pre-computing of a data strategy chosen for item-based offers applicable to large-scale applications. The objective is to create a matrix describing the pairwise similarity of all available items in advance. A prediction is generated at runtime for Item I and a user 'U' by identifying the most similar items to I and calculating the weighted sum of the current user ratings for these ratings in the neighborhood. The potential number of neighbors is limited to the number of items rated by the current user.

Current RS uses multicriteria attributes. The goal of multicriteria recommender systems is the same as that of single-rating recommender systems: to identify items that maximize the utility of each user. However, compared to single-rating rating systems, multicriteria rating systems contain more information about users and things that can be used effectively in making recommendations [8].Recent multicriteria RS allows users to express their opinions simply by providing multiple segments in the form of contexts. As a result, users in this digitalization era have many choices in selecting items. Due to this, multicriteria RS faces unpredictable situations where users may rate high overall ratings even though they may satisfy over a few contexts. They may rate high for a few segments and low or not rated for the remaining things, causing a curse of dimensionality and sparsity.

Sparsity is a significant concern in multicriteria RS since users are not interested in rating all the contexts.Therefore, replacing missing contexts with the available ones is a big challenge. One approach is a regression framework used by severalstandard models for multiple imputations to generate plausible missing values. In addition, several imputation methods are used to replace missing attributes. However, only some imputation methods excel in all applications [9].

This paper uses memory-based collaborative filterings of user-user(Pearson's correlation) and item-item(Adjusted cosine similarity) in multicriteria RS to impute missing contexts to avoid sparsity.Furthermore, ordinary Least Squares(OLS) with backward elimination are used to obtain the significant contexts to avoid the curse of dimensionality. As a result, predicted overall ratings can obtain with significant contexts. Finally, several machine learning regression models were applied to assess the significance of overall and predicted overall ratings.Furthermore, their performance can evaluate with cost functions using the TripAdvisor dataset.

2. RELATED WORK

Traditional CF techniques rely on user ratings as a source of suggestion input, but they cannot convey the fine-grained analysis behind users' behaviors. It is impossible to determine why the user chose these ratings, and it is impossible to identify the user's particular preferences. Multiple-criteria decision analysis is merged with RSto generate a multicriteria recommender system (MCRS). MCRS is based on multiple factors rather than one element[10].

When a user gives a product a high rating, it does not necessarily mean they are not interested in it completely. However, they may still dislike specific characteristics or aspects of the item. Therefore, when assigning overall ratings, the user places varying emphasis on different elements, influencing their final decision[11].

MCRS is a trend in RS research that improves recommendation accuracy by using user-generated reviews or ratings to represent the item's criteria[12].

3.RELATED LITERATURE:

Wasid and Ali[13] propose a multicriteria RS employing a clustering technique to expand the suggestion set by discovering comparable users within a user's cluster. Euclidean distance and Mahalanobis distance determine the closest C for each user, while the anticipated rating for an item is calculated using similar neighbors from the same cluster. The drawback of this approach is that if two data vectors share no attribute values, they may have a smaller distance than other pairs[14].

Zheng[15] developed a multicriteria, utility-based recommender system that suggests items to a user based on the utility function of each item for that user. The utility score computes with the metric cosine similarity.However, cosine similarity only considers the vectors' direction, not magnitude. Mean value differences are not fully considered in recommender systems.

Recommender systems represent the utility of an item to a user in the two-dimensional Users Items space using a single criterion rating. The recommendations are inferred explicitly or implicitly with the specification of the user's initial set of ratings. Multicriteria rating systems contain more information about users and items that can be utilized effectively in making recommendations. The general form of a rating function in a multicriteria recommender system is R: Users Items R0 x R1x... x Rk, where R0 is the overall rating, and Rk is the set of possible rating values for each criterion[16].

Similarity-based and aggregation function-based techniques are categorized as multicriteria-based recommendation methods [17]. Several solutions are proposed for the limitations of Multicriteria Collaborative Filtering(MCCF) [18]. The majority of current research focuses on enhancing precision. Several techniques, including Mahalanobis distance[19], Euclidean distance [13], and grey relational analysis[20], are proposed for computing similarities to increase the accuracy of referrals in similarity-based methods. In addition, an attempt has been made to generate more precise criterion-based predictions to improve the accuracy of forecasts in aggregation function-based MCCF systems[21]. The fuzzy Bayesian approach [22] and Matrix factorization [23] produce predictions based on predetermined criteria. The remaining researchers attempt to improve the accuracy of their forecasts by incorporating more precise aggregation functions. Support vector regression, feed-forward neural networks[24], and adaptive genetic algorithms [25] are utilized in aggregation function learning without addressing the curse of dimensionality, a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

significant worry of recommender systems[26]. Some work has been done to avoid insignificant contexts to predict overall ratings [27][28].

4. PROPOSED FRAMEWORK METHODOLOGY

Traditional CF employs a two-dimensional user-item rating matrix in which users have rated a set of system-entered items. Traditional recommender systems are widely usedbuthave limitations in recommending things in real-world scenarios. Overall rating is the primary factor in these systems to recommend. Nowadays, users seek other criteria besides overall ratings before selecting things. Now leading recommender systems incorporate multi-criterion to generate practical recommendations in multicriteria RS containing multiple contextual segments. Higher dimensionality is the greatest obstacle for these recommendation systemsfor causing sparsity and the curse of dimensionality. This article proposes a framework to avoid sparsity andthe curse of dimensionality by incorporating memory-based user-user and item-item collaborative filterings. These two methods are usedto impute the missing contexts. Then Ordinary Least Squares Regression with backward elimination is used to identify the significant contexts to avoid the curse of dimensionality. The significant contexts are used to predict the overall rating. Both overall and predicted ratings are analyzed using various machine learning models using the below steps Asian continent tourism hotels on star rating wise data collected from TripAdvisor with web scrapping. The dataset was extracted from all the major tourism cities ranked by Visa and Mastercard.

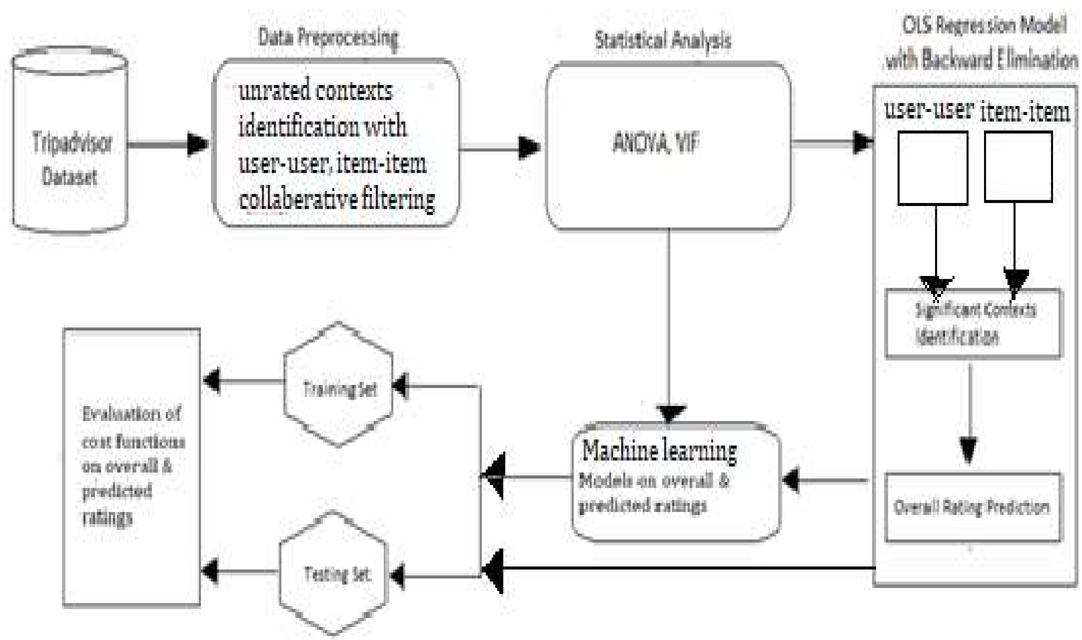

Fig 1:Framework For The Proposed Prediction Model.

1. Acquisition of data from TripAdvisor.
2. Using Pearson correlation and Adjusted Cosine similarity, identify and impute the unrated contexts with user-user and item-item collaborative filtering.
3. Identification of outcome and inputvariables.
4. Conduct the statistical analysis with Analysis of Variance(ANOVA), Variance Inflation Factor(VIF).

Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. a) Build the regression models with Ordinary Least Squares(O.L.S), Decision Tree(D.T.R), Random Forest, and Support Vector.
b) Prediction of outcome with OLS using backward elimination by eliminatinginsignificant contexts in user-user and item-item multicriteria.
6. Assessment of overall and predicted overall ratings with cost functions using a holdout cross-validation approach.

In the dataset, the contextual multicriteria ratings, Cleanliness, Location, Room, Value, Service, and Sleep-Quality are considered input variables, and the outcome variable is the overall rating. The regression models M.L.R, R.F, D.T.R, and SVR were applied between outcome and input variables.

6. RESULTS AND DISCUSSION

Table 1:ANOVA and VIF results on User-User multicriteriaCF

User-User multicriteria	Cleanliness	Location	Value	Rooms	Service	Sleep Quality
Df	1	1	1	1	1	1
Sum sq	7106	5654	7357	7160	7539	6872
Mean sq.	7106	5654	7357	7160	7539	6872
F value	10724	7292	11438	10875	11984	10093
Pr(>F)	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***
VIF	2.677532	2.052300	2.681547	2.584797	2.426462	2.609547

The above table displays the ANOVA and VIF results on User-User multicriteriaCF.All contextual segments indicate a significant impact on the overall rating, and VIF results exhibit low multicollinearity among input variables which is negligible.

Table 2: ANOVA and VIF results on Item-Item multicriteria

Item-Item multicriteria CF	Cleanliness	Location	Value	Rooms	Service	Sleep Quality
Df	1	1	1	1	1	1
Sum sq	4519	3165	4894	4579	7416	4232
Mean sq.	4519	3165	4894	4579	7416	4232
F value	4046	2634	4477	4114	7927	3730
Pr(>F)	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***	<2e-16 ***
VIF	1.468878	1.280293	1.439917	1.446691	1.594897	1.384570

The above table displays the ANOVA and VIF results on Item-Item multicriteriaCF.All contextual segmentsindicate a significant impact on the overall rating, andVIF results exhibit low multicollinearity among input variables which is negligible.

Table 3: Accuracy results of User-User Multicriteria CF (overall rating)

User-User Multicriteira CF	Accuracy	MAE	MSE	RMSE
MLR	0.639	0.449	0.434	0.659
RF	0.822	0.297	0.213	0.461
DTR	0.872	0.218	0.154	0.39
SVR	0.690	0.357	0.368	0.606

The above table indicates the accuracy and different cost function results obtained on various machine learning approaches MLR, RF, DTR, and SVR. For example, the Root Mean Squared Error ranges from 0.39 to 0.65, which is low for DTR and high for MLR.

Table 4: Accuracy results of Item-Item Multicriteria CF (overall rating)

Item-Item Multi criteria CF	Accuracy	MAE	MSE	RMSE
MLR	0.458	0.618	0.757	0.870
RF	0.758	0.395	0.338	0.581
DTR	0.825	0.289	0.245	0.495
SVR	0.551	0.512	0.631	0.794

The above table indicates the accuracy and different cost function results obtained on various machine learning approaches MLR, RF, DTR, and SVR. For example, the Root Mean Squared Error ranges from 0.49 to 0.87, which is low for DTR and high for MLR

Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

User-User Multi criteria CF	Accuracy	MAE	MSE	RMSE
MLR	1.0	8.814	1.596	1.263
RF	0.999	0.007	0.000	0.024
DTR	1.0	6.408	2.577	1.605
SVR	0.994404	0.056	0.004	0.065

The above table indicates the accuracy and different cost function results obtained on various machine learning approaches MLR, RF, DTR, and SVR on predicted overall rating (outcome variable) received with OLS after removing insignificant contexts. The Root Mean Squared Error for all of these models is negligible.

Table 5: Accuracy results of Item-Item Multicriteria CF (Predicted overall rating)

Item-Item Multi criteira CF	Accuracy	MAE	MSE	RMSE
MLR	1.0	6.591	9.934	9.967
RF	0.998687	0.010491	0.000830	0.028811
DTR	1.0	6.27e-4	1.19e-7	1.094
SVR	0.991	0.067	0.005	0.073
SVR	0.991	0.067	0.005	0.073

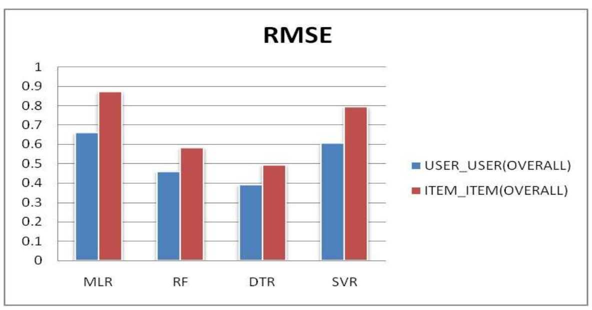

Fig 2: RMSE graph results on User-User and Item-Item multicriteria CF.

The above figure is the comparison metrics of RMSE over user and item multicriteria collaborative filterings on the overall rating.

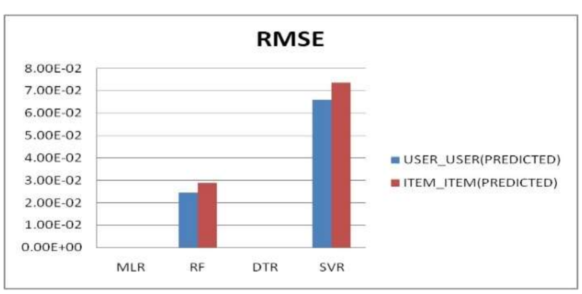

Fig 3: RMSE graph results on User-User and Item-Item multicriteria CF.

The above figure is the comparison metrics of RMSE over user and item multicriteria collaborative filterings on the predicted overall rating after removing the insignificant contexts to avoid the curse of dimensionality. The better results of item-item indicate low RMSE than user-user multicriteria collaborative filterings.

Before building the models initially ANOVA test conducted between overall rating with the other independent variables to find the significant impact.

The results indicate that all the independent variables have significant impact on overall rating. Multicollinearity also checked here to find the association among independent variables. The results indicate that no significance.

Machine Learning regression models of multiple regression, Random Forest, Decision Tree, Support Vector Machine are built upon the dependent variables overall and predicted ratings.

The accuracy results are evaluated with the cost function RMSE between multicriteria recommender systems of user-user and item-item collaborative filtering methods. Both these collaborative filterings works successfully to replace the unrated contexts. A major aim of this paper is to avoid the curse of dimensionality, which is a regular problem faced by current recommender systems. To provide solution for these only significant contexts are considered to predict the rating, and with this predicted rating as the outcome variable several machine learning models like multiple regression, Random Forest, Decision Tree, Support Vector Machine regressions are applied.

5. CONCLUSION

Modern Multicriteria RS suffers from sparsity and the curse of dimensionality since users are not interested in rating all the contextual segments due to higher dimensions; also, they rate high gratification even satisfied with few contexts. As a result, it affects overall gratification. However, memory-based collaborative filterings cannot exhibit the curse of dimensionality with a single criterion. Therefore, this paper incorporates the memory-based collaborative filtering approaches in multicriteria RS.First, unrated contexts are replaced with user-user and item-item collaborative filterings. Next, ordinary Least Squares with backward elimination identify the significant contexts which are used to predict the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

overall rating. Then several regression techniques are applied to both the target and the predicted outcomes.The results signify that memory-based collaborative filterings combined with OLS outperform the curse of dimensionality and sparsity. Furthermore, the predicted overall rating obtained with preprocessing of item-item collaborative filtering and OLS regression tested on other machine learning models signifies the lowest error rate of RMSE than with user-user collaborative filtering.

FUTURE SCOPE

This study focuses on avoiding the curse of dimensionality in multicriteria collaborative filterings. In the future, regularization approaches will apply to the overall and predicted overall ratings without removing the insignificant contexts of these methods.

REFERENCES

[1]. D. Jannach, M. Zanker, A. Felfernig, G. Friedrich,"Recommender Systems, An Introduction," Cambridge University Press, (2011).

[2]. Shen, J., Deng, C., Gao, X., 2016,"Attraction recommendation: Towards personalized tourism viacollectiveintelligence,"Neurocomputing173 (P3),789–798. http://dx.doi.org/10.1016/j.neucom.2015.08.030.

[3]. Daniel Valcarce et al.," Collaborative filtering embeddings for memory-based recommender systems," Engineering Applications of Artificial Intelligence 85 (2019) 347–356.

[4]. G. Tseng and W. Lee, "An enhanced memory-based collaborative filtering approach for a context-aware recommendation," Proceedings of the World Congress on Engineering 2015 Vol I WCE 2, London, U.K.

[5]. R. Sharma, D. Gopalani, and Y. Meena," Collaborative Filtering-Based Recommender System: Approaches and Research Challenges," 3rd International Conference on Computational Intelligence and Communication Technology (CICT), India.

[6]. Adomavicius G, Tuzhilin A, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE Transactions on Knowledge and Data Engineering. 2005;17(6):734–749.

[7]. Al-Bashiri H, Abdulgabber MA, Romli A, Salehudin N. "A Developed Collaborative Filtering Similarity Method to Improve the Accuracy of Recommendations under Data Sparsity." International Journal of Advanced Computer Science and Applications (IJACSA). 2018; 9(4):135–42. http://dx.doi.org/10.14569/IJACSA.2018.090423

[8]. FerdaousHdioud et al., "Multi-Criteria Recommender Systems based on MultiAttribute Decision Making”,iiWAS2013, 2-4 December, 2013, Vienna, Austria.

[9]. Chhabra G, Vashisht V, Ranjan J,"A comparison of multiple imputation methods for data with missing values,"Indian J Sci Technol.2017;10(19):1–7. https://doi.org/10.17485/ijst/2017/v10i19/110646.

[10]. Krishna, C.V.M., Rao, G.A. & Anuradha, S, "Analysing the impact of contextual segments on the overall rating in multicriteria recommender systems," J Big Data 10, 16 (2023). https://doi.org/10.1186/s40537-023-00690-y.

[11]. C. Yang, X. Yu, Y. Liu, Y. Nie, and Y. Wang, "Collaborative filtering with weighted opinion aspects," Neurocomputing, vol. 210, pp. 185–196, Oct. 2016.

[12]. G. Adomavicius, N. Manouselis, and Y. Kwon, "Multicriteria recommender systems in Recommender Systems Handbook," F. Ricci, Ed. Boston, MA, U.S.A.: Springer, 2011, pp. 769–803.

[13]. M. Wasid and R. Ali, "An improved recommender system based on multicriteria clustering approach," Procedia Comput. Sci., vol. 131, pp. 93–101, Jan. 2018.

[14]. Legendre P, Legendre L. Numerical ecology. Elsevier; 2012.

[15]. Y. Zheng, "Utility-based multicriteria recommender systems," in Proc. 34th ACM/SIGAPP Symp. Appl. Comput., 2019, pp. 2529–2531.

[16]. FerdaousHdioud et al., "Multicriteria Recommender Systems based on Multi-Attribute Decision Making," Proceedings of International Conference on Information Integration and Web-based Applications & Services, DOI:10.1145/2539150.2539176.

[17]. Adomavicius G, Kwon Y. "New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems

2007"; 22 (3): 48-55. doi: 10.1109/MIS.2007.58

[18]. T. Wei-Guang and L. Hsin-Hsien, "Collaborative Recommendation with Multicriteria Ratings," 2007.

[19]. Yalcin E, Bilge A,"Binary multicriteria collaborative filtering". Turkish Journal of Electrical Engineering & Computer Sciences 2020; 28 (6): 3419-3437. Doi: 10.3906/elk-2004-184.

[20]. Wasid M, Ali R."Multicriteria clustering-based recommendation using Mahalanobis distance," International Journal of Reasoning-based Intelligent Systems, 2020; 12 (2): 96-105. Doi: 10.1504/IJRIS.2020.106803.

[21]. Hu Y-C, Chiu Y-J, Liao Y-L, Li Q, "A fuzzy similarity measure for collaborative filtering using nonadditive grey relational analysis." The Journal of Grey Systems 2015; 27 (2): 93-103.

[22]. Batmaz Z, Kaleli C. AE-MCCF: "An autoencoder-based multicriteria recommendation algorithm," Arabian Journal for Science and Engineering 2019; 44 (11): 9235-9247. Doi: 10.1007/s13369-019-03946-z.

[23]. Kant V, Jhalani T, Dwivedi," Enhanced multicriteria recommender system based on fuzzy bayesian approach," Multimedia Tools and Applications 2018; 77 (10): 12935-12953. Doi: 10.1007/s11042-017-4924-2.

[24]. Majumder GS, Dwivedi P, Kant V,"Matrix factorization and regression-based approach for multicriteria recommender system," In Proceedings of International Conference on Information and Communication Technology for Intelligent Systems; Ahmedabad, India; 2017. pp. 103-110.

[25]. Hassan M, Hamada M. "A neural networks approach for improving the accuracy of multicriteria recommender systems," Applied Sciences 2017; 7 (9): 868. Doi: 10.3390/app7090868.

[26]. Kaur G, Ratnoo S, "Adaptive genetic algorithm for feature weighting in multicriteria recommender systems,"Pertanika Journal of Science & Technology 2019; 27 (1): 306-314.

[27]. ChintaVenkataMuraliKrishnaetal.2022ECSTrans,2022,107(1), PP19813-19825 "Gratification in Indian Tourism Domain on Hotel Classes and Trip-Types."

[28]. Krishna, C.V.M., Appa Rao, G., Kadaru, B.B., AnuRadha, S. (2022). Impact of Contextual Segments in the Prediction of Overall User Gratification in Asian and European Continental Hotel Tourism Sector. In: Kumar, A., Mozar, S. (eds) ICCCE 2021. Lecture Notes in Electrical Engineering, vol 828. Springer, Singapore Singapore https://doi.org/10.1007/978-981-16-7985-8_11921.