Enhancing User Profile by Combining User and Item Based Collaborative Filtering

K. Reka, T.N. Ravi

Abstract: Product selection has been made more complex and tedious due to the huge number of selections available online. Personalized recommendations can effectively solve this problem by presenting more related products to users. Such recommendations can enable huge profits for organization and less stressful for users. This paper presents an effective model that can handle cold start problem to provide more reliable predictions. User and item based collaborative filtering approaches are combined to identify highly related items. These items are then used to train the proposed bagged ensemble model to provide the final recommendations. The performance of the proposed model was evaluated through experiments and the results obtained were compared with that of the state of art models. The results indicate effective performances in terms of MAE and RMSE.

Index Terms: Recommendation systems; ensemble model; voting; collaborative filtering; user based collaborative filtering; item based collaborative filtering

---

1. INTRODUCTION

Recommendation systems are used mainly for the process of information filtering, which in turn aids in predicting items for the users. This helps in providing personalized suggestions for the users [1]. The increased dependence and usage of e-commerce has made recommendation systems an indispensable feature. Selection fatigue associated with the too many products is another main reason for usage of recommendation system models. Recommendation systems aim to eliminate such issues, making the decision process easier for the users. Further, competitions in the e-commerce industry has made recommendations a much necessary feature for any organizational model. Effective suggestions and recommendations not only increase the Quality of Service (QoS) for the users, but also increase the Quality of Experience (QoE) for the users [2]. Recommendation systems can be implicit or explicit. Implicit systems aim to observe the user's behavior and perform predictions based on the behavioral nature of the users [3]. The explicit systems perform recommendations based on the details of the users. Personal and product based details of users are explicitly collected and these details are used as the base to train models for the recommendation system. Recommendation systems can be divided into two major categories; content based systems and collaborative systems [4]. The implicit systems come under the category of content based systems. These systems analyze the movie descriptions to identify the movies that are of interest to the users. They are simple models, however capturing the current complexities is much difficult for such models. Further, these models are static and hence are not sensitive to the change in interest of the users. The introduction of Web 2.0 has resulted in an enhanced model that uses additional details of users to perform recommendations. These models are termed as collaborative filtering models. They form explicit systems. Collaborative filtering models are based on the similarity or the dissimilarity metric. However, the model tends to suffer from cold start problem [5]. Major applications of recommendation models include movie recommendations, music recommendations, video recommendations, product recommendations for e-commerce and several other recommendation models. The rest of this paper is structured as follows; section 2 provides the related works, section 3 presents the proposed model, section 4 presents the results and discusses them and section 5 concludes the work.

- *K.Reka, Research Scholar in P.G & Research Department of Computer Science, Periyar E.V.R College(Autonomous),Trichy,TamilNadu, India, E-mail:rekaccw@gmail.com*
- *T.N.Ravi, Assistant Professor in P.G & Research Department of Computer Science, Periyar E.V.R College (Autonomous),Trichy,TamilNadu, India, E-mail:proftnravi@gmail.com*

2 RELATED WORKS

Increase in e-commerce applications has resulted in recommendation systems becoming a popular choice for organizations and researchers alike. This section deals with recent and most prominent contributions in this domain.A data clustering based recommendation system model was proposed by Katarya et al. [6]. A cuckoo search based K-Means clustering model has been proposed for the recommendation process. This is a collaborative filtering based model. Usage of metaheuristic models effectively results in faster processing. A Hidden Markov Model (HMM) for movie recommendation was proposed by Trabelsi et al. [7]. This method is particularly based on movie recommendation. Recommendations are provided by modifying the HMM model to incorporate user's rating profiles during the analysis phase. A genre correlation based collaborative filtering model was proposed by Choi et al. [8]. This is similar to item based collaborative filtering approach, where genre details are used as the base for the recommendation process. The authors have enhanced the model to provide recommendations even with profiles containing very low data. Several other methods exist to avoid the cold start problem; such as; category correlation model by Choi et al. [9], associative retrieval techniques by Huang et al. [10] and information diffusion approach by Ishikawa et al [11]. A data-sparsity handling technique was proposed by Li et al. [12]. This model is based on handling two major issues in the recommendation system; handling data sparsity and identifying and handling concept drift. This method proposes a hybrid algorithm, whichcombines movie features and user's ratings to provide better recommendations. It generates a user interest vector, which combines the attributes associated with the movies and the user's ratings provided for those movies. The model also identifies the long term and short term interests of the user, using which concept drift is handled effectively. Metaheuristics is another major model used for implementing a recommendation system. A metaheuristics based collaborative filtering model was proposed by Chen et al. [13]. This model proposes new formulae for calculating the affinity between an antigen and an antibody, and also between an antigen and an immune network. The model also proposes a Pearson correlation based formula for identifying similarity between items. A model dealing with user experience based movie recommendation was proposed by Lee et al. [14]. The model discusses the enhancements that can be achieved by incorporating user satisfaction and the intent to use such a system. A multiattribute based recommendation system was proposed by Son et al. [15]. The model utilizes multiple attributes for the recommendation process to effectively perform item recommendations to users.In this system items in the network are linked based on their similarity levels. This is used to formulate the network. Other similar methods include distance measure basedmodels by Choi et al. [16], Lops et al. [17] and Lew et al. [18]. A computational intelligence based collaborative recommendation system was proposed by Wang et al. [19]. This is a model based movie recommendation system that uses K-Means Clustering model and Genetic Algorithm to identify the best recommendations. The model also employs PCA as a data reduction technique to reduce the data size for faster recommendations. Other similar methods include models by Li et al. [20], Kim et al. [21],Kohrs et al. [22] and Wei et al. [23].

3 OUR APPROACH

Recommendation systems require quality and sufficient base data to provide effective predictions. Cold-Start is one of the major set-backs faced by recommendation systems. Cold-Starts occur due to unavailability of sufficient data for the recommendation process. This work aims to solve the cold-start problem by building an enhanced profile of the user.

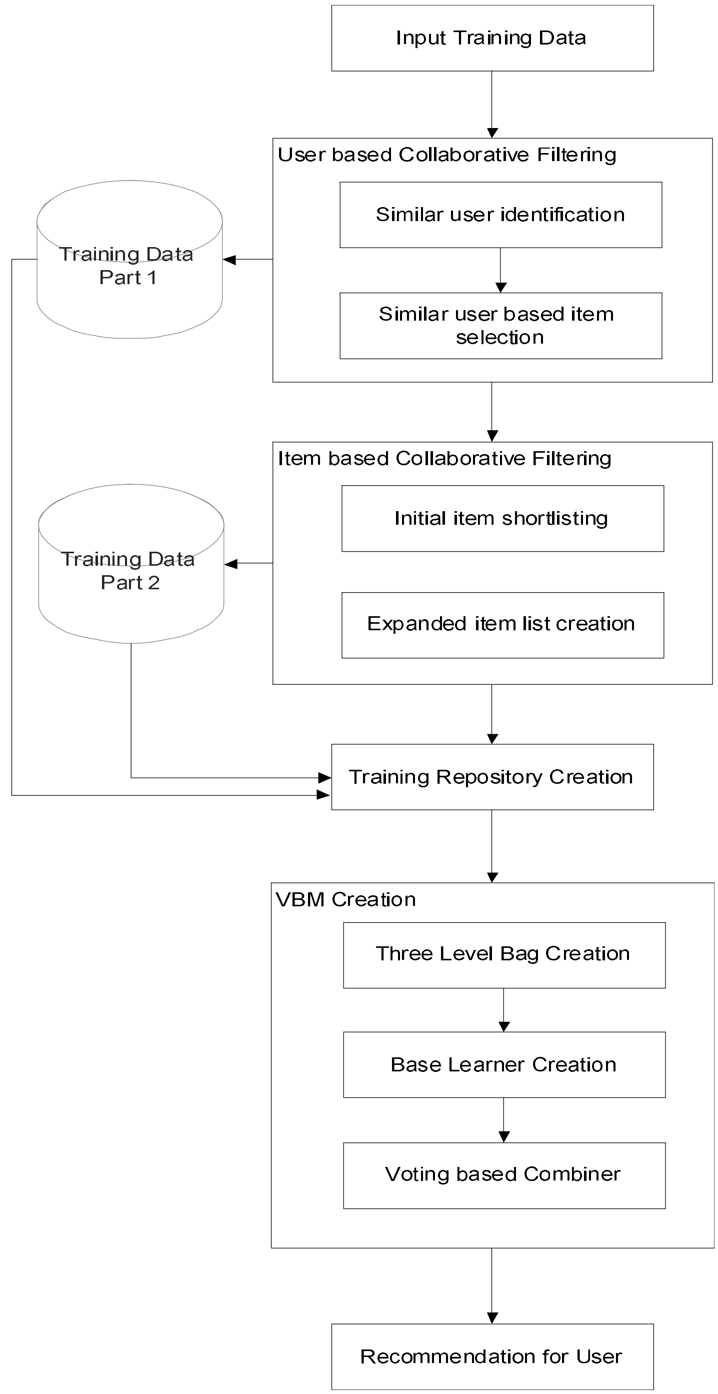

Figure 1:*Architectural view of the proposed model*

The profile enhancement is performed by using user based collaborative filtering and item based collaborative filtering. The enhanced profile is used by the recommendation system for the prediction process. This model is best suited and can be used to overcome the cold-start problem which occurs for a new user and also for a new item. The proposed model is composed of four major phases; user based collaborative filtering, item based collaborative filtering, training repository creation and recommendation. The overall architecture of the proposed model is shown in figure 1.

3.1. User based Collaborative Filtering
User based collaborative filtering is the process of recommending items based on the user's personal profile. Recommendation systems operate on per-user basis. Items are selected based on users similar to the current user.

3.1.1 User Profile based Similar User Identification
Recommendations are usually based on a single user. User's personal profile is identified and other users similar to the current user are shortlisted for the process. Profile analysis and attribute selection is performed prior to this process. The process of attribute shortlisting is performed by the domain expert. Similarities are identified by constructing the correlation matrix based on the shortlisted attributes. Users with correlation greater than 90% with the current user are selected as similar users.

3.1.2 Similar User based Item Selection
This phase proceeds with identifying most preferred items of similar users. Highly rated item for every similar user is selected and aggregated to form the secondary profile of the current user. Secondary profile comprises the items that have not been ranked, but are highly likely to be ranked high by the current user.

3.2 Item based Collaborative Filtering
Item based collaborative filtering is the process of recommending items to the user based on their item profile. Items highly ranked by the user is used for analysis. Other items that are similar to the highly ranked items are considered for analysis. This process is also performed on a per user basis. The operation is performed in two phases: initial item shortlisting and expanded item list creation.

3.2.1 Initial Item Shortlisting
The initial process is to identify the items that have high preference, from the current user. This is performed by selecting the items that are provided with highest ratings, by the selected user. Rating levels are defined by the domain expert and is based on the domain requirements and the size of the current user's profile. In case of a small profile, medium to high ratings are considered and in case of users with larger profiles, only the high preferences are selected. The selected items are labelled as the primary preferences of the current user. Hence these terms form the primary profile of the user.

3.2.2 Expanded Item List Creation
The expanded item list is created by identifying the correlation matrix representing the correlation levels of each item with every other item. Every item in the primary profile is taken and all items that represent high correlation (>90%) with these items are selected. This might even result in selection of duplicates. Hence distinct items are chosen and added to the secondary profile of the user.

3.3 Training Repository Creation
The secondary profiles created in the process of user based collaborative filtering and item based collaborative filtering are combined together to form the aggregated secondary profile of the user. The aggregated secondary profile, along with the primary profile of the user forms the training data repository. Training data plays a vital role in defining the performance of a classifier. Sufficient and qualitative training data results in an effective classifier. The size of the training data is adjusted based on the size of the user's profile. Most of the data are accepted in case of user with a smaller profile, while specific selections are entertained for a user with a larger profile.

3.4 Voting based Bagged Model (VBM) for Recommendation
Recommendations are performed by the voting based bagging model, VBM. This model uses bagging as the major operational architecture. Bagging is the process of using multiple base learners with varied input data to provide multiple predictions. These predictions are finally aggregated to provide the final results. The major advantage of using a bagging model is that varied data are passed to the multiple base learners. Hence every base learner identifies different signatures. Hence all types of patterns can be effectively identified. The proposed VBM architecture is constructed in three major phases; the bag creation phase, the base learner training phase and the voting phase.Bag creation is the first phase of the process. The proposed model creates multiple base learners for the training phase. Every base learner is to be provided data for training. The bag creation phase performs the process of data selection. Data is selected such that every base learner has overlaps with other base learners. This work modifies the bag creation process by performing data selections independently based on the primary and the secondary user profiles. Three bag sets are created. The first bag set is created by selecting data only from the primary profile. The second bag set is created by selecting data only from the secondary profile. The final set is created by selecting equal parts of data from the primary and the secondary profiles. The final set contains mixed data. Every set can contain multiple bags. This depends on the bag size and the data size. The next phase is the process of creating the base learner for training. This work treats the recommendation system as a one class classification model. Predictions are usually performed on a large number of items. It becomes computationally less intensive when movies are selected directly, rather than filtering after rating. Since training data is composed only of the movies with high ratings, all training data correspond to one class only. Hence one-class SVM is used as the base learner for the prediction process. Multiple instances of the one-class SVM model is created. The number of instances depends on the number of bags created in the previous phase. Each created bag is provided to a single base learner and the training process is performed. The training process results in creation of rules that corresponds to the highest rated items. All the items that are not contained in the primary profile are used for the prediction process. Each of the item is passed to all the base learners and the predictions are performed. Every base learner provides its own predictions. Multiple predictions are generated for each item. Hence voting is used to combine the generated predictions to a single final prediction. The highest voted class for each instance is identified as the final prediction. Items that are classified as positive are passed as recommendation to the user.

4. RESULTS AND DISCUSSION
Experiments were performed by implementing the proposed architecture using Python. Movie Lens dataset is used as the base for the recommendation process. Comparisons were performed with weighted strategy based model (SW I, MLR and CM II) proposed by Fremal et al. [24] and K-Means

Cuckoo Search based model proposed by Katarya et al. [25]. Comparisons were based on Mean Absolute Error (MAE) and Root Mean-Squared Error (RMSE) values. A comparison of the MAE values in shown in figure2. It could be observed that the proposed VBM model exhibits MAE levels of ~0.7. The lowest MAE is exhibited by the K Means Cuckoo model, followed by the proposed VBM model and then by SW I, MLR and CM II. Although the proposed model does not exhibit the best MAE values, variation level when compared with the best performer was observed to be very low.

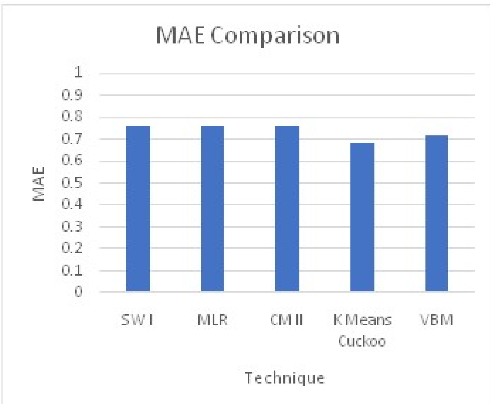

Figure2: *Comparison of MAE levels of the VBM Model*

A comparison of the RMSE levels is shown in figure3. It could be observed that the proposed VBM model exhibits the lowest RMSE levels when compared with the existing models. Further, the RMSE variations were also observed to be quite high, hence making the proposed model an effective model for the process of recommendatio

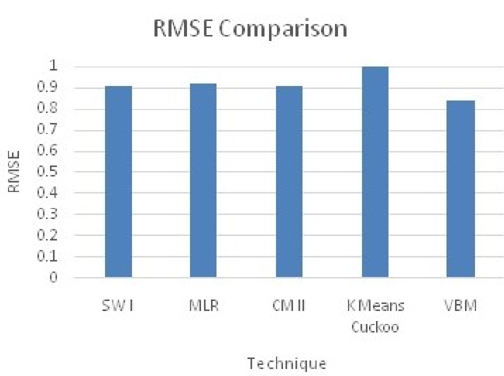

Figure 3*: Comparison of RMSE levels of the VBM Model*

A tabulated view of the results is shown in table 1. The best results are shown in bold. It could be observed that the proposed model exhibits the best RMSE levels. Although the MAE levels were found to be slightly higher, the very low RMSE levels indicate effective performances.

Table 1: Performance Analysis of VBM in Terms of MAE and RMSE

Model	MAE	RMSE
SW I	0.7616	0.9096
MLR	0.7611	0.9212
CM II	0.7615	0.9086
K Means Cuckoo	0.6842	1.231
VBM	0.72	0.84

4. CONCLUSION

Recommendations system plays a vital role in today's competitive industry. This work presents an effective model that can be applied to any customer based organization for providing item recommendations. Bagging based model is used for the prediction process. This effectively enables identification of varied decision criterion. Experiments indicate that the proposed model exhibits effective results in terms of MAE and RMSE. The major advantage of the proposed VBE model is that the training data can be limited according to the computational capacity. This results in low training time and also reduces the complexity levels to a large extent. Further, the predicted results correspond to highly associated items, hence rating based filtering is not necessary. Although the RMSE levels are low, MAE levels were found to be slightly higher. Future enhancements will be directed towards reducing the MAE levels.

REFERENCES

[1] J.S. Bridle, "Probabilistic Interpretation of Feedforward Classification Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: a survey. Decis Support Syst 2015;74:12–32.
[2] Chen L, Chen G, Wang F. Recommender systems based on user reviews: the state of the art. User Model. User-Adapt Interact 2015;25:99–154.
[3] Katarya R, Verma OP. Recent developments in affective recommender systems. Phys A Stat Mech Appl 2016;461:182–90.
[4] Fahad a, Alshatri N, Tari Z, Alamri a, Khalil I, Zomaya a, et al. A survey of clustering algorithms for big data: taxonomy & empirical analysis. IEEE Trans Emerg Top Comput 2014;2. 1–1.
[5] Hartigan Ja, Wong Ma. A K-means clustering algorithm. J R Stat Soc 1979;28:100–8.
[6] Katarya, R., & Verma, O. P. (2017). An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal, 18(2), 105-112.
[7] Trabelsi, C., & Yahia, S. B. (2016). Harnessing the Potential of HMM for Movie Rating Recommendation. Procedia Computer Science, 96, 1543-1550.
[8] Choi, S. M., Ko, S. K., & Han, Y. S. (2012). A movie recommendation algorithm based on genre correlations. Expert Systems with Applications, 39(9), 8079-8085.
[9] Choi, S. M., & Han, Y. S. (2010). A content recommendation system based on category correlations. In: The fifth international multi-conference on computing in the global information technology (pp. 1257–1260).

[10] Huang, Z., Chen, H., & Zeng, D. D. (2004). Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Trans. Inf. Syst., 22, 116–142.
[11] Ishikawa, M., Géczy, P., Izumi, N., Morita, T., & Yamaguchi, T. (2007). Information diffusion approach to cold-start problem. In: Web intelligence/IAT workshops (pp.129–132).
[12] Li, J., Xu, W., Wan, W., & Sun, J. (2018). Movie recommendation based on bridging movie feature and user interest. Journal of computational science, 26, 128-134.
[13] Chen, M. H., Teng, C. H., & Chang, P. C. (2015). Applying artificial immune systems to collaborative filtering for movie recommendation. Advanced Engineering Informatics, 29(4), 830-839.
[14] Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies, 103, 95-105.
[15] Son, J., & Kim, S. B. (2017). Content-based filtering for recommendation systems using multiattribute networks. Expert Systems with Applications, 89, 404-412.
[16] Choi, S. S. , Cha, S. H. , &Tappert, C. C. (2010). A survey of binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics, 8 (1), 43–48 .
[17] Lops, P. , De Gemmis, M. , & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. Recommender systems handbook . Springer US
[18] Lew, M. S. ,Sebe, N. , Djeraba, C. , & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multime- dia Computing, Communications, and Applications (TOMM), 2 (1), 1–19 .
[19] Wang, Z., Yu, X., Feng, N., & Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence. Journal of Visual Languages & Computing, 25(6), 667-675.
[20] Li, Q., & Kim, B. M. (2003, October). Clustering approach for hybrid recommender system. In Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003) (pp. 33-38). IEEE.
[21] Kim, K. J., &Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. Expert systems with applications, 34(2), 1200-1209.
[22] Merialdo, A. K. B. (1999). Clustering for collaborative filtering applications. Intelligent Image Processing, Data Analysis & Information Retrieval, 3, 199.
[23] Wei, S., Zheng, X., Chen, D., & Chen, C. (2016). A hybrid approach for movie recommendation via tags and ratings. Electronic Commerce Research and Applications, 18, 83-94.
[24] Frémal, S. and Lecron, F., (2017). Weighting strategies for a recommender system using item clustering based on genres. Expert Systems with Applications, vol.77, pp.105-113.
[25] Katarya, R. and Verma, O.P., (2016). An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal.