DOI: https://doi.org/10.31341/jios.44.1.7
Open Access
UDC 004.62:004.891:791.43
Original Scientific Paper

A Novel Rule based Data Mining Approach towards Movie Recommender System

Mugdha Sharma *mugdha.sharma145@gmail.com*
Research Scholar, Amity School of Engineering &Technology, Amity University, Noida, India

Laxmi Ahuja *lahuja@amity.edu*
Professor, Amity Institute of Information Technology, Amity University, Noida, India

Vinay Kumar *vinay5861@gmail.com*
Associate Professor, Vivekananda Institute of Professional Studies, Guru Gobind Singh Indraprastha University, Delhi, India

Abstract

The proposed research work is an effort to provide accurate movie recommendations to a group of users with the help of a rule-based content-based group recommender system. The whole approach is categorized into 2 phases. In phase 1, a rule- based approach has been proposed which considers the users' viewing history to provide the Rule Base for every individual user. In phase 2, a novel group recommendation system has been proposed which considers the ratings of the movies as per the rule base generated in phase 1. Phase 2 also considers the weightage of every individual member of the group to provide the accurate movie recommendation to that particular group of users. The results of experimental setup also establish the fact that the proposed system provides more accurate outcomes in terms of precision and recall over other rule learning algorithms such as C4.5.

Keywords: Data Mining, Movie Recommender System, Group Preferences, MovieLens, Classification Rules, Rule Base.

1. Introduction

In today's digital world, one of the major day-to-day challenges is to find the appropriate information from the vast amount of data that is available online. There are various personalization techniques available in the market to overcome this problem; one aspect of personalization is recommendation system which is explained in detail by Moses and Babu [1]. The primary objective of a recommender

system is to suggest the correct item to a user based upon description of item that matches with the profile of the user's requirements and interests. Recommender systems are generally categorized into two categories namely personal recommendation system and group recommendation system. Personal recommender systems provide the recommendations based on the interest of an individual user whereas a group recommender system provides the recommendations which can satisfy the need of all the group members.

Finding the most relevant piece of information from the various sources that are available online is a difficult task. Similar is the case with recommendations related to movie and entertainment industry and therefore, people need more efficient and intelligent recommender system for this category. Any general movie recommender system can recommend the movies to the user in two possible ways: one by providing a general/ common list of movies to a single user and another, by providing the most relevant movies to a group of users according to user's viewing history and group preferences. This paper focuses on the second part only because when a user search for any particular movie recommendation, they usually have their own preferences as per the group with whom they are going to watch the movie.

So, a novel content-based group recommendation algorithm has been proposed in this paper which uses the Decision Rule Set Learning (D.R.S.L) Algorithm and generates the Rule Base for an individual user. Then final movie recommendation is provided to the whole group of users based on that Rule Base and assigned weightage of each group member using Repeat Group Rule Approach (R.G.R.A). The whole idea of the proposed algorithm revolves around the fact that - for a particular user, there is a different rule base which holds peculiar relevance. Processing of available information through assigning weightage and repeat group rule-based approach will render the more relevant & suitable recommendations for a group of users.

The increasing demand in developing new recommender systems, challenges in implementing the efficient recommendation algorithms and to evaluate the potential of rule base in improving the accuracy of group recommender systems motivated the authors to write this paper. The research work is constructed as follows: next section explains the recent work done in the area of recommender system and various other methodologies like group recommendation system. Further the architecture of the proposed approach has been elaborated. The implementation and experimental evaluation of the proposed recommender system has been covered in the next section whereas last section concludes the outcomes and results of the research work.

2. Related Work

Lu et al. [2] mentioned in their work that recommender systems are being utilized in numerous applications like music, online shopping sites, news, study material, travel & tourism etc. Few authors [3][4] explained the recommender systems that are based on user as well as product reviews and analyzed that recommender systems typically

work on user's preferences which are pertained by every individual in context to various items or products. Presulli [5] stated that quality and efficiency of the work suffers due to lack of regular flow of complete information and consistent demonstration of the requisite data. So, he proposed an approach to resolve that problem by collaboratively acquiring the necessary key factors related to work.

Recommender system employs various techniques and several researchers have explained all the techniques in detail in their research work. Collaborative filtering is a type of recommendation technique that measures the interest of a user by using preferences of other related users. Zhang et al. [6] explained the main idea behind this technology that if two users can be considered as having the same tastes and interests, then it can be assumed that both the users will have the same opinion about any new item. Other than collaborative filtering, recommendation systems can also be built with the help of content of items, and a matching profile to those items. This technique is named as content-based filtering approach. Hyung et al. [7] mentioned that the content-based recommendation approach primarily suggests the same kind of items that the users have preferred in the past after matching an individual's profile with content of the items.

There are demographic recommender systems also that recommends the items based on demographics (language, country) of users as explained by Chen and He [8]. This technique provides more personalized recommendations because much detailed information about the user is involved in determining the recommendations. Apart from this, there are knowledge-based recommendation systems that make use of possible user-item knowledge displaying how a peculiar item meets user demand to develop or predict the recommendations [9].

The main objective of a group recommender system is to provide the recommendations as per the satisfaction of all the group members. But the main challenge here is - how to cater the interests and preferences of various different members of the group. Whereas in case of personal recommender systems, it is easy to measure the user tendency and item tendency and finally predict the ratings of the items those are unrated [10]. There are two further algorithms under collaborative filtering - Memory-based and Model-based algorithms. Memory-based technique has been used in [11] to improve the results of collaborative filtering algorithm. In this research paper, authors discovered the possible similarity relationships between users or items with the help of user-item interaction.

In case of model-based collaborative filtering algorithms, various authors have explored different models such as latent factor and matrix factorization where additional visual features have been used to improve the movie recommendations [12]. When it comes to Clustering-based recommendation methods, a novel multi-type clustering approach [13] was proposed which improves the accuracy of recommendations with the help of trust-based user clustering. Another collaborating filtering approach put forward in [14] explores the recommendations on signed social rating networks with the help of Bayesian probabilistic modeling approach. However, there is one issue with the collaborative filtering methods that it considers

all items and users as individual entities, but recommendations are provided without considering the specifications of those individual items or users. Various content-based recommendation methods have been deployed which provides recommendations after exploring all the details about an item, like actors or genre of a movie or about a user, like demographic details [15]. There are other content-based recommendation systems also which explores the various domains like news article recommendations [16] which provide proper reasons and explanations for each and every single recommendation they make.

The main focus of this research paper is on group recommendations since it is a less explored segment in comparison to other recommendation methods which have been discussed till now. Currently, group recommendations are provided by taking individual ratings about an item and then combining the recommendations of all the members of the group, rather than investigating the relationships between the group members in detail. And that's why these recommendation systems face the problem of data sparsity. A group recommendation system based on random walk and restart method [17] is a novel group recommendation approach. At an initial stage, it detects the various information in detail with respect to users, items, and groups, and then develop a tripartite graph representing the relationship amongst the three entities. Then it provides the recommendations to the groups based on relevance degrees of the graph. TV Recommendation and personalization Systems [18] is a web-based group recommendation system which recommends the TV program for more than one user by merging user profiles with the help of total distance minimization technique. Clustering methods can be clubbed within the group recommendation systems to detect and form a group with the help of various group modeling strategies. [19] Explains that the performance of a group recommendation system is strongly affected by the group modeling strategy that is used to build that group.

In this paper, a novel approach has been proposed for a group recommendation system which uses the Decision Rule Set Learning Algorithm and generates the Rule Base for an individual user. Then final movie recommendation is provided to the whole group of users based on that Rule Base and assigned weightage of each group member using Repeat Group Rule Approach. And that's why the proposed approach is better than current group recommendation systems wherein items are recommended only on the basis of combination of individual preferences of the group members or by combining the group's rating of similar items with the help of collaborative filtering approach.

3. Proposed Recommender System Design

The primary objective of the proposed work is to develop an effective movie recommender system that is capable to provide movie recommendations to an individual user as well as group of users. The whole proposed approach revolves around generating the rules and Repeat Group Rule Approach (RGRA) to provide group recommendations. The rule base generated by DRSL is the final input for RIRA to recommend the movies to the group of users. One last factor that is

weightage of every user in a group is also taken into account to improve the recommendations to the group.

A real dataset from MovieLens has been used to implement the proposed recommendation system. This dataset includes total 100000 ratings from 671 users on 9125 movies. All the movies have been classified according to their different genres and these genres are the primarily are covered under all the training sample set. This procedure is done for each and every user who will be new to the proposed system. In this way, system is able to get a unique profile of every user. Ratings are collected from all the users for the movies they have seen. This information about the ratings and movies are stored in the system as the viewing history against every individual user.

Once the ratings and viewing history of the users gets stored in the system, D.R.S.L uses this information and learns the rules for each and every user, also known as RuleBase. Now, when a new movie will be introduced to the system, it will automatically get classified as per the features of the movies data set in the proposed approach. Here, the rating is 5-scaled one. 1 indicates poor, 2 indicate average, 3 indicate above average, 4 indicate a good, and 5 indicate excellent.

At the initial stage of the proposed system, it won't be able to provide any recommendations to the users because it doesn't have any information about the users and their preferences. For this purpose, few external rules must be added to the system to generate the recommendations. So, system collects the ratings for every individual user for few weeks which are called the training sample set for the learning algorithm. With the help of these training sample set, the learning algorithm learns the patterns of rules that user's RuleBase and system will ultimately get the predicted ratings for that movie. These ratings will be stored in the system and help to generate the movie recommendations to the individual users as per their RuleBase. Finally, R.G.R.A uses the set of rules generated by the D.R.S.L and provides the movie recommendations to a group of users.

The flowchart in Figure 1.depict all the steps involved in the process to recommend an appropriate movie to a group of users with the help of proposed recommender system.

In the proposed approach, ratings are collected from users for each and every movie they have seen. But there might be a situation when some users are unlike to provide the ratings. For this purpose, the proposed system generates the ratings for those movies by own as per that user's viewing history. System calculates the probability of occurrence of the movies similar to that movie in the viewing history and generates the ratings based on that. If similar kind of movie occurs more in the viewing history, then the rating will be higher and vice versa.

3.1 Decision Rule Set Learning Algorithm

In content-based recommendation system, learning algorithms are used to learn the profiles of users. The learning algorithm used in the proposed recommendation system is D.R.S.L that is based on two multi-class rule learner algorithms – RIPPER and FOIL. D.R.S.L is a multi-class learning algorithm where five different classes

have been used i.e. poor, average, above average, good and excellent.
Initially, the training sample sets are classified into 2 different data sets – Train and Prune Data Set. Train data set has been used to rules learning and Prune data set is for reducing the complexity of those set of rules and improve the predictive accuracy.

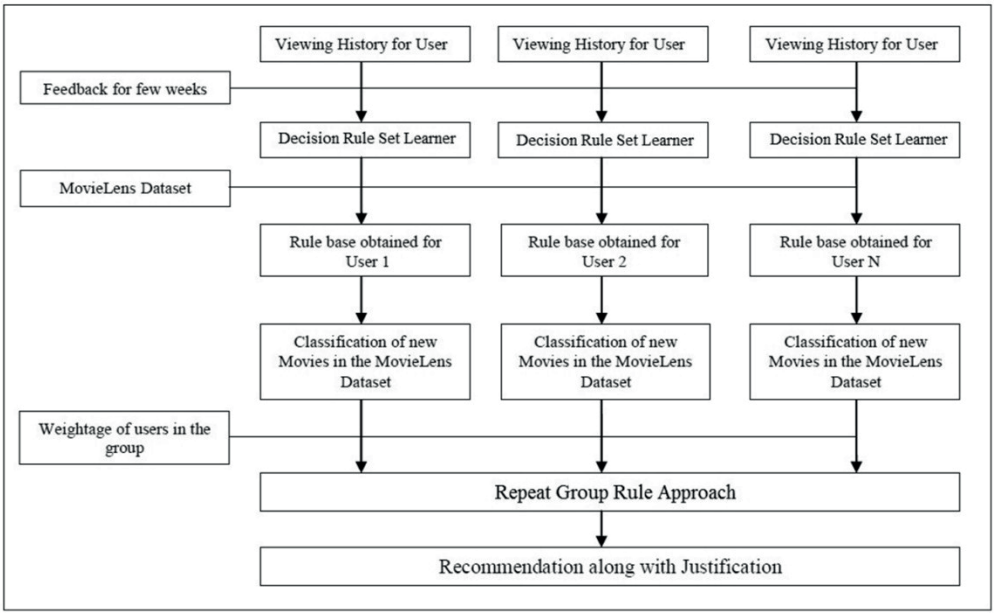

Figure 1. Framework of the proposed group recommender system.

FOIL is a rule learning approach that uses a special performance parameter called FOIL Information Gain given as below:

$$FoilGain(R', R) = s\left[(\log_2 \frac{P1}{P1+N1} - \log_2 \frac{P0}{P0+N0})\right] \tag{1}$$

Where R is the original rule and R' is the candidate rule after adding a condition. N0 and P0 is the number of negative and positive entities of R respectively. N1 and P1 is the number of negative and positive entities of R + R' respectively. S is the number of true positive entities in R and R'.

The pruning criteria which is known as Rule Value Metric is given as follows:

$$RuleValueMetric(R) = \frac{p-n}{p+n} \tag{2}$$

Where n and p are the number of negative and positive examples respectively, present in the prune data set that are covered by the rule (R).
In RIPPER Algorithm, conditions are added to the set of rules to improve the

FoilGain until it reaches its maximum value. Similarly, pruning criteria deletes the final sequence of conditions to maximize the RuleValueMetric.
Once the rule base is obtained for the user, new movies from the MovieLens data set are classified according to that rule base for that particular user. And this list will act as the input to the next step of the proposed approach i.e. R.G.R.A.

3.2 Repeat Group Rule Approach to generate Group Recommendations

When a rule based recommender system is supposed to provide recommendations to a group, an individual approach alone is never sufficient to provide the precise recommendations to that group. So, a combined group rule- based approach has been proposed in this paper that depends on mainly three factors. The first factor is when the least number of group members like the particular item. Second factor is when most of group members are satisfied with the particular item. And the last one is when all the members of the group are satisfied with the item.

These three factors have been combined in the proposed group recommender system to provide accurate recommendations to a group based on the individual choices of the group members. The proposed strategy in this paper has been named as Repeat Group Rule Approach (R.G.R.A).

Let's consider a group G with N number of users within it. The next step is to obtain the two factors mentioned above i.e. when the least number of group members are satisfied and when all the group members are satisfied from every movie which has been classified and forwarded by D.R.S.L. Both the factor values are added for each movie and the maximum of those values is considered to recommend the movie to that particular group. If the same maximum value is observed for more than one instance, then the whole process is repeated without taking least happiness factor into consideration. And this process is applied for every single instance which has been forwarded by D.R.S.L. for each and every group member repeatedly. That's why this algorithm has been named as Repeat Group Rule Approach.

To improve the accuracy of recommendations, weight factor has been assigned to each member of the group since there are always few members in a group whose preferences are considered more than the other members of the group. They might be the decision makers of that particular group, like father or any other senior person in a group of family. There might be a person who has his/her birthday on a particular day and the whole group will consider their preferences rather than other group members on that day. So these members should've assigned more weightage than the other group members.

This weightage factor has been classified into three categories – low, medium and high. After assigning the weightage factor to all the users, the instances are modified specifically which act as the primary input to the R.G.R.A Algorithm. Suppose if a user in a group has rated a particular instance the lowest, then his rating for that instance is divided by his weightage factor. And for other users of the group, their ratings are multiplied with their weightage factor. After doing this modification to the instances, the same R.G.R.A algorithm is applied to obtain the

most accurate recommendations for a group of users.

All the steps involved in the proposed movie recommender system are mentioned in following Algorithm:

Algorithm: Movie Recommender System using D.R.S.L and R.G.R.A

1. Start
2. For every individual class, identify the number of training examples.
3. Now classify all the training examples into positive and negative examples when there is no class left.
4. For all the identified positive examples:
   - 4.1 Take an empty Rule.
   - 4.2 Add conjucts to rule if there is an improvement in the FOIL Information Gain.
   - 4.3 Perform rule pruning by eliminating the final condition sequence.
   - 4.4 Denote the explored positive examples by this rule and add to the RuleSet.
5. Consider a Group G with N numbers of users in it.
6. Numbers of users, their weightage and Instances with Rating of the movies are the inputs to the R.G.R.A algorithm.
7. Based on the number of instances (i) in the Instance Rating Vector(I):
   - 7.1 Eliminate the instances in I with user rating of"0".
   - 7.2 If all the entries in I has "0" user rating, then directly go to next step.
8. Based on the repetition of the number of instances (i) in I:
   - 8.1 If there is no repetition of I in I, then Ci and max is calculated as follows:
     - 8.1.1 Ci = Sum of ratings of all the users and minimum user ratings for instance "i"
     - 8.1.2 Max = maximum value inCi.
   - 8.2 If all the instances have the same ratings in a group, then that item is recommended to the whole group.
   - 8.3 If there are few repetitions in the number of instances and max appears various times in Ci, then eliminate the entries with lowest rating, for that particular instance and repeat the step8.
9. Provide a relevant movie recommendation to the group of users (G).

4. Results and Discussions

The performance of the proposed system will be evaluated under this section. In order to ensure the accuracy of the proposed recommender system, two different groups of users are selected and studied for a period of 1 month. Ratings are

collected from all the users for 20 movies they have already seen. These set of movies with their corresponding ratings are stored in the system under user viewing history section. Now with the help of D.R.S.L, proposed system learns the Rule Base for all the users. Whenever a new released movie is entered in the system, its general information like genre, cast, director, etc. is also stored in the system. So, this new movie gets classified automatically by the system on the basis of past ratings and Rule Base of each user. And hence, system provides a recommendation list for each individual user. Then weightage is assigned to each user of the group and finally, R.G.R.A provides the movie recommendations to the whole group as mentioned in Table 1.

Captain America: Civil War has been recommended to group A since it consists of features such as Action, Adventure, Sci-Fi, and Thriller which are already included in the previous items that the group members have rated higher in the past. Similarly, Kingsman: The Secret Service has been recommended to Group B. The proposed system provided accurate recommendations that match with the interest of whole group.

Group	Recommended Movie	Justification of Recommendation
A	Captain America: Civil War (2016)	Action, Adventure, Sci-Fi, Thriller
B	Kingsman: The Secret Service (2015)	Action, Adventure, Comedy, Crime

Table 1. Recommended Movies to two different Groups.

4.1 Performance of the Proposed System

To evaluate the performance of the proposed movie recommender system, following metrics are used:

4.1.1 Recall

Recall defines the total relevant items retrieved by the proposed algorithm. It can be described as follows:

$$\text{Recall (R)} = I_{relevant} / R_{relevant} \quad (3)$$

$I_{relevant}$ = Number of relevant items retrieved by the proposed system.
$R_{relevant}$ = Total number of relevant items.

4.1.2 Precision

According to Wen et al. (2012), precision is defined as the probability of recommended items that are relevant for the user. So, precision is used as one of the performance evaluation criteria to measure the efficiency of the proposed system. The value is measured as:

$$\text{Precision (P)} = \text{Irecmend} / \text{IRuleSet} \quad (4)$$

Irelevant = Number of relevant items retrieved by the proposed system.

IRuleSet = Number of items retrieved for a particular user (RuleSet).

The proposed recommender system consistently provided the accurate recommendations for various groups of different sizes. The results on MovieLens data set with respect to different values of Pruning Factor (Rule Value Metric) is shown in Figure. 2 below:

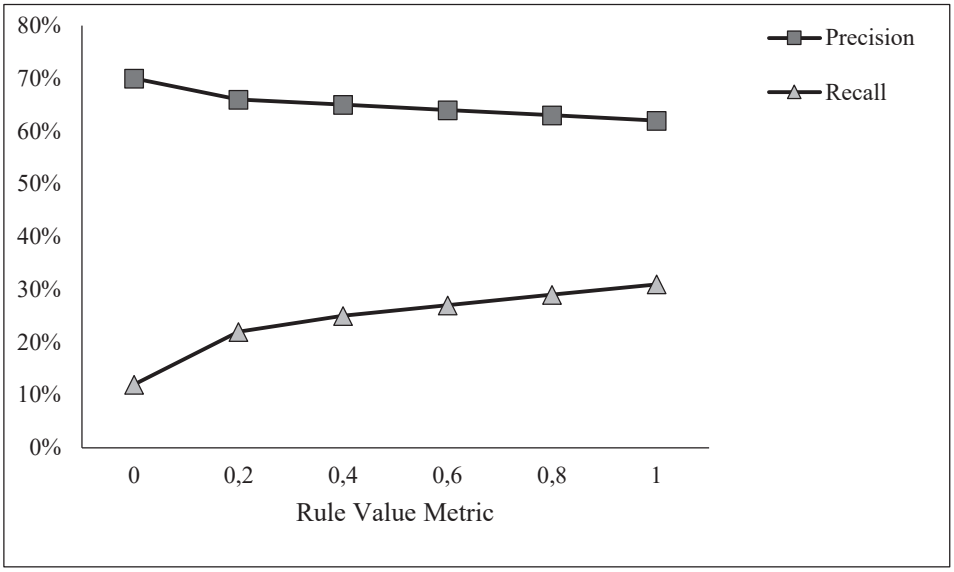

Figure 2. Precision & Recall Metric for MovieLens Data Set.

As depicted by the results, the precision value decreases when Rule Value Metric increases. This is also observed that the decrease is not that steep because in the proposed system, the user's viewing history matches with their rating pattern. The higher precision value is achieved through the proposed recommender system since the accurate recommended items is provided by the system and most relevant ones.

4.2 Outcomes and Comparison

Figure 3 represents the comparison among various learning algorithms such as C4.5 decision tree rule learning approach, cluster-based learning method along with the proposed learning algorithm, which can be used for group recommender system. The graph clearly demonstrates that the performance of the proposed approach is better than the other two learning algorithms. It also proves that D.R.S.L algorithm takes less number of conditions than the C4.5 learning algorithm and cluster based approach that are required for the training data classification. Cluster based methods anyway consumes more time to save all the samples in order to provide the

recommendations. So, the graph of cluster-based approach is having higher values than the other two approaches. The graph represents the difference in time taken by different methods to learn the number of different training datasets. It is also clear from the comparison that the performance of the proposed approach is almost in-line with the training sample datasets.

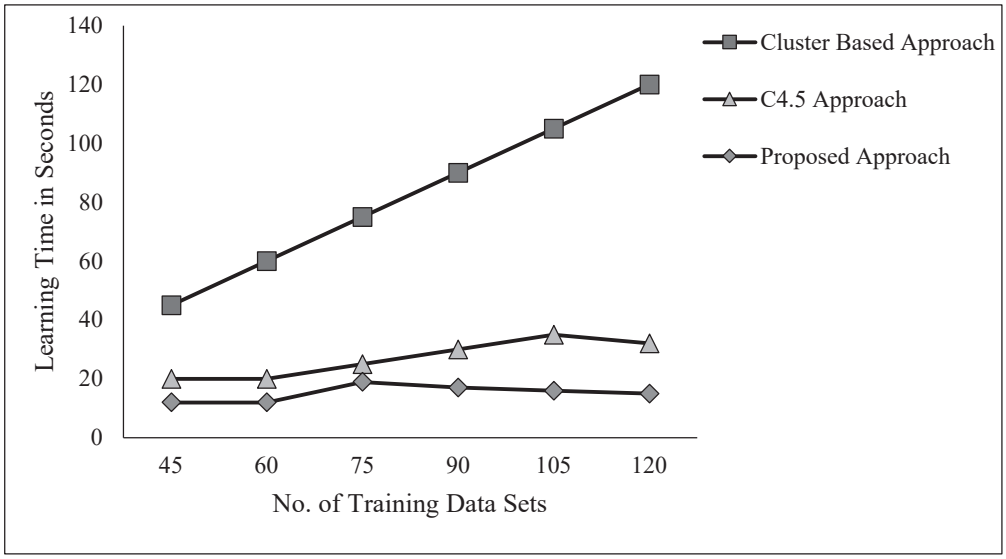

Figure 3. Comparison between C4.5, cluster based approach and the proposed approach.

5. Conclusion

A wide variety of research and development has been performed in the field of group recommender systems; but most of them primarily focus on enhancing the recommendation accuracy by only fine-tuning the algorithm. At the same time, it is important to provide the justification for the respective recommendations that a system provides. Through this research work, a novel content-based group recommendation algorithm for movies has been proposed, which is a combination of two unique algorithms. One is D.R.S.L., used to generate the rule base for the users and this rule base is used by another algorithm called R.G.R.A to provide the movie recommendations to a group of users. The proposed group recommendation system strictly operates according to the preferences of users, their previous ratings and weightage of each and every user of the group. Performance has been enhanced over the other traditional learning algorithms such as C4.5 and cluster-based approach. Same has been expressed in term of precision value and the time taken to learn the set of rules for a user.

6. Current & Future Developments

In recent years, a lot of progress has been made with new studies on the movie recommender systems. The proposed group recommendation system along with its findings and better results indicates towards a prosperous future of new and improved movie recommender systems that could revolutionize many applications such as movie ticketing apps and online video content providers.

Acknowledgements

The authors would like to express their gratitude to Dr. Ashok K. Chauhan, the Founder President of Amity University for his keen interest in promoting research in the Amity University and have always been an inspiration for achieving great heights.

References

[1] J. S. Moses and L. D. Babu, "Evaluating Prediction Accuracy, Developmental Challenges, and Issues of Recommender Systems," *International Journal of Web Portals*, vol. 10, no. 2, Jul., pp. 61-79, 2018.

[2] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments," *Decision Support Systems*, vol. 74, Jun., pp. 12–32, 2015.

[3] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: the state of the art," *User Modeling and User-Adapted Interaction*, vol. 25, no. 2, Jun., pp. 99–154, 2015.

[4] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, Nov., pp. 261-273, 2015.

[5] M. Presulli, "Studying Collaborative Approaches to Improving the Quality of Management Information in IT Projects," *International Journal of Information Technology Project Management*, vol. 7, no. 4, Oct., pp. 1-20, 2016

[6] J. Zhang, Y. Lin, M. Lin, and J. Liu, "An effective collaborative filtering algorithm based on user preference clustering," *Applied Intelligence*, vol. 45, no. 2, Feb., pp. 230-240, 2016.

[7] W. K. Hyung, H. Keejun, Y. Y. Mun, C. Joonmyun, and H. Jinwoo, "MovieMine: Personalized Movie Content Search by Utilizing User Comments," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, Nov., pp. 1416-1424, 2012.

[8] T. Chen and L. He, “Collaborative Filtering based on Demographic Attribute Vector,” in Proc. ETP International Conference on Future Computer and Communication FCC’09, 2009, pp. 225-229.

[9] S. V. Aciar, G. I. Aciar, C. A. Collazos, and C. S. González, "User Recommender System Based on Knowledge, Availability, and Reputation From Interactions in Forums," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 11, no. 1, Feb., pp. 18-22, 2016.

[10] R. Sreepada, B. Patra, A. Chakrabarty, and S. Chandak, “Revisiting tendency based collaborative filtering for personalized recommendations,” in Proc. of the ACM India Joint International Conference on Data Science and Management of Data ’18, 2018, pp. 230–239.

[11] Y. Hu, W. Shi, H. Li, and X. Hu, “Mitigating Data Sparsity Using Similarity Reinforcement-Enhanced Collaborative Filtering,” *ACM Transactions on Internet Technology*, vol. 17, no. 3, Jul., pp. 1-20, 2017.

[12] L. Zhao, Z. Lu, S. J. Plan, and Q. Yang, “Matrix factorization+ for movie recommendation,” in Proc. of the Twenty-Fifth International Joint Conference on Artificial Intelligence’ 16, 2016, pp. 3945-3951.

[13] X. Ma, H. Lu, Z. Gan, and Q. Zhao, “An exploration of improving prediction accuracy by constructing a multi-type clustering based recommendation framework,” *Neurocomputing*, vol. 191, May, pp. 388-397, 2016.

[14] G. Costa and R. Ortale, “Model-Based Collaborative Personalized Recommendation on Signed Social Rating Networks,” *Transactions on Internet Technology*, vol. 16, no. 3, Aug., pp. 1-21, 2016.

[15] N. A. ALRossais and D. Kudenko, “ISynchronizer: A Tool for Extracting, Integration and Analysis of MovieLens and IMDb Datasets,” in Proc. Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization’ 18, 2018, pp, 103-107.

[16] M. Kompan and M. Bieliková, “Content-Based News Recommendation,” in *E-Commerce and Web Technologies*, F. Buccafurri and G. Semeraro, Eds. Lecture Notes in Business Information Processing, vol 61. Springer, Berlin,Heidelberg, 2010, pp. 61-72.

[17] S. Feng and J. Cao, “Improving group recommendations via detecting comprehensive correlative information,” *Multimedia Tools and Applications*, vol. 76, no. 1, Jan., pp. 1355-1377, 2017.

[18] M. Soares and P. Viana, “TV Recommendation and Personalization Systems: Integrating Broadcast and Video On demand Services,” *Advances in Electrical and Computer Engineering*, vol. 14, no. 1, Feb., pp. 115-120, 2014.

[19] L. Boratto, S. Carta, and G. Fenu, "Discovery and representation of the preferences of automatically detected groups," *Future Generation Computer Systems*, vol. 64, Nov., pp. 165-174, 2016.