International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

An Analysis on Recommendation Systems in Machine Learning

Tarun Singh[1], Anees Ahmed Choudhury[2], Akash[3]

[1,2,3]*Student, Department of Computer Applications, Maharaja Surajmal Institute, GGSIP University, Delhi (India)*

Abstract - *With the increase in growth of the digital world and the influx of huge amount of information availability, a method to find the right choice in selecting and filtering an item to choose the right product from the pool of information has become necessary. To solve this problem, recommendations systems have been developed which can help the users in getting recommendations utilizing their previous searches and purchases. Recommendation systems are used in many areas which include and not limited to online shopping, news media, music, movie recommendation. They are highly deployed and utilized in companies like Amazon, Netflix, and YouTube.*

Key Words: Recommendation Systems, Filtering, Machine Learning, ML Algorithms, Collaborative Filtering, Content-based Filtering, Hybrid Filtering

1. INTRODUCTION

In this digital age we have all kinds of applications, software, and websites which have personalised content made to suit the needs, requirements, or tastes of their target users in one way or the other. This is where machine learning emerges in the picture. All the recommendation systems available online have been built using machine learning by deploying different learning methods, techniques and algorithms to achieve their required goals and objectives to match their user's perception's and needs for better availability of content and to provide potentially useful data that might meet the user's necessities and desires.

Apart from being useful for the users it is much more beneficial for the organizations, developers who deploy their software and applications to the users. It effectively helps them identify and classify their users through their data processed by their algorithms and recommend needful and resourceful information on products and services that stand to mutually benefit both the user and the organization itself. It can also be used as a feedback mechanism for the manufacturers and organizations as it lets them understand the needs and requirements of their users and lets them fund relevant research and innovations to improve the quality and functionality of their products and allows them to have a better grasp at the market for the consumers itself.

Such a system works by processing a huge amount of database collected through the application about the user and then an algorithm is developed which filters the necessary information through the database and then starts making relevant connections with the user and their data about their preferences and interests which further helps in making pertinent comparisons and links to the available services and products available online. There are multiple ways to accomplish this task and many techniques are used for the same. These techniques employ different and specific algorithms to accomplish their needs and classify the data into understandable, resourceful, and usable format which helps them identify important points of interest and match it with their content of available of products, applications, etc.

2. LITERATURE REVIEW

In this section, we will be discussing about the previous work that has been done in the field of recommendation systems in machine learning. A lot of work has been done by researchers which will help us in understanding and realizing the need of recommending systems in this modern era of technologies using machine learning.

G. Chidambaram, A. Dhanush Ram, G. Kiran, P. Shivesh Karthic, Abdul Kaiyum (2021) proposed a paper titled "Music Recommendation system using Emotion Recognition" where the proposed system was mainly aimed at detecting human emotions for developing emotion-based music players. They used the Deep Neural Networks (DNN) to study the relevant features abstracted directly from the data taken in an uncontrolled environment and managing the limitations of handcrafted features. The proposed system architecture comprised of 5 modules that include the client, the user, the server, the VGG16 model, and a Spotify API. The proposed system could detect the user's facial expressions and the model used in this system was VGG16 CNN. The proposed system had 7 types of emotions that includes Happy, Sad, Neutral, Disgust, Angry, Fear, and Surprised. The system worked by clicking the user's picture using the camera, and after that, the user's emotion is detected. Once the previous step is done, then next it plays the songs subject to the user's mood by applying the Spotify API [1].

Bushra Ramzan, Imran Sarwar Bajwa, Noreen Jamil, Riaz Ul Amin, Shabana Ramzan, Farhan Mirza, and Nadeem Sarwar

(2019) proposed a paper titled "An Intelligent Data Analysis for Recommendation Systems Using Machine Learning" where the proposed system helped the users in finding the hotels features matrix using polarity identification and it also recommends about the guest type for user's personalized recommendation. The proposed system used a common CF approach for recommendation using machine learning in which they had used the opinion-based technique of sentiment analysis to achieve a new hotel feature. The proposed system used this approach which included lexical, syntax, and semantic analysis that helped in understanding the sentiment of the user's asked hotel filters. The proposed system stated that the system provided accurate and quality recommendations to users [2].

Minal Zope, Tanmay Harshe, Sahil Dhanvij, Sanket Agarwal, Rajamala Jambhulkar (2021) proposed a paper titled "Food Recommendation System for Auto-Immune Diseases" where the proposed system helped the user's in recommending the types of food according to the disease they had. They built a user-friendly system for users for all over the world to minimize the inflammation by suggesting some anti-inflammatory food and recipes which could help people in curing their autoimmune diseases [3].

S Anusha, M Bindu, G Navya (2021) proposed a paper titled "Crop Recommendation System with Comparative Analysis of Different Machine Learning Algorithms" where the proposed system explained the use of ML algorithms used in predicting the crop according to the area with specific temperature, humidity, pH, rainfall, and potassium, nitrogen, phosphorus levels in the soil. The proposed system also conducted a comparative analysis of different machine learning algorithms to get the best accuracy. The steps involved in their system started by collecting statistical data, then doing the preprocessing of that data, then came building a model using various ML algorithms, and training them, and finally testing them. They had also used the graph for getting clearer results which further indicated that the accuracy of random forest and naive Bayes algorithms were high, and the accuracy of logistic regression and support vector machine algorithms were less when compared with the accuracy of other algorithms [4].

3. CLASSIFICATION IN RECOMMENDATION SYSTEMS

We can classify these recommendation systems into 3 types: Collaborative filtering, Content Based filtering, and Hybrid Filtering, which are discussed below:

3.1 COLLABORATIVE FILTERING

Collaborative filtering is a technique used in machine learning which helps in recommending items based on the user's idea or its opinion. Collaborative filtering requires a large set of data or a huge amount of user's information that is used in the process of filtering information using various algorithms and various techniques. Collaborative filtering helps in predicting what the user wants by enabling providers to select subsets of information to be displayed for the user's screen.

It recommends the items or information that is based on the user's previously purchased item or is based on his/her interest. For recommending items to its user, this technique requires active user's involvement, and his/her interests for active filtering. A classic example of this type of filtering is given below:

A user purchased an item from Amazon say, a book. Once, the user purchased that item and goes to the Amazon website and there he/she adds a review for his purchase and according to his/her added review, the system checks the person's added review with some other reviews of their liking. So, now, next time the user will find the recommended items of interest which haven't been liked by the user.

3.1.1 METHODS FOR COLLABORATIVE FILTERING

In Collaborative filtering, basically there are two methods for performing this technique and those are given below:

A. **User-based Filtering-** In this filtering method, the recommendation system recommends the item to its target user based on the user's previous purchasing history or preferences related to the same users as them. In this filtering, the system finds and uses the ratings of like-minded users who had given the same rating as the user to whom it is predicting the new items. In the Item-based filtering method, an algorithm is commonly used named Nearest Neighbor Algorithm. For example, if a user has never seen a movie but his/her peers had watched that movie and rated it, the recommendation system will recommend that same movie to this user based on its peer's observations.

B. **Item-based Filtering-** In this filtering method, the recommendation system recommends the item based on the ratings of an item but not from the ratings of the users. In this filtering method, recommendations for an item must be based on the

user's ratings for similar items [5]. For example, if a user A has rated an item positive, and others have also given the same rating, and the user A is expected to give a positive rating to the next item as well, but others haven't rated that item yet, so the system predicts that others should also give it a try.

3.1.2 APPROACHES FOR COLLABORATIVE FILETERING

1. **Memory-based-** It is a type of approach in which it stores the users' ratings, items, and its data to predict similar items. In Memory based approach a very standard method is used for Collaborative filtering known as Nearest Neighborhood Algorithms. There are two neighborhood algorithms i.e., Item-based, and User-based which are also discussed above.
   Looking at the Nearest Neighborhood Algorithms which predicts the items by calculating the weighted average of the ratings of the users, here, two approaches are used for finding the correlation of similarity which are discussed below:
   A. **Pearson Correlation-** It uses 2 users (say **p, q**) for finding the similarity by using the below formula: [6]

$$s(a,u) = \frac{\sum_{i=1}^{n}(r_{a,i} - \overline{r_a})(r_{u,i} - \overline{r_u})}{\sqrt{\sum_{i=1}^{n}(r_{a,i} - \overline{r_a})^2}\sqrt{\sum_{i=1}^{n}(r_{u,i} - \overline{r_u})^2}}$$

   B. **Vector Cosine-** It uses 2 items *u, v* for finding the cosine-similarity by using the below formula: [6]

$$s(\vec{u},\vec{v}) = \frac{\vec{u}\cdot\vec{v}}{|\vec{u}| * |\vec{v}|} = \frac{\sum_i r_{u,i} r_{v,i}}{\sqrt{\sum_i r_{u,i}^2} \times \sqrt{\sum_i r_{v,i}^2}}$$

2. **Model-based-** It is the approach in which it predicts the items based on the user's unrated items which are calculated from the similarity of the same users as them. As this is a model-based approach, so for this, models are built which can be done using two techniques i.e., data mining technique and techniques of machine learning.
   In model-based techniques, we have various techniques that include various CF algorithms like Cluster model algorithm, Bayesian model, Artificial Neural network, Decision Tree, Singular Value Decomposition, Matrix Completion Technique, etc.

3.1.3 PROS AND CONS OF COLLABORATIVE FILTERING

Now let's see some pros and cons of using collaborative filtering in recommendation systems.

Pros:

Collaborative filtering consists of some pros as compared to Content-based filtering in which the CF can also work in those conditions where there is less or difficult content for analyzing like user ideas and users' opinions. Another advantage of this technique is that it can provide recommendations that are suitable for the user even without the content being in the user's profile.

Cons:

If a technique has some pros, it also includes some cons as well, for instance, this technique cannot handle fresh items which can lead to a cold start problem (not having an adequate amount of information to recommend an item). Another con is that we can't add some more side features in query and items, and there are other cons like scalability, synonymy, etc [6].

3.2 CONTENT BASED

The information, products, or attributes of those information that a user interacts with on the internet can be termed as content and so content based recommendation system is based on the basic concept of finding links with the user desired content. Such a system uses such content and finds other similar 'content' that the user might be keen to explore and potentially valuable or beneficial in one way or the other. The similar content here can be referenced to any service or product available on the internet. For e.g., suppose that a user might watch a movie at a streaming service portal, the portal will store data about the user's searches such as the genres of the movies viewed such as action, sci-fi, animated, etc.,, the duration of movies viewed as well and the ratings given as well. The recommendation system will process this information and find similar movies with the same genres and durations and store it as potential interests for the user. These recommendations are viewed to the user and as the user selects and views more movies from these recommendations the recommendations get updated and keeps improving and providing more accurate and potential recommendations for the user. This is the overview of how most content-based recommendation systems work.

3.2.1 CONCEPTS AND TECHNIQUES

Such a recommendation system recommends web pages, news articles, publications, products, services, etc. to the users and so it requires a detailed analysis of features and characteristics of different content viewed by the user and so

it generates a different profile for each user [7] which makes it as personalised as possible. CBF also adds in the possibility that if the user profile is dynamically changing then the recommendations will change as well according to the user and so it is versatile and flexible and remains relevant to each user in the long term as well.

Several models are used to find out the relevancy of articles or content to deliver meaningful recommendations. A TF-IDF score [8] may be used to rank the content according to its similarity and then comparing it with the user profile by arranging them in an n-dimensional vector space. Such a model is called a Vector Space Model [9]. Other probabilistic models such as Neural Networks [10] or Decision trees [11] can also be used to attain a document relationship within a corpus.

There are quite a few benefits to using a content-based recommender system. A CBF can recommend unique items that may not be suitable for other users but the concerned user alone and so it has the potential to enthral all kinds of users and of all age groups. Since these recommendations are tailored for each person, other user profiles are not required as compared to other recommendation systems like Collaborative filtering system and so the potential for upscaling to a larger user base is always present. Also, it is possible to recommend newly released or fresh items which have not been reviewed to be recommended to suitable users. Moreover, it's possible to present recommendations for users without the need to share their profile and thus it has capability to insure privacy [12]. The effectiveness and quality of recommendation is dependent on the metadata present about the user preferences and so the amount and quality of the descriptive data is proportionate to the accuracy of the recommendation.

3.3 HYBRID SYSTEMS

After reviewing Content based and Collaborative filtering systems, the need for Hybrid systems becomes all the clearer. Everyone on the internet uses products from all around the world and so finding some inaccuracies in those items suggested to oneself or somebody else and so to resolve these inaccuracies a hybrid system is required. So, for the better output of the product and to avoid inaccuracy in the product the collaborative filtering method and content-based filtering are merged together. The combination of the two filtering systems gives a much more accurate and potential recommendation to the user. Using this technique can improve the down-factor of the individual systems in a combined model. The merging of the two approaches can be done in many of the following ways: separating implementation of algorithms and mixing of the result, utilizing one of content-based filtering in collaborative approach, making use of some collaborative filtering in content-based approach, creating a unified recommendation system that brings together both approaches. And so the use of a hybrid system can overcome the limitations of the other filtering systems. The different types of Hybrid filtering systems is listed below [13].

3.3.1 Weighted hybridization

There are various recommenders whose recommendations are combined and accommodate the scores from each technique by a linear formula. On this basis one example is the P-tango, which is having both type content-based recommenders and collaborative recommenders. It is executed by giving an equal amount of data at the initial stage but is not stable as it would change or get adjusted depending on the predictions being done. The benefit of using this technique is that the strengths of both recommendation systems are utilized greatly during the recommendation process [14].

3.3.2 Switching hybridization

In this type, the system would change or interchange to one of the recommendation techniques as per the guess of recommender ability to produce a good rating. This method can underestimate the problems specific to a single method/way. As the problem in content-based recommender are swapped to the collaborative recommendation system. The advantage of this is that it is responsive to the strong points and weak points of its recommenders. But the disadvantage of this is that mostly it injects complexity into the recommendation process. The reason for this is that the swapping process increases the number of frameworks to the recommendation system. A good example of this model is Daily Learner. It uses both system content-based and collaborative types. The content-based is deployed at the beginning and if it doesn't work for any reason then the second method is initiated automatically [15].

3.3.3 Mixed hybridization

This type of recommendation system combines recommendation results of different techniques at the same time. Each recommendation system has many recommendations linked to it from different techniques. In this type, the individual preference does not affect normal performance. A real-life example of this is the PTV system, where a TV schedule is recommended for the user which is combined with content-based and collaborative systems. Also, some others who utilise this technique are Profinder and PickAFlick [16].

3.3.4 Feature-combination

In this type, the features of one recommendation system is transferred to another technique. For example, suppose a user has rated something which is collaborative filtering which will be used to identify the similarity between the ratings. A real-life example of this is the Pipper which is using a collaborative filter's rating for the recommendation of movies [5]. The merit of using this technique is that it does not always just rely on collaborative information [17].

3.3.5 Meta-level

In this model there is an internal model generator whose recommendation is used as input for the others. The generator always has more amount of information as compared with the single rating technique. The problem of sparseness in collaborative methods is solved in this method (Meta-level) as it's using the input of the first technique for the second technique. The real-life example of the meta-level technique is Labour, which uses learning on the spot to create a content-based user portfolio that is then compared collaboratively [18].

4. CONCLUSION

The need for recommender systems has become indispensable and therefore is now used in most aspects of online platforms. Such systems have made the internet more immersive, useful, and time saving and effortless to browse through the selections and make an informed choice from the palette of recommendations provided by such systems. However, as we have discussed several limitations of all these techniques still exist today and so by discussing such constraints, we wish to highlight theses inadequacies and hope for future research to overcome these challenges and devise many more algorithms with better efficacy and accuracy and present a more superior system in every than we have today.

REFERENCES

[1] G. CHIDAMBARAM, A. DHANUSH RAM, G. KIRAN, P. SHIVESH KARTHIC, ABDUL KAIYUM "MUSIC RECOMMENDATION SYSTEM USING EMOTION RECOGNITION" IRJET (2021)

[2] Bushra Ramzan, Imran Sarwar Bajwa, Noreen Jamil, Riaz Ul Amin, Shabana Ramzan, Farhan Mirza, Nadeem Sarwar, "An Intelligent Data Analysis for Recommendation Systems Using Machine Learning", *Scientific Programming*, vol. 2019, Article ID 5941096, 20 pages, 2019. https://doi.org/10.1155/2019/5941096

[3] Prof. Minal Zope1, Tanmay Harshe2, Sahil Dhanvij3, Sanket Agarwal4, Rajamala Jambhulkar "Food Recommendation System for Auto Immune Diseases" IRJET (2021)

[4] S Anusha1, M Bindu2, G Navya "Crop Recommendation System with Comparative Analysis of Different Machine Learning Algorithm" IRJET (2021)

[5] Schafer J.B., Frankowski D., Herlocker J., Sen S. (2007) Collaborative Filtering Recommender Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_9

[6] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal, Volume 16, Issue 3, 2015, Pages 261-273, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2015.06.005.

[7] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowledge-Based Systems, Volume 46, 2013, Pages 109-132, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2013.03.012.

[8] Lukáš Havrlant & Vladik Kreinovich (2017) A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation), International Journal of General Systems, 46:1, 27-36, DOI: 10.1080/03081079.2017.1291635

[9] Venkat N. Gudivada, Dhana L. Rao, Amogh R. Gudivada, Chapter 11 - Information Retrieval: Concepts, Models, and Systems, Editor(s): Venkat N. Gudivada, C.R. Rao, Handbook of Statistics, Elsevier, Volume 38, 2018, Pages 331-401, ISSN 0169-7161, ISBN 9780444640420, https://doi.org/10.1016/bs.host.2018.07.009.

[10] Bishop, Christopher M. "Pattern recognition." Machine learning 128.9 (2006).

[11] Duda, Richard O., and Peter E. Hart. *Pattern classification*. John Wiley & Sons, 2006.

[12] Shyong, K., Dan Frankowski, and John Riedl. "Do you trust your recommendations? An exploration of security and privacy issues in recommender systems." International Conference on Emerging Trends in Information and Communication Security. Springer, Berlin, Heidelberg, 2006

[13] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal, Volume 16, Issue 3, 2015, Pages 261-273, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2015.06.005.

[14] Souvik Debnath, Niloy Ganguly, and Pabitra Mitra. 2008. Feature weighting in content based recommendation system using social network analysis. In Proceedings of the 17th international conference on World Wide Web (WWW '08). Association for Computing Machinery, New York, NY, USA, 1041–1042.
DOI:https://doi.org/10.1145/1367497.1367646

[15] Andrea, Jonathan & Bindner, Philippe & Besdel, Patrick & Bournat, Marc. (2013). Switch Hybridization: Micro-Arcing Model and Color Analysis. Proceedings of the 59th IEEE Holm Conference on Electrical Contacts, HOLM 2013. 1-7. 10.1109/HOLM.2013.6651429.

[16] Shah, Lipi & Gaudani, Hetal & Balani, Prem. (2016). Survey on Recommendation System. International Journal of Computer Applications. 137. 43-49. 10.5120/ijca2016908821.

[17] Xu-Cheng Yin, Chang-Ping Liu, Zhi Han, Feature combination using boosting, Pattern Recognition Letters, Volume 26, Issue 14, 2005, Pages 2195-2205, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2005.03.029.

[18] Vilalta, R., Drissi, Y. A Perspective View and Survey of Meta-Learning. Artificial Intelligence Review 18, 77–95 (2002). https://doi.org/10.1023/A:1019956318069

BIOGRAPHIES

Tarun Singh

Final year student pursuing a Bachelor of Computer Application from Maharaja Surajmal Institute, Delhi. His current research interests include Machine Learning, Mobile Development.

Anees Ahmed Choudhury

Currently pursuing Bachelor of Computer Applications at Maharaja Surajmal Institute, New Delhi and learning different computer technologies at the time with keen interest towards networking, programming and software development.

Akash

Currently pursuing BCA course from MAHARAJA SURAJMAL INSTITUTE from GGSIPU and keen interest in the various technologies used in the world, and mostly of the Machine learning.