International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2022, Vol 01, Issue 02, 19-29

Science Transactions © 2022

Original Paper

THE NOVEL APPROACH OF RECOMMENDATION ON MOVIE DATA

Manisha Valera[1] and Rahul Mehta[2]
[1]Department of Computer Engineering, Ph.D. Scholar, Gujarat Technological University, Ahmedabad
valeramanisha@gmail.com
[2]Department of Electronics & Communication Engineering, GEC Rajkot, Gujarat Technological University, Ahmedabad
rdmehta@hotmail.com

Abstract

With constant advancement of web applications around the world, it is a challenge to find the suitable information needed for the user in a limited time. Without a proper recommender system, it is very cumbersome to get essential information from the web applications. For different kinds of requirements different types of recommender systems have been planned. This paper identifies crucial areas of research openly available for new researchers. After analyzing this paper new researchers can understand problems of recommender systems which need improvement and hence, they can make those problems as their area of research. The rating or preference that is given to an item, can be predicted using recommendation systems. The observations in this paper will directly support researchers to better understand present developments and new directions in the field of recommender systems using Artificial Intelligence.

Keywords

Recommendation System, Machine Learning, Deep Learning, Hybrid Filtering

1. Introduction

The recommender system is able to provide suggestions (recommendations) to the users, in multiple contexts like when they are making a choice amid a large list of items or whenever they want to receive suggestions. [1]
recognizes 4 key features:

- **Help to Decide:** forecasting a rating for a user for an item
- **Help to Compare:** rank a list of items in a custom-made way for a user
- **Help to Discover:** provide a user with unfamiliar items that will be valued
- **Help to Explore:** give items alike to a given target item

Actually, there are several motives because service providers employ recommendation engines:

Increase the revenue: it means increase the numbers of things that are sold. The goal here is to actually vend more items than there would have been without any recommendations. To meet this objective, recommended items are expected to satisfy the user's choice and need.
Increase diversity of items sold: The goal of this function is to incite users to select items that would remain unfamiliar without recommendation. For example, the service provider wants to sell books from all its catalogue and not only the top 10 most widespread ones.

Improve the user experience: If the system works properly and is planned in a proper manner, it can rise the user satisfaction. Indeed, by receiving interesting, diverse and relevant recommendations, the user will gain the experience on the web.

Understanding users: Another foremost purpose of a RS is to be able to describe users' likings. These likings may have been collected explicitly or by predicting them. This data might be used to manage its production or stock.

Burke's classification [2] offers a very comprehensive classification of present recommendation techniques by identifying each method's input data and its algorithm used.

He has defined five types of recommendation methods:

- Collaborative filtering
- Content-based
- Demographic
- Utility-based
- Knowledge-based

2.Classification of Recommendation System

Based on how the systems study and filter the data according to the user necessities, there are numerous recommendation systems[3].

2.1 Content Based Recommendation System (CBR):

It is based on description of the item and the profile of the user which is build as per their preferences. The items are defined using keywords and user profile covers all the items that the user likes. This kind of recommendation system recommends items to the user as stated by his profile. The user profiles are restructured automatically based on their feedback. The filtering techniques for information retrieval is used, in which some user items are matched with those items that the user had previously rated and the recommends the best item.

Advantages:

- Give recommendation based on user likings.
- No sparsity issue.
- No cold start issue.
- Automatic updation

Disadvantages:

- It depends on the past knowledge of the user to give recommendation.

2.2 Collaborative Filtering Recommendation System (CFR):

It depends on gathering and analyzing the data which is based on the user preferences and make a prediction of items based on their similarity with other users, this similarity is considered by comparison of their ratings with the ratings of other users for the identical item.

There are two types of collaborative filtering methods – item based and user based.

Item based collaborative filtering system In this system; item relationships are recognized using user item matrix and use this to compute recommendation for users indirectly. It recommends item to the user, depending on other items with high correlations.

Advantages:

- High performance
- Decent prediction quality

User based collaborative filtering system User based collaborative system recommends an item to the user based on other similar minded users' opinion who has common interests for that item. The users with common interests are considered as the nearest neighbours. After that the prediction for the item will be made based on the rating that the user's neighbours give for the same item.

Advantages:

- Simple
- Efficient
- Accurate

Disadvantages:

- Cold start issue
- Scalability issue
- Data sparsity issue

2.3 Hybrid Recommendation System

Hybrid recommendation system is demonstrated to increase the efficiency which combines the above two recommendation systems.

It can be implemented in 3 ways –

- by separately making predictions and then combining them,
- adding functions of content-based system to collaborative system and combining the models into one model.

The example is Netflix in which recommendations are provided by comparing the similar users as in collaborative system and recommends the movie that is highly rated by the users as in content-based systems.

Advantages:

- Avoids scalability issue
- Minimize data sparsity

Disadvantages:

- Cannot adapt to increasing number of users or items.

2.4 Context Aware Recommendation System

Context characterizes the situation of an entity such as person, place or object. Tour guides heavily depend on on location and ignore other context types. As a result, information overload problem arises. This can be evaded by using appropriate personalization and content filtering method. This system studies five main types of contextual data such as location, time, weather, personalization and social media entertainment.

Advantages:

- It detects the level or variance between data.
- Reduces the dimensionality of data.
- More accurate

Disadvantages:

- High prediction errors.

2.5 Demographic Recommendation System

They can recommend items according to the demographical information of users. This type of system depends on the theory that each demographic class should be recommended differently. Websites can easily grasp the displaying language according to the user's country or language. Recommendations can take into consideration the user's age.

2.6 Knowledge Based Recommendation System

This system recommends products based on the necessities of the user and his/her preferences. It needs a functional knowledge about the user for satisfying them by recommending a particular item. So, there is a great use of domain knowledge. Its decision is self-determining, so there is no need to gather information about a particular user.

Advantages:

- Avoids early rated problem.
- Avoids sparse rating problem.
- More efficient

Disadvantages:

- High cost
- Less accuracy
- Asymmetric model

3. Research Problems

Recommendation systems have a boundless future. But some problems are yet to be solved by the research community to make research more effective. Figure 1 Shows all the research problems which can give ideas to work in the area [4].

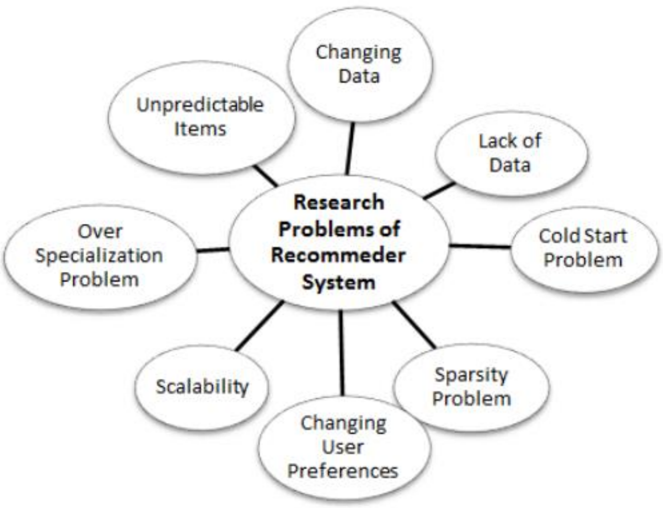

Figure 1. Research Problems of Recommender System

This section examines issues and challenges in recommender systems and offers different solutions provided by researchers to handle these issues. [5]

2.7 Cold Start Problem

It happens when new users enter the system or new items are added to the list. In this scenario, neither the taste of the new users can be forecast nor can the new items be rated or purchased by the users results into poor recommendations. The cold-start problem can be solved by:

- asking the user in the starting to rate some items
- asking the users clearly to state their taste in collective manner
- suggesting items to the user based on the gathered demographic information.

2.8 Synonymy

It arises when an item exists with two or more different names or entries having same meanings. In such cases, the recommender can't recognize whether the terms represent different items or the same item. For example, a memory-based CF approach will treat "Action movie" and "Action film" differently. the extreme practice of synonym words declines the performance of CF recommenders. To ease the problems of synonymy, different techniques including ontologies, the Single Value Decomposition (SVD) techniques, and Latent Semantic Indexing (LSI) could be used.

2.9 Shilling Attacks

What happens if a mean user or competitor enters into a system and starts giving wrong ratings on some items to rise the item popularity or to degrade its popularity. this can break the trust on the recommender system and degrade the performance and quality of system. This risk is of more concern in CF techniques but lesser to the item-based CF technique.

2.10 Privacy:

Feeding personal information to the recommender systems may cause issues of data privacy and security. whether CB or CF recommender system, should build trust amid their users, however CF recommenders are more susceptible to such privacy issues For this purpose, cryptographic Techniques can be used by providing recommendations without involving third parties and peer users.

2.11 Limited Content Analysis and Overspecialization

Content-based recommenders depend on on content about items and users. The limited availability of content leads to difficulties including overspecialization. Here, items are characterized by their subjective attributes, where selecting an item is based on their subjective attributes. relevant items cannot be recommended unless the analyzed content contains enough information to be used in distinctive items liked/disliked by the user. In order to recommend innovative items along with familiar items, we need to introduce randomness by using genetic algorithms that brings variety to recommendations. [6]

2.12 Grey Sheep

It happens in pure CF systems where views of a user do not match with any group ,therefore, is unable to get advantage of recommendations . Pure CB filtering can solve this matter where items are suggested by manipulating user personal profile and contents of items being recommended. Mixing CB with CF techniques may also produce more novel recommendations.

2.13 Sparsity

The availability of huge size of data about items list and the reluctance of users to rate items make a disseminated profile matrix results into less precise recommendations. To deal with this situation, several methods can be used including multidimensional recommendation model, SVD techniques, demographic filtering, content-boosted CF algorithm.

2.14 Scalability

It becomes problematic for a typical recommender to process large-scale data. For example, Amazon.com recommends more that 18 million items to more than 20 million customers. Different techniques have been planned including clustering, reducing dimensionality, and Bayesian Network. The problem can be addressed by using clustering algorithms that search users in small clusters instead of entire database.

2.15 Latency Problem

CF recommender Systems face latency problem when new items are added more frequently to the database, where the recommender suggests only the already rated items as the newly added items are not yet rated. Using CB filtering can decrease waiting times but it may create overspecialization. To handle this situation, the category-based approach in combination with the user can be used.

2.16 Evaluation and the Availability of Online Datasets

Evaluating a recommender system determines its quality. The design of evaluation criteria and selection of suitable evaluation metrics is tricky in recommender systems. Most traditional recommender systems evaluate system results and apply metrics like MAE, Precision, and F-Measure for evaluation. These are not applicable in different domains especially in evaluating context-aware recommenders. To evaluate them, contextual precision and contextual ROC can be used. another evaluation methods contain questionnaires or interviews. Yet, these approaches are expensive and time-consuming.

2.17 Context-Awareness

Context-awareness aggregates all categories that represent the setting in which recommender is deployed, e.g., the current location, the current activity, and the time. It is proposed that the forthcoming recommender systems will use contextual information got through mobile services infrastructure and will include the user's short and long term history, location, entries in the calendar, and the information that the user provides to social networks .Finding out user likings and context-related information is crucial to come up with appropriate recommendations for the user.

Data for recommendation system are of 2 types: implicit and explicit. Implicit data are logged from user click streams, hyperlink navigation whereas explicit data are in form of ratings or feedback provided by user for an item. Automatic recommendation systems are specialized data processing systems that are enhanced for collaboration with customers in place of marketers [7].
Feedback Collection is a vital part of the recommendation. This data can be collected two different ways:
Explicitly: by directly requesting the user to provide feedback, usually by providing a rating mechanism to the user ("rate this film on a scale from 0 to 10") or a questionnaire with which the user can express his/her satisfaction and view about the content.
Implicitly: by inertly stalking the user's activity. It is less accurate as it tracks user's actions while he/she is viewing the selected content item.[8]

4. Related Work

In this paper[12] They have proposed a MOVREC movie recommendation system which works on collaborative filtering approach that uses information provided by users, analyzes them and then recommends the movies to the user at that time. The recommended movie list is sorted according to the ratings which were given by previous users and K-means algorithm is used. This system was developed in PHP using Dreamweaver 6.0 and Apache Server 2.0. The recommendations are generated using various types of knowledge and data about users, the available items, and past transactions stored in customized databases. The user will be able to browse the recommendations to find a movie of their choice. Based on informal evaluations that were carried out over a small set of users which got a positive response. They would like to have a bigger data set that will enable more meaningful results using system. The future work is to perform different machine learning and clustering algorithms and study the comparative results. Ultimately, they would like to implement a web-based user interface that has a user database, and has the learning model personalized to each user.
In this paper[13], They have analyzed the two techniques UBCF and IBCF on MovieLens dataset and provided the results of the performance of each algorithm and found which algorithm provides better results.They implemented two techniques UBCF and IBCF with the implementation metrics, namely, True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN), True Positive Rate (TPR), False Positive Rate (FPR), Precision (Pr), and Recall (Re), Three datasets (MovieLens 100K, 1M and, 10M) were used to compute the error metrics, and two datasets (MovieLens 1M and 10M) were used to compute the implementation metrics. It was detected that IBCF provided improved results as compared to UBCF. It shows that when recommendations are made based on items previously liked by the user, the efficiency of such recommendations is more as compared to the recommendations made based on users who all like a similar item. As a future scope, this work can further be implemented using deep learning and machine learning techniques.
In this paper[14], seven clustering algorithms were used to cluster performance comparison methods for movie recommendation systems, such as the K-Means algorithm, birch algorithm, mini-batch K-Means algorithm, mean-shift algorithm, affinity propagation algorithm, agglomerative clustering algorithm, and spectral clustering algorithm. They were used to compare the best algorithms with regard to the similarity groupings of users on movie genre, tags, and rating using the MovieLens dataset. Then, optimizing K for each cluster did not

significantly increase the variance. To validate the quality of the recommender system evaluation measures, the mean squared error (MSE), Dunn Matrix as cluster validity indices and social network analysis (SNA) were used. additionally, average similarity, computational time, association rule with Apriori algorithm and clustering performance evaluation measures were used for comparing the methods of performance systems.
In this paper MovieREC[15], a recommender system for movie recommendation is introduced. It lets a user to select his choices from a specified set of attributes and then recommend him a movie list based on the cumulative weight of different attributes and applies K-means algorithm. Moreover, they would like to incorporate different machine learning and clustering algorithms and study the comparative results. In future, they would like to implement a web-based user interface that has a user database, and has the learning model tailored to each user.

This paper [16] demonstrates the role of Data Mining in Recommendation System and proposes a workflow of RS. Also defines the review of techniques, challenges of RS & compares recommendation systems of numerous e-commerce websites. The proposed workflow of RS has two phases: a) Information gathering phase b) analysis and recommendation. user's information is collected regarding searching, buying habits etc. and this gathered information about user is accommodating in analyzing user taste. Hybrid Techniques is a mixture of more than two approaches to obtain discrepancies & optimization sole systems. Some tricks to achieve hybridization are like, Weighted Method where score of various items is integrated to build solitary recommendation. Switching Method which switches between RS techniques based on simultaneous state. Mixed Method in which Numerous recommendations from different system are represent at same time. Feature Combination where Several attributes from divergent knowledge sources are combine together in order to generate single algorithm. Feature Augmentation where output of one approach is considered as the input of next approach. Cascade Method where recommender is given rigid importance rather than the lower importance, before breaking the chain of higher scoring. Meta-Level is pertained to build a model whose input used by next approach.
All the algorithms described in this paper [17] are compared with respect to their precision rates which are Content Based Filtering, Collaborative Based Filtering, Hybrid Content-Collaborative Based Filtering, k-mean clustering and Naive Bayes classifier. with the intention of achieving the best possible precision they have presented a comprehensive comparative analysis. This portrays the strength and the weakness of each one of them in different versions of the MovieLens dataset. this paper adapts that out of all these approaches Naive Bayes gives the finest precision.
This paper [18] focuses on the movie recommender system based on movies' genres and actors/actresses themselves as the input tags, or tag interpolation of the MovieLens datasets. We apply tag-based filtering and collaborative filtering that can effectively predict a list of movies that is similar to the movie that a user has been watched. Due to not reliant on users' profiles, our model has eliminated the effect of the cold-start problem. The training and test sets have been split into five-fold cross-validation. The experimental results provide well recommendations to users because it utilizes the underlying relation between movies based on their similar genres and actors/actresses. The idea can be further executed in any datasets that emerge the identical characteristics.
This Paper [19] estimates the effectiveness of adding movie trailer responses as the side information to the movie rating data for movie recommendation. To integrate the trailer data, they have used three approaches: integrating all of them as movie features, treating sentiment scores as a rating matrix to integrate with the movie rating matrix and others as the movie features, only integrating the sentiment rating matrix with the movie rating matrix. Inclusive, the evaluation results show that if They include movie trailer data, it decreases the prediction error and rises the recommendation accuracy. if all the trailer feedback data is combined as the movie features, our recommender system provides the most precise result. They also discovered that deep neural network model performs better than the matrix factorization model. In future, it

can be extended by adding temporal signals as the side information to the system . As users' conditions to find a movie and user likings on a movie may change over time, if we can add the temporal signals in our model, we might be able to recommend movies based on users' fluctuating interests. different DNN models can be tried to implement the recommender system. For example, instead of MLP, we may try the CNN model, or the RNN model if we consider the sequence information (time of ratings).

5. Proposed System

The MovieLens data set is mainly made up of three data tables, respectively is a user table (user), the movie table (item), scale (rates). This experiment using the MovieLens own training set and testing set for testing.[20]

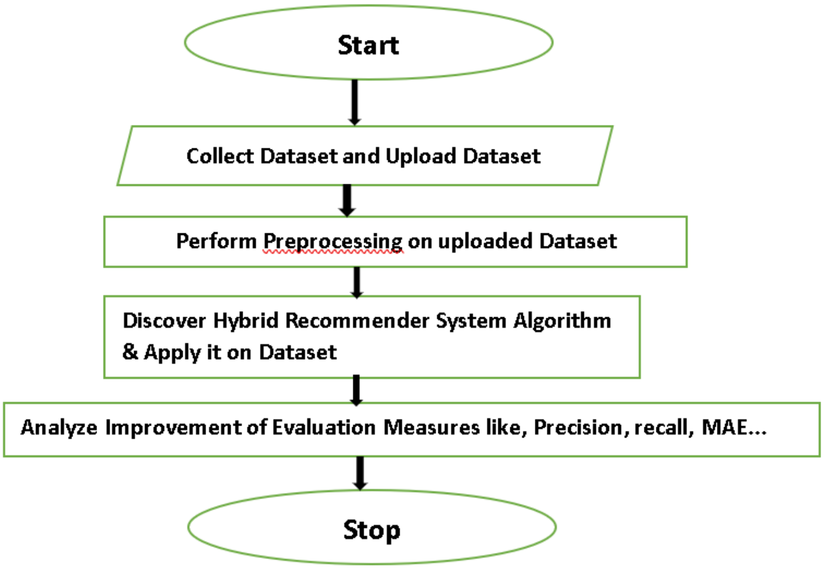

Figure 2. Proposed Recommender System

In The Proposed System, first we will collect MovieLens dataset and upload for the system implementation, followed by Data cleaning. Then We will Discover the Hybrid Recommender System Algorithm which can be applicable on the dataset, which can provide improvement in Evaluation metrics like Precision, Recall, MAE etc.

6. Conclusion

This paper has presented recommender systems to new research studies. This paper has also identified key problems which is needed to resolve in recommender systems. This paper can help PhD Scholars and Masters students in picking their area of research. The research breach is already presented in this paper to form diverse problems of recommender systems. As these problems get solved more and more useful recommendation systems will become more intellectual and usable.

REFERENCES

[1] Meyer., 2012 F. Meyer., 2012. Recommender systems in industrial contexts. Thèse de Doctorat, University of Grenoble, France. 6, 21

[2] R. Burke, 2002. Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction

[3] Gigimol S and Sincy John, "A Survey on different types of recommendation systems," *International Research Journal of Advanced Engineering and Science*, Volume 1, Issue 4, pp. 111-113, 2016.

[4] Mishra, N., Chaturvedi, S., Vij, A., & Tripathi, S. (2021). Research Problems in Recommender systems. *Journal of Physics: Conference Series*,1717, 012002. doi:10.1088/1742-6596/1717/1/012002

[5] Shah Khusro, Zafar Ali, Irfan Ullah, Recommender Systems: Issues, Challenges, and Research Opportunities, *Information Science and Applications (ICISA) Springer*, 2016 Volume 376, doi:10.1007/978-981-10-0557-2_112

[6] Binu Thomas , Amruth K John ,Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis To cite this article: 2021 *IOP Conf. Ser.: Mater. Sci. Eng*. 1085 01201

[7] Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). Ecommerce recommendation applications. In Applications of Data Mining to Electronic Commerce (pp. 115-153). *Springer US*.

[8] F.Ricci, L. Rokach, B. Shapira, P. Kantor (Eds.), Recommender systems handbook, Vol. 1, *Springer*, 2011.

[9] Pradeep Kumar Singh, Pijush Kanti Dutta Pramanik, Avick Kumar Dey, Prasenjit Choudhury, Recommender Systems: An Overview, Research Trends, and Future Directions, January 2021, *International Journal of Business and Systems Research* 15(1):14–52

[10] Qian Zhang, Jie Lu, Yaochu Jin, Artificial intelligence in recommender systems, *springer Complex & Intelligent Systems* (2021) 7:439–457 https://doi.org/10.1007/s40747-020-00212-w

[11] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay, Deep Learning based Recommender System: A Survey and New Perspectives, *ACM Comput. Surv*. 1, 1, Article 1,July 2018,DOI: 0000001.0000001

[12] Manoj Kumar , D.K. Yadav , Ankur Singh , Vijay Kr. Gupta , A Movie Recommender System: MOVREC, *International Journal of Computer Applications* (0975 – 8887) Volume 124 – No.3, August 2015

[13] Garima Gupta, Rahul Katarya, Recommendation Analysis on Item-based and User-Based Collaborative Filtering, *International Conference on Smart Systems and InventiveTechnology(ICSSIT)IEEE*,November2019,DOI: 10.1109/ICSSIT46314.2019.8987745

[14] Debby Cintia, Ganesha Putri , Jenq-Shiou Leu, Pavel Seda, Design of an Unsupervised Machine Learning-Based Movie Recommender System, *Symmetry 2020 Multidisciplinary Digital Publishing Institute (MDPI)*,January 2020, doi:10.3390/sym12020185 (Scopus)

[15] Ashrita Kashyap ,Sunita.B, Sneh Srivastava,Aishwarya,Anup Jung Shah, A Movie Recommender System: MOVREC using Machine Learning Techniques, *International Journal of Engineering, Science and Computing*, volume 10, June 2020

[16] Neha Verma, Devanand, Bhavna Arora, Experimental Analysis of Recommendation System in e-Commerce *International Journal of Innovative Technology and Exploring Engineering*, ISSN: 2278-3075, Volume-8 Issue-8S3, June 2019(Scopus)

[17] Satya Prakash Sahu, Anand Nautiyal, Mahendra Prasad, Machine Learning Algorithms for Recommender System - a comparative analysis, *International Journal of Computer Applications Technology and Research* ,February 2017

[18] Quynh Nhut Nguyen, Nghia Duong-Trung , Dung Ngoc Le Ha, Xuan Son Ha,Tan Tai Phan,Hien Xuan Pham, Hiep Xuan Huynh, Movie Recommender Systems Made Through Tag Interpolation, *ICMLSC 2020, ACM: Proceedings of the 4th International Conference on Machine Learning and Soft Computing*,January 2020 Pages 154–158,https://doi.org/10.1145/3380688.3380712

Authors

Manisha Valera

Manisha Valera is a leading researcher in the field of Machine learning and AI. She has been published extensively as author and co-author of over 10 papers in highly regarded, peer-reviewed journals.

Rahul Mehta

Rahul Mehta is a leading researcher in the field of Networking and AI. He has been published extensively as author and co-author of many papers in highly regarded, peer-reviewed journals.