NIET Journal of Engineering & Technology (NIETJET)
Volume 12, Issue 1, 2021 ISSN: 2229-5828 (Print)

Study on Movie Recommendation System Using Machine Learning

Aditya Narayan Singh[1], Rahul Kumar Sharma[1], Aman Agarwal[1], Amarendra Mishra[1] Mukul Pundhir[1], Anant Vijay[1]
[1]Department of computer science& Engineering, Noida Institute of Engineering and Technology, Greater, Noida 201306, India.

Abstract : ***As network technology and entertainment production have evolved, the types of movies available have become increasingly diverse, leaving viewers bewildered as to how to choose among them. The recommendation system was intended to make the selection process more efficient. Users are given product or service recommendations by these systems. These types of solutions have also improved the quality and efficiency of decision-making. A variety of machine learning algorithms may be used to create the recommendation system. Among these methods, we're using Content Based Filtering, Collaborative Based Filtering, k-mean clustering, and the Naive Bayes classifier. The experimental investigation with a cold start revealed the potential of all of these strategies by demonstrating a large improvement in accuracy.***

Keywords - Recommended System, Collaborative Based, Content Based, Cluster, Correlation.

1. INTRODUCTION

Recommendation systems are computer-based systems that provide suggestions to individuals based on a number of factors. Based on the user's data as well as other factors such as preferences and interests, the recommendation system chooses the most important information from a large quantity of data. In order to give suggestions, it determines the user-item match and infers the similarity between users and items [1]. Opinion mining focuses on attitudes, whereas traditional text mining focuses on facts [3]. The main research topics include sentiment classification, feature-based sentiment classification, and opinion summarizing. Sentiment classification investigates how individuals feel about a certain object. Feature-based categorization [4] is concerned with categorizing and assessing items based on their properties. Opinion summarizing varies from standard text summarizing in that it focuses entirely on the product aspects about which customers have expressed their opinions, rather than assessing a piece of a review and rewriting some of the original words to capture the main idea.

The Movie Recommendation System recommends films that are similar to ones you've already seen. There's also a collaborative content screening system that makes recommendations based on what other users have seen. More information about people and/or goods is used in the content-based approach. Item characteristics are used in this form of filtering to suggest other things that are comparable to the one you're looking at. One of the problems with sentiment analysis is that a phrase that has a positive connotation in one context may have a negative connotation in another. Second, different people react to the same circumstance in various ways. In classical text processing, a little change in the meaning of two bits of text has no impact. "It's a good picture" is not the same as "It's a good image."

2. ALGORITHMS

Content Based Filtering: The Content Based Filtering system takes into account the articles that a user has evaluated in order to make future recommendations when surfing the web. A person is more likely to give anything a positive or negative rating if it is something he enjoys or dislikes. The replies to the question have an impact on his evaluations. He gives anything a higher grade if he likes it, and a lower rating if he doesn't. These graded chunks are used as the 'content' in Content Based Filtering. Based on this data, the user is given recommendations for future things he could enjoy. The client is presented films that fall into a certain genre that he enjoys [3].

Input: users X, movies m, rating r, movie genre mg , Number of movies to be recommended(μ).
Output: Recommended movies R

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

1. for all users do
2. Select seen movies s, unseen movies s', association of unseen movies asi' w.r.t X, association of each genre agj w.r.t s' , where i is 1 to n and j is 1 to m.
3. Calculate scorej .
4. Select highest three scorej
5. Select m' ⊂ s' according to highest three score j
6. Calculate score me' where e ∈ m'
7. Return top μ score recommendations.
8. end for

The notations used in this method have the following meanings: The entire number of users who rated movie I s' is represented by the association of each movie as, and the total number of movies belonging to genre j is represented by the association of each genre age.
Score j = ag j / m
score(me') = ame / total count of m'

Collaborative Filtering: Many individuals are likely to rate something in the same way that the user intended. Collaborative Filtering uses the same format as their user evaluations. The concept of Collaborative Filtering [4] is to propose a product based on the preferences of others who share similar interests.: Many individuals are likely to rate something in the same way that the user intended. Collaborative Filtering uses the same format as their user evaluations. The concept of Collaborative Filtering [4] is to propose a product based on the preferences of others who share similar interests.

Collaborative Filtering Algorithm 2
Inputs: users X, movies m, rating r, and the number of movies to suggest().

1. Do this for all users
2. Choose between s seen movies' and unseen movies'.
3. Find simi (similarity) with respect to s, where I = 1 to n.
4. Choose the simi user with the highest score.
5. Choose m' s of the user from step 4 and s' of the ith user.
6. Determine the weight W(me').

K-Mean Clustering: The k-mean classification approach is a non-parametric classification technique. It divides the objects into k clusters based on how close they are to one another. In this study, the Euclidean distance is employed to assess how close they are. We utilised binary rated and unrated movies to calculate the Euclidean distance. The centroid of a cluster is the average of all its objects. With each iteration, all members of a cluster go closer to the centroid, which is updated. Iteration continues until the centroid does not change any more, at which point it stops. The search space is decreased as a result of this strategy, resulting in lower processing complexity. These calculations are done in the background, which gives you more freedom.

Users X, movies m, rating r, number of movies to suggest, k value
R is a film that is highly recommended.

1. get started
2. Pick k centroids at random.
3. From k centroids, calculate the euclidean distance (eucd) for X.
4. Use eucd to assign X to the kth cluster.
5. Use (summation(ki) from 1 to p)/p to update the centroid for each cluster, where p is the number of members in the ki cluster.
6. Repeat steps 3–5 until the centroid(t) equals the centroid (t+1).
7. Do it for all users
8. Select'seen movies s' and 'unseen movies s' from the drop-down menu.
9. Find simi (similarity) with respect to s, where I = 1 to p.
10. Choose the simi user with the highest score.
11. choose m' s from the highest simi and s' from the ith user.
12. Calculate weight W (me') where e C m'

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

13. Return top μ weight recommendations.
14. end for
15. end

Naive Bayes: The Naive Bayes technique is built on the Bayes theorem. When it comes to determining classification probability and recognising model ambiguity, the Naive Bayes Classifier employs a probabilistic approach. It's a quick-learning technique that relies on prior data understanding. The basic premise is that the attributes are conditionally independent [6].

Naive Bayes Algorithm
Users X, movies m, rating r, number of films to suggest()
Recommendations for movies R.

1. Do this for all users
2. Select seen movies s and unseen movies s' from the drop-down menu.
3. Find simi (similarity) with respect to s, where I = 1 to n.
4. Choose x' X where simi is more than ten.
5. Determine the relationship between unseen films and x'
6. Determine the score (se').

Future Scope

For the past decade or two, recommender systems have been a prominent research area, and they continue to amaze. Despite the fact that recommender systems have come a long way since their humble beginnings with content-based and collaborative filtering techniques, there is now a lot of research being done to enhance output accuracy and recommender systems in general [7]. The search is focused on a few areas to enhance the RS's usability and applicability in real-world circumstances. The following are some of the RS areas where significant research is now being conducted, and these efforts will likely influence the future of recommender systems [8].

Conclusion

On the Internet, recommender systems expand the reach of customized information retrieval. It also aids in the reduction of information overload, which is a key issue with information retrieval systems, by allowing users to access goods and services that aren't readily available to other users. We create a system that is tailored to the user's individual preferences, and we provide suggestions based on his previous evaluations. This strategy contributes in improving the accuracy of recommendations. Each user has their own profile, which contains information about themselves, likes, ratings, comments, and password management options. It also helps collect correct and authentic data, which improves the system's trustworthiness.

References

[1] Tewari A.S., Kumar A., and Barman A.G.,. Book Recommendation System Based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining, IEEE, 978-1-4799-2572-8, pp 500-503, 2014.
[2] Wanaskar U.H., Vij S.R., and Mukhopadhyay D. A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm, Journal of Software Engineering and Applications, 6, pp 396-404 2013.
[3] Shinde U., and Shedge R. Comparative Analysis of Collaborative Filtering Technique, IOSR Journal of Computer Engineering, Volume 10, pp 77-82 2013.
[4] CHENG X., WANG J., Danqian LU. Research of Question Analysis Based on HNC and K Nearest Neighbor, Journal of Computational Information Systems, 6:10, pp 3449-3455, 2010.
[5] Campos P.G., Bellogín A., Díez F., and Chavarriaga J.E. Simple Time-Biased KNN-based recommendations, ACM, 978-1-4503-0258-6, 2010.
[6] Puntheeranurak S., and Pitakpaisarnsin P. Time-aware Recommender System Using Naïve Bayes Classifier Weighting Technique, International Symposium on Computer, Communication, Control and Automation, 3CA, pp 266-269 ,2013.
[7] Bellogín A. Castells P. and Cantador I. Precision-Oriented Evaluation of Recommender Systems: An Algorithmic Comparison, ACM, 978-1-4503-0683-6, 2011.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

[8] Puntheeranurak S. and Sanprasert S. Hybrid Naive Bayes Classifier Weighting and Singular Value Decomposition Technique for Recommender System, IEEE, 978-1-4244-9698-3, pp 473-476, 2011.
[9] Rohan Kumar,Shyam Lal, Sudhir Kumar; 'A Review on Manufacturing of Composite Materials by Electromegnetic Stir casting Method',Volume No.1,Issue No.2,2013,PP.033-039,ISSN :2229-5828
[10] Rajiv Ranjan, V.K.Giri; 'ECG Signal Enhancement Using Wavelet Transform',Volume No.1,Issue No.2,2013,PP.040-045,ISSN :2229-5828
[11] Pali, Harveer Singh; 'Performance Charaterstics of Biodeisel Blend in CI Engine using Artificial Neural Network,(Karanja Oil)',Volume No.1,Issue No.2,2013,PP.046-053,ISSN :2229-5828
[12] Gyanchandra Sharma, Shashi Prakash Dwivedi and Ashok Kumar Yadav; 'Analysis of Four Stroke Single Cylinder Compression Ignition Engine Operated Mith Blends of Waste Cooking Oil Biodiesel/Diesel',Volume No.1,Issue No.2,2013,PP.054-058,ISSN :2229-5828
[13] Deepak Kumar Tyagi,Anuraag Awasthi, Ritesh Rastogi; 'Analysis and Design of Metrics for Autonomic Computing',Volume No.1,Issue No.2,2013,PP.059-063,ISSN :2229-5828
[14] Sudev Das Sagnik Pal, Harinder Singh and Swapan Bhaumik; 'Correlation for Prediction of Heat Transfer Coefficient for Pool Boiling Using TiO2 Nanofluid',Volume No.1,Issue No.2,2013,PP.064-066,ISSN :2229-5828.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.