

A Recommendation System: Trends and Future

Shefali Gupta, Meenu Dave



Abstract---Recommendation system plays a key role in e-commerce universe and is used in many applications, websites and more. It has led to synergies between applications, created global village and growth of information. This paper represents the overview of approaches and techniques generated in recommendation systems. Recommendation system is categorized in two classes: Personalized and Non-personalized, which is further divided into various approaches and techniques. This paper discusses each of the methodology in detail highlighting their strengths and weaknesses.

Keywords: Recommendation System, Personalized recommendation system, Non-Personalized recommendation system, Content Based Filtering, Collaborative filtering, Knowledge Based Filtering, Hybrid Filtering

I. INTRODUCTION

“Recommender systems are the software tools and techniques that provide suggestions, such as useful products on e-commerce websites, videos on YouTube, friends' recommendations on Facebook, book recommendations on Amazon, news recommendations on online news websites, and the list goes on” [1]. Over many years significant work has been done in academics and industry to develop new approaches to recommender systems, but still interest remains high because it helps user to handle surplus information in order to provide personalized recommendations and services to them.

II. BACKGROUND & RESULTS

Various techniques are there which use rating or content knowledge to provide recommendation, but most approaches suffer from one limitation or the other. A Hybrid technique was suggested by combining different techniques to overcome the limitations of each technique individually. Some of the key techniques have been shared below [2].

2.1 Non-personalized Recommendation System

A non-personalized recommender system is independent of the user so it makes same recommendation to everyone. For example, if you visit any website as an unidentified user, it recommends items that are currently observed by other customers. In Non-personalized recommendation system following algorithms are used:

Aggregated Opinion Approach.

Today, various ecommerce websites are using average customer ratings to display top rated items as

recommendations to a user. These ratings signify popularity of an item. The depiction of these ratings generally ranges from 0 to 5 and is calculated as $\text{Score} = \text{round}(\text{MEAN}(\text{ratings}) * 10)$ [3]. On the other hand, some websites may display graphs with top-N items based on highest average ratings, while others use percentage of the people who rated an item good or bad.

Merits.

- It is easy for a user to interpret the results [3]
- As only the highly rated items gets displayed, it becomes easy to implement [6]
- Data collection is relatively easier in this technique due to limited number of variable required [6]

Demerits.

- It lacks personal appeal as it does not account for user specific attributes [3]
- These systems face challenges in clustered diverse population [3]

Product Association Approach.

Product association technique identifies the best possible combination of products or services which are frequently bought by customers from transactional dataset. With the multiplication in the volume of data being gathered from daily activities, product association technique has gained major popularity as a tool of data mining [4]. Product Association analysis is done based on an algorithm named “Apriori Algorithm”.

Association rule is an inference rule of the form $X \rightarrow Y$, where X and Y are itemsets. For example, in the sales data, association rule can be defined as $\{\text{item1}\} \rightarrow \{\text{item2}\}$ which implies that if ‘item1’ is bought, customer buys ‘item2’ as well.

Following rule evaluation metrics are used, in which higher the ratio higher the probability of item being purchased:

Support: This signifies popularity of an item. For example, in fig. 3, item 1, 2 and 3 are bought together in 2 of 5 cases hence 40% support.

Confidence: This signifies the possibility of item Y to be bought when item X is bought. It defines as percentage of transactions in which item X appears and item Y also appears.

Merits.

- It calculates extra sets of frequent items and takes into account huge item set property [7].
- Apriori algorithm is implementable and can be parallelized easily [7].

Manuscript published on 30 September 2019.

* Correspondence Author (s)

Shefali Gupta, Jagannath University, Jaipur, Rajasthan, India.

Dr. Meenu Dave Jagannath University, Jaipur, Rajasthan, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Demerits.

- It assumes transaction database is memory resident thereby consuming huge memory [7].
- It involves numerous database scans as it iterates through the transaction set each time [7].
- Process of frequent itemset generation could be extremely slow as it generates too many candidates item set [8].

2.2 Personalized Recommendation System

A personalized recommender system is dependent on the user so it makes recommendations based on user preferences. For example, if you visit any website, it recommends items that matches your interest and characteristics.

Content Based Filtering System.

A content-based filtering system provides recommendation of an item to a user based on content or description of the item liked by the user in the past (user’s profile). It, then builds a profile of the item’s user might be interested in purchasing based on the content of those rated items. Using the profile, item suggestions are provided that would fit for the user [9].

Techniques.

- *Term Frequency (TF) and Inverse Document Frequency (IDF).* In this Technique, Terms are allocated a weight that is based on how regularly it appears in a specific document and how regularly it occurs in the entire document collection [10].

Term Frequency (TF) is given by:

$$TF(t) = \frac{\text{Number of times term (t) appears in a document}}{\text{Total number of terms in a document}}$$

Inverse Document Frequency (IDF) is given by:

$$IDF(t) = \log \frac{\text{Total number of documents}}{\text{Number of documents with term(t) in it}}$$

Merits.

- *No Cold Start.* It recommends new items immediately before being rated by significant users [14].
- *User Independence.* This method only uses items and user profile for providing recommendation [19].
- *Transparency.* User can know, based on which feature recommendation was given [19].

Demerits.

- *Limited content analysis.* If content is limited enough to distinguish between two different items with same set of attributes, the recommendation will be not be correct [14].
- *New user.* New users don't get correct recommendations as ratings are not sufficient when they arrive on website [14].
- *Over-specialization.* Users are delimited to getting recommendations which looks similar to those already known or defined in their profiles [19].

Collaborative filtering system.

Collaborative filtering system technique provides recommendation of an item to a user based on the behavior

and interest of other users. Feedback for an item is collected from the user in the form of ratings and these ratings are utilized are provide recommendation for a new item [11].

Techniques.

- *Memory Based Filtering System.* This approach uses entire user-item rating database to be stored in memory to provide recommendations. It aims at finding a set of users, called as neighbors. This neighborhood is then used to provide recommendations to the new user [6].

There are two primary types of memory-based filtering system based on neighborhood approach [2]:

– *User Based Filtering.* This technique provides recommendations using similarity between users. It compares users’ ratings on the same item and recommendation for the item by the active user is made using weighted average of the ratings of the item by users similar to the active user. Here, weights are the similarities of the users with the target item.

– *Item Based Filtering:* This technique provides recommendations using similarity between items. It looks into user-item matrix and extracts items rated by an active user and then uses this retrieved item to determine how similar they are to the target item. It then selects the most similar items and recommendation is made by taking weighted average of the active users rating on the similar items.

- *Model Based Filtering System:* This approach first develops a parametric model on training dataset of ratings and then uses this model to provide recommendations to the user for unlabeled item, given his/ her ratings for other items without using complete dataset every time [2].

Machine learning models used for model building are as follows:

– *Cluster Based Filtering.* This technique works by grouping items into clusters. Grouping is done so that the items in same cluster are more similar than items in other clusters thereby forming meaningful clusters from the data. Pearson correlation coefficient is used to provide recommendation to the user. Here, recommendation can be given using small groups (cluster) instead of the complete database [12].

– *Bayesian network model.* Bayesian-network model uses Naive Bayes to calculates probability that the user will buy item given users and items history of transaction. The probability can be determined by: [13]

$$p(k_1 | m_1, m_2, m_3, \dots) = p(k_1) \cdot \left(\frac{p(m_1, m_2, m_3, \dots | k_1)}{p(m_1, m_2, m_3, \dots)} \right)^{\frac{c_{k_1}}{n}}$$

Here, k1 is the item, ‘m1, m2, m3, ...’ are the transaction history of an active user, ‘n’ is the number of items in transaction history and ‘n’ is a constant value of 3 determined by Wang & Tan [13] experiment.

Merits.

- Collaborative filtering works well where content analysis of item is difficult and expensive [2].
- Memory based filtering techniques does not taken content of item in account and new item can be added easily [6].



- Model based filtering techniques are easily scalable and overcomes performance and sparsity problem [11].

Demerits.

- *Grey Sheep.* Opinions of a user do not contest with any group and the user gets classified in more than one group. Therefore, is unable to get benefit of recommendations [6].

- *Synonymy.* It means an item is signified with two or more different names or entries having alike meanings. In such cases, this technique is unable to classify whether the terms represent different items or the same item [6].

- *Shilling attacks.* This attack occurs when a malicious user starts giving false ratings in order to increase or decrease popularity of an item. Such biased ratings negatively affect collaborative filtering algorithms [6].

- *New user.* New users don't get correct recommendations as ratings are not sufficient when they arrive on website [14].

Knowledge Based Filtering.

In this type of recommendation system, domain knowledge is used to identify user preference. In order to achieve this recommendation system needs to employ three types of knowledge – knowledge about the item, knowledge about the user and need matching between user and item. In other words, knowledge-based filtering system is applied based on aggregated knowledge of the ‘learners’ and the ‘learning material’ [14].

Merits.

- *New item.* New items can be recommended immediately before being rated by significant users [20].

- *Grey sheep.* User gets classified into one group. Thereby, gets accurate recommendation [20].

- *Domain knowledge.* Noise free domain knowledge leads to recommendations that are accurate and reliable [20].

Demerits.

- *Knowledge Engineering.* It requires some knowledge engineering to find out the products which matches user requirements [14].

- *Static suggestion.* These systems lack the ability of learning from user preferences. So, user has to provide its preferences every time in order to get recommendation [15].

Hybrid Based Filtering System.

This technique integrates different methods in order to provide resolution to the limitations witnessed by each technique individually. The idea is to combine two techniques which gives more precise recommendations than single technique as disadvantage of one algorithm will be overcome by another [17, 24].

Merits.

- *No cold Start.* It is inefficient in recommending items not being rated by the user [21].

- *No sparsity.* The problem that majority of the users do not rate maximum of the items and consequently the ratings matrix becomes very sparse is not present in this technique [21].

- *Robust.* If recommendation from one technique fails, it can switch to the other one thereby building a robust system [21].

Demerits.

- *Complexity and expensive.* Since this technique merges two or more techniques, the result system being complex and implementation becomes expensive [22].

Sentimental Product Recommendation.

This technique is based on features. For each feature, we extract the sentiment by looking in a sentiment lexicon that list sentiment words. We then form an opinion pattern and calculate the frequency of these extracted patterns. At the end frequently appearing features are noted. Here, stochastic learning algorithm is implemented to analyze reviews, ratings, and emoticons [18].

Merits.

- *Improve customer service.* Manage complaints to avoid leaving customers feeling ignored and angry [23].

- *Lead generation.* Loyal and happy customers will bring new customers [23].

Demerits.

- Irony, humor or sarcasm are discernible [23].

III. CONCLUSION AND FUTURE WORK

Recommender systems made a significant progress over the last decade when numerous Content, Collaborative and Hybrid methods were proposed and several algorithms have been developed. However, despite all these advances, the current generation of recommender systems discussed in this document still requires further improvements to make recommendation methods more effective in a broader range of applications. In this document, we reviewed various limitations of the current recommendation methods and discussed possible extensions that can provide better recommendation capabilities. We further propose to develop an algorithm that solves the limitations faced by above developed recommender systems hoping that the issues presented in this paper would advance the discussion in the recommender systems community about the next generation of recommendation technologies.

REFERENCES

1. Suresh K. Gorakala and Michele Usuelli, Building a Recommendation System with R, Packt Publishing, 2015.
2. Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, "Collaborative Filtering Recommender Systems," in Foundations and Trends in Human-Computer Interaction, ACM, vol. 4, no. 2, pp. 81-173, 2011
3. Sneha Khatwani and M.B. Chandak, " Building Personalized and Non-Personalized recommendation systems," in Proceedings of IEEE International Conference on Automatic Control and Dynamic Optimization Techniques, 2016, pp. 623-628
4. Jugendra Dongre, Gend Lai Prajapati and S. V. Tokekar, "The role of Apriori algorithm for finding the association rules in Data mining," in Proceedings of IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, 2014, pp. 657-660

5. Karthiya Banu.R, Dr. Ravanan.R and Gopal.J, "Analysis and Implementation of Association Rule Mining," in Proceedings of IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, 2014, pp. 657-660
6. Najdt Mustafa, Ashraf Osman Ibrahim, Ali Ahmed and Afnizanfaizal Abdullah, "Collaborative filtering: Techniques and applications," in Proceedings of IEEE International Conference on Communication, Control, Computing and Electronics Engineering, 2017
7. Xiao-Feng Gu, Xiao-Juan Hou, Chen-Xi Ma, Ao-Guang Wang, Hui-Ben Zhang, Xiao-hua Wu and Xiao-Ming Wang, "Comparison and improvement of association rule mining algorithm," in Proceedings of IEEE 12th International Computer Conference on Wavelet Active Media Technology and Information Processing, 2015, pp. 383-386
8. Yu Cheng and Ying Xiong, "Research and Improvement of Apriori Algorithm for Association Rules," in Proceedings of IEEE 2010 2nd International Workshop on Intelligent Systems and Applications, 2010
9. Joeran Beel, Bela Gipp, Stefan Langer and Corinna Breitingner, "Recommender Systems: A Literature Survey," in International Journal on Digital Libraries, Springer, vol. 17, no. 4, pp. 305-338, 2016
10. Teguh Bharata Adji, Zainil Abidin and Hanung Adi Nugroho, "System of negative Indonesian website detection using TF-IDF and Vector Space Model," in Proceedings of IEEE International Conference on Electrical Engineering and Computer Science, 2014, pp. 174-178
11. Ruisheng Zhang, Qi-dong Liu, Chun-Gui, Jia-Xuan Wei and Huiyi-Ma, "Collaborative Filtering for Recommender Systems," in Proceedings of IEEE Second International Conference on Advanced Cloud and Big Data, 2014
12. Liu Xiaojun, "An improved clustering-based collaborative filtering recommendation algorithm," in Cluster Computing, Springer, vol. 20, no. 2, 2017
13. Kebin Wang and Ying Tan, "A new collaborative filtering recommendation approach based on naive Bayesian method," in Proceedings of ACM Second international conference on Advances in swarm intelligence, 2011, pp. 218-227
14. John K. Tarus, Zhendong Niu and Ghulam Mustafa, "Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning," in Artificial Intelligence Review, Springer, vol. 50, no. 1, pp. 21-48, 2018
15. A. Felfernig and R. Burke, "Constraint-based Recommender Systems: Technologies and Research Issues," in Proceedings of ACM 10th international conference on Electronic commerce, 2008
16. Anna Gatzoura and Miquel Sánchez-Marrè, "A Case-Based Recommendation Approach for Market Basket Data," in Proceedings of IEEE Intelligent Systems, 2014, pp. 20-27
17. Emre Sezgin and Sevgi Özkan, "A Systematic Literature Review on Health Recommender Systems," in Proceedings of IEEE E-Health and Bioengineering Conference, 2013
18. Saman Shishehchi, Seyed Yashar Banihashem and Nor Azan Mat Zin, "A proposed semantic recommendation system for e-learning: A rule and ontology-based e-learning recommendation system," in Proceedings of IEEE International Symposium on Information Technology, 2010
19. Anitha Anandhan, Liyana Shuib, Maizatul Akmar Ismail and Ghulam Mujtaba, "Social Media Recommender Systems: Review and Open Research Issues," in Proceedings of IEEE Access (Volume: 6), 2018, pp. 15608 - 15628
20. Muhammad Galih Wonoseto and Yusep Rosmansyah, "Knowledge based recommender system and web 2.0 to enhance learning model in junior high school," in Proceedings of IEEE International Conference on Information Technology Systems and Innovation, 2017, pp. 168-171
21. Bansari Patel, Palak Desai and Urvi Panchal, "Methods of recommender system: A review," in Proceedings of IEEE International Conference on Innovations in Information, Embedded and Communication Systems, 2017
22. Sonali R. Gandhi and Jaydeep Gheewala, "A Survey on Recommendation System with Collaborative Filtering using Big Data," in Proceedings of IEEE International Conference on Innovative Mechanisms for Industry Applications, 2017, pp. 457-460
23. Paolo Montuschi, Fabrizio Lamberti, Valentina Gatteschi and Claudio Demartini, "A Semantic Recommender System for Adaptive Learning," in Proceedings of IEEE IT Professional, 2015, pp. 50-58
24. Passi R., Jain S., Singh P.K, "Hybrid Approach for Recommendation System," in Proceedings of the 2nd International Conference on Data Engineering and Communication Technology, vol 828, Springer, 201