Original Article

Movie Recommendation System Using Deep Learning Techniques

S. Shaik Salam[1], Dr.P. Dhanalakshmi[2], Dr.N. Laksmipathi Anantha[3], Dr.A. Sandeep[4]
[1]Associate Professor, Dept of CSSE, SVEC, A. Rangampet.
[2]Professor, Dept of CSSE, SVEC, A. Rangampet.
[3]Professor, Dept of CSE, GCET, Hyderabad.
[4]Senior Project Manager, Infosys, Hyderabad.

Abstract

Due to increase of data in online web resources Recommendation System plays a vital role. In many platforms such as music, movie, books, videos, ecommerce for recommending the useful data. Video recommendation system provides users a very highly satisfying recommendations which in turn increase the user stickiness to the site. A personalized recommendation system with collaborative filtering technique is implemented to increase the user satisfaction. This method is used to recommend movies to users depending on their historical data on the streaming site and also by exploring the content watched by the users with similar preferences. A personalized recommendation system using collaborative filtering with multi-layer perceptron is implemented to increase the accuracy of the system.

Keywords: Deep Learning Techniques, Multi-layer Perceptron, Preferred Movies.

DOI: 10.47750/pnr.2022.13.04.144

INTRODUCTION

The e-commerce market is growing worldwide with the development of information technology and the popularization of mobile devices, various types of products are released on a daily basis. Hence people suffer from time consuming as people need to surf through the whole site to get the product they need. Significantly, the issue of information overload leads to dissatisfaction of the user. Therefore, there is a need for a personalized recommendation system. This is not new to the industry the staff sales to retail stores that is recommended to the customers for the purpose of upselling and cross-selling, and this maximize their profits hence the same strategy is used. The recommendation system can be defined as a model which tries to predict the user preferences and to provide suggestions based on them. The main objective of the any recommendation systems is to achieve customer satisfaction by providing relevant recommendations and increasing the time spent by a user on the site. Personalized recommendations help users to easily get what they wanted by seeing recommendations, also help in sales increase, to retain customers, and improve their shopping experience. Netflix, YouTube, Amazon, e-commerce, are all examples for personalized recommendation systems. Streaming applications like Netflix, Amazon Prime usage is increased by the consumers who enjoy the video content.

Many applications used in daily life use personalized recommendation systems for recommending books, videos, movies, news. Particular characteristics of the datasets. Movie or video Recommendation systems are very powerful and important systems. Movie Recommendation systems help the user to look for the movies that they may like and also reducesthe time to find their preferred movies.

The first step here is to see the movies that the user watched and the pages that the user visited in the past. Then these are analyzed and then the recommends movies according to the user history. There are two filtering techniques to recommend videos to the users. One is a content- based filtering approach and the other one is collaborative filtering technique.

Content-based filtering is one of the recommendation approach which tries to predict what the user might like by taking input as user's activity. Content based system filters by keywords and attributes assigned to site content and by matching them to the individual user profile.

Collaborative filtering is also known as social filtering. Here this filtering technique used is to filterthe data from the past users based on their reviews it is used to make personalized recommendations for the new users with similar preferences.

Hybrid approaches are done in a number of ways: demonstrate by combining the outputs of content-based and interactive collaborative outputs and integrating them; by adding content- based skills to a collaborative-based approach or by merging methods into a single model. Several studies strongly compare hybrid performance with pure content-based and content-based methods and have shown that mixed systems can provide the accurate recommendations than non-mixed approaches. These methods eliminate cold start and sparsity problem as they are the most commonly faced issues. Inputs are taken and sent to the collaborative and content recommendation systems then the outputs are combined in the combiner then recommendations are made using them.

To enhance the performance of the recommenders and to make it much more effective the DNN (deep neural networks) are brought into picture. Deep learning makes the recommendations more effective and efficient. Several recommendation approaches like content, collaborative and knowledge-based filtering techniques can be combined with deep learning techniques which improve their presentations using other dn approaches. The approaches include auto encoder, Convolution neural network, multilayer perceptron.

LITERATURE SURVEY

F. Furtado et.al [1] proposed a web-based collaborative filtering movie recommendation system. This approach computes the connection between different clients using similarity cosine i.e collaborative filtering This technique depends on ratings to recommend movies to users who have similar tastes and also allows users to explore more. It is web application so it displays all the features provided to the users and by clicking the generate recommendation button it will recommend movies based on previous ratings. If the user is a new user and has not rated any movies then user is required to search a random movie on their interest and to give ratings to atleast six movies. So that their interest will be known. This approach works for limited users and depends on ratings.

Ehsan Aslaniana et.al [2] developed a unique technique to suggest things to users supported Content Feature Relationship victimization hybrid filtering approaches that may be a combination of content-based and item-based cooperative filtering. This technique extracts the connection between content options. There are 2 intuitive justifications for this approach. The primary is though Associate in item content options are the supply of preference for a user towards that item, not all of the content options have a similar result. The second is that content options are semantically associated with one another. A second matrix is employed to model content options relationships. To coach the content options relationship matrix from the rating information and also the method of least squares method is employed. This technique is barely appropriate for restricted information and extremely rely on past info of things.

Anmol Chauhan et.al [3] proposed a movie recommendation system using sentiment analysis. This approach records the sentiment of the user in the form of good and bad. If user likes the movie they can give good smiley and if the experience is not good they can give sad reaction based on this movies are recommended and in order to recommend the movies the KNN collaborative filtering algorithm is used. The motive work of algorithm is to recommend movies that are nearest neighbor as if most of the movies that are close to the movies that it belongs to will be put or comes under the same cluster in which distance of the movie shortest to the neighbor movie. After finding the near neighbor movies the similar movies based on users past history are filtered and recommended. But this approach come up with low accuracy and high running time. These approaches only applicable to static data or for limited instance and has low efficiency and accuracy. To overcome the issues in all techniques a novel technique KNN is used[4,5]. This systemtakes input from user's history and ID of watched videos and output will be a video set is recommended based on content of the information

EXISTING SYSTEM

Video recommendation system is extremely helpful in user's day to day life to cut back user time and problem for looking a selected pic because the user desires. The KNN is one in every of the foremost necessary techniques to suggest the films. the most motive of KNN rule is to search out nearest neighbour as if the foremost of thing that are on the brink of the item that's belongs to are going to be place or comes below an equivalent cluster during which the gap of the item is shortest to the neighbour and similar content items.

Step 1: Import knowledge from movies lens repository.
Step 2: Pre-process the info to get rid of ambiguity.
Step 3: Calculate the user's historical activity from user IDs of movies watched by the user.
Step 4: Generate user because 2nd matrix of the ratings within the sort of Rmxn.
Step 5: Apply the formula of circular function Similarity that identifies the similar things to the users WHO need to observe then, it'll generate the matrix that's similar to the user read.
Step 6: Prediction of high rated neighbor things of comparable content.
Step 7: Calculate prognostic accuracy metrics for analysis of performance.

First import.csv files for pre-processing to get rid of redundant and ambiguous knowledge. Next users' activity history is obtained from IDs of movies being watched. During this step victimization user historical watch data makes it straightforward to find videos that user watched recently. And then a user 2nd matrix is generated during which videos as columns and user as rows wherever ratings given to every pic by user as values. After, circular function Similarity is applied on the generated matrix wherever comparison on content {and identifies the neighbor items belongs to same cluster/content of things. From the results of circular function similarity, the extremely rated videos of comparable content are expected. At last the performance of system the prognostic accuracy, precision, recall and MRR are calculated.[6,7]

The working principle of existing system is defined in its modules\phases. The existing system hasfour modules:

1. Data Preprocessing
2. User History
3. Content Behaviors
4. Prediction/Recommendation module

Data Pre-processing

Pre-processing refers to all the transformations on the raw data where all the redundant, missing, null values are eliminated. The movie lens dataset is filtered in pre-processing phase. A good recommendation system starts with pre-processing phase. The Main motive of this phase is to filter the data and identify the ambiguous data.

User History

The section collects all the information about user history. Gets the previous activity history and the videos ID watched by the user. From the user history it will be easy to figure out the user's interest and similar movies with other users. Collecting the user history and analysing it helps in optimized and related result.

Content Behaviour

The existing recommendation systems primarily worked on content filtering measures (CFM). The content filtering live refers to a psychological feature filtering technique that engages in recommending things supported the weather of the things themselves. The information of the things and user info area unit taken into thought supported the content, that the user has viewed on the system. It recommends things supported similarity among the content of the things and a user profile outline. The content of each item is sense as a collection of attributes and words that arise in an exceedingly document. The user profile is sense with similar attributes and engineered up by evaluating the content of things that are viewed or watched by the user. Things that area unit typically like the fully rated things area unit counseled to the user. CFM employs numerous algorithms to search out the similarity among items/documents to come up with vital recommendations. It doesn't need the profile of additional users since they are doing not influence recommendations. And if generally the user profile changes, the CFM approach still will alter its new recommendations among a really.

Prediction Module

To predict and advocate the similar content of things the KNN technique is applied. K Nearest Neighbour formula falls underneath the supervised Learning class and is employed for classification (most commonly) and regression. It's a flexible formula conjointly used for imputing missing values and resampling datasets. because the name (K Nearest Neighbour) suggests it considers K Nearest Neighbours (Data points) to predict the category or continuous price for the new Datapoint. The principle behind K Nearest Neighbours. Here, nearest neighbours are those purpose |datum| information} s that have minimum distance in feature house from the actual new knowledge point. And K is that the range of such knowledge points we have a tendency to take into account in implementation of the formula. Therefore, distance metric and K price are 2 vital issues whereas victimization the KNN formula. Euclidian distance is that the most well-liked distance metric. Even playing distance, Manhattan distance, mathematician distance may be as per the necessity. For predicting class/ continuous price for a replacement datum, it considers all the info points within the coaching dataset. Finds new knowledge point's 'K' Nearest Neighbours (Data points) from feature house and their category labels or continuous values. Once finding the closest neighbours things filtering the similar things is that the vital step. Trigonometric function Distance is outlined as distance between 2 points in High Dimensional house. it's outlined because `the price equals to one – Similarity (A, B). Therefore, the vary of the trigonometric function Distance ranges from zero to one further. If the trigonometric function Distance is zero (0), whichmeans the things are identical. If the trigonometric function Distance is one (1) which means the things are undoubtedly completely different. Victimization trigonometric function similarity, the similar content things are filtered and suggested to the user.[7,8].

Existing system trust heavily on content of item and generating recommendations for various users and might handle solely restricted knowledge[9,10]. This requires systems to scale expeditiously. The projected system is in response to handle the problem of measurability and to boost recommendations [11]. The projected system maps users and things to a latent house wherever the similarity between users and their chosen things is maximized employing a Deep Learning technique.

```
movie selected:  Toy Story (1995) Index:  0
searching for recommendation......
0                                                  NaN
2353                              'night Mother (1986)
418                               Jurassic Park (1993)
615                Independence Day (a.k.a. ID4) (1996)
224              Star Wars: Episode IV - A New Hope (1977)
314                                Forrest Gump (1994)
322                              Lion King, The (1994)
910     Once Upon a Time in the West (C'era una volta ...
546                          Mission: Impossible (1996)
963                                        Diva (1981)
Name: title, dtype: object
```

Fig 1: output for existing system

Demerits

- Limited datasets can only be handled.
- Inflexible to the naïve users.

- Limited data availability in the movie description.
- Scalability.

Proposed System

Mostly the recommendations are based on Collaborative Filtering (CF), that is that the most typically used methodology for creating recommendations [12]. In Collaborative filtering approaches, users and things with similar patterns of rating area unit is taken under consideration to supply the recommendations for the particular user [13]. The fundamental entity in Collaborative Filtering is that the user-item ratings matrix, composed of collection of things I = i1,...in and a collection of users U = u1,...um. The ratings matrix R ∈ Rn×m contains the ratings given by users to things, wherever n represents range |the amount| the quantity} of things and m number of use.The motive of the planned system is to be ready to predict ratings for movies a user has not however watched. The movies with the best expected ratings will then be suggested to the user.

Step 1: Import data from movies lens repository.
Step 2: Preprocess the data to remove ambiguity.
Step 3: Generate respective user and movie matrices.
Step 4: Map user ID and movie ID to respective embedded matrix.
Step 5: Perform the dot product between the user vector and movie vector.
Step 6: Recommendation of users based on predicted ratings.

First import.csv files for pre-processing to remove redundant and ambiguous data. Next users' activity history is analysed. In order to improve recommendation, it should be made sure to test thedataset after training it with certain required attributes to see whether it obeys the constraints or not. After, the model of the system will be generated where further computations for recommendation are done. The embedded matrix is generated for respective user and items/videos in which types of movies, genre, ratings, no. of. views...etc are gathered and same for movie matrix [14].In both matrix ID of user and movie are defined to match respective user interest. so the user ID and movie Id aremapped based on interest of similar uses and ratings to the movie are predicted. The top 10 highlyrated movies are recommended to the user [15].

The implementation of the proposed system can be explained in the form of modules/phases. The proposed system comprises of four modules:

1. Data Preprocessing
2. User History
3. Candidate recommendation module
4. Ranking and Similarity

Data Pre-processing

Pre-processing refers to all the transformations on the raw data where all the redundant, missing, null values are eliminated. The movie lens dataset is filtered in pre-processing phase. A good recommendation system starts with pre-processing phase. The Main motive of this phase is tofilter the data and identify the ambiguous data.

User History

The section collects all the information about user history. Gets the previous activity history and thevideos ID watched by the user. From the user history it will be easy to figure out the user's interest and similar movies with other users. Collecting the user history and analysing it helps in optimized and related result.

Candidate recommendation module

Candidate recommendation module is that the part wherever implementation of the code takes place. CRM produces optimum and accurate results for records which are extremely related and suggests videos exploring preferences of users and allows the users to control data overload problem. This proposed technique looks the things or pictures which are seen by the user and likable by user within the past victimization cooperative filtering technique. Collaborative filtering (CF) is a method which will filtrate things that the user would like.

The approach is to form the clusters which is a group of items or movies and find out User interested movies or item using algorithms and recommend the movies to the users. In order to formthe clusters, the matrix factorization is performed. Matrix factorization is a collaborative filtering method to find the relationship between items and user's entities. While performing dimension of the matrix will be reduced according to the similarities between the movies.
Once all the user interested movies are found. The top 10 movies which are not watched by the user are recommended[17].

Ranking and Similarity

It is the ultimate part of the system wherever filtering the similar items\video on the prediction results of the higher than module\phase and predicting the very best hierarchic item\videos. The highest K-number of items\videos are suggested to the user. The similarity isn't calculated on based on victimization factors. The ranking is based on ratings given by the user to the movies at last, it's necessary to verify the experiment results. It's in measured within the kind of metrics referred to as accuracy, foreseen values...etc. the technique to find the performance accuracy is finding RMSE value. Root Mean Square error is the square root of the mean of the square of all the error. The use of RMSE is very common, and it is considered an excellent general purpose error metric for numerical predictions [16].

The proposed system provides customized recommendations that facilitate users notice prime quality videos relevant to their interests and tier of comfort and interacts with user in higher by recommending videos to observe that cater to their wants. Many of us have downside choosing the choice item of picture thanks to lack of your time and thanks to search problems. Also, picture fromfriends is time intense.

```
==========================
movies with high rating from user
------------------------
Fight Club (1999) : Action|Crime|Drama|Thriller
Dogville (2003) : Drama|Mystery|Thriller
Arizona Dream (1993) : Comedy|Drama|Fantasy|Romance
Idiots and Angels (2008) : Animation|Drama|Fantasy
Baby Driver (2017) : Action|Crime|Thriller
```

Fig 2: Movies with high rating from user

```
--------------------------------------------------
Top 10 recommendations for 125
--------------------------------------------------
Willy Wonka & the Chocolate Factory (1971) : Children|Comedy|Fantasy|Musical
Goodfellas (1990) : Crime|Drama
Shining, The (1980) : Horror
Out of Sight (1998) : Comedy|Crime|Drama|Romance|Thriller
Exorcist, The (1973) : Horror|Mystery
American History X (1998) : Crime|Drama
Meet the Parents (2000) : Comedy
Lord of the Rings: The Fellowship of the Ring, The (2001) : Adventure|Fantasy
Grave of the Fireflies (Hotaru no haka) (1988) : Animation|Drama|War
Lord of the Rings: The Return of the King, The (2003) : Action|Adventure|Drama|Fantasy
```

Fig 3: Top 10 recommendations for the user

Merits

- Less time consuming
- More accurate

Conclusion and Future Work

Personalized recommendation system is proposed for recommending the movies by considering ratings and user history for high user satisfaction. The system uses multi-layer perceptron and provides accuracy 0.32.

In future, this model can also be applied using hybrid recommendation systems on largedatasets.

References

Francesco Ricci and Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook Recommender Systems Handbook, Springer, 2011, pp. 1-35.

"Playboy Lead Rise of Recommendation Engines - TIME". TIME.com. 27 May 2010. Retrieved 1 June 2015.

Hosein Jafarkarimi; A.T.H. Sim and R. Saadatdoost A Naïve Recommendation Model for Large Databases, International Journal of Information and Education Technology, June 2012.

R. J. Mooney & L. Roy (1999).Content-based book recommendation using learning for text categorization. In Workshop Recom. Sys. Algo. and Evaluation.

Prem Melville and Vikas Sindhwani, Recommender Systems, Encyclopedia of Machine Learning, 2010.

ChenHung-Hsuan; Chen Pu (2019-01-09)."Differentiating Regularization Weights -- A Simple Mechanism to Alleviate Cold Start in Recommender Systems". ACM Transactions on Knowledge Discovery from Data (TKDD).13: 1–22.

Rubens, Neil; Elahi, Mehdi; Sugiyama, Masashi; Kaplan, Dain (2016)."Active Learning in Recommender Systems". In Ricci, Francesco; Rokach, Lior; Shapira, Bracha (Eds.). Recommender Systems Handbook (2 ed.). Springer US.

Karlgren, Jussi (October 2017)A digital bookshelf: original work on recommender systems". Retrieved 27 October 2017.

Shardanand, Upendra, and Pattie Maes. "Social information filtering: algorithms for automating "word of mouth"." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 210-217. ACM Press/Addison-Wesley Publishing Co., 1995.

Resnick, Paul, and Hal R. Varian. "Recommender systems." Communications of the ACM 40, no. 3 (1997): 56-58.

Lakshmi Haritha. M & K. Ramani. Impact of Deep Learning on Localizing and Recognizing Handwritten Text in Lecture Videos. International journal of Advanced Computer Science and Applications; 12(4), 2021.

Sakthivel M, Sivanantham S, Kamalraj R & Krishnamoorthy V (2022). An Analysis of Machine Learning Depend on Q-MIND for Defencing the Distributed Denial of Service Attack on Software Defined Network. International Journal of Early Childhood Special Education. 2022; 14(05): 3769 – 3776.

P. Lakshmi Sagar (2022).Resolving sets and Dimension in Bidiakis Cube and Durer Graphs. International journal of Neuro Quantology; 20: 4988-4992.

CH Prathima, R. Anusuya, M. Ram Kumar Prabhu (2022). Comprehensive Design Analysis of Digital Marketing in Agriculture Sector. International Journal of Early Childhood Special Education; 14(5):2022.

Silpa C, Ram Prakash Reddy Arava, K.K. Baseer. Agri Farm: Crop And F International Journal of Early Childhood Special Education ertilizer Recommendation System for High Yield Farming Using Machine Learning Algorithms; 14(5):2022.

Kamalraj R, Sakthivel M (2018). A hybrid model on child security and activities monitoring system using iot. In2018 International Conference on Inventive Research in Computing Applications (ICIRCA) 2018 Jul 11 (pp. 996-999). IEEE.

Siva Kumar Depuru & Dr.K. Madhavi (2019). Autoencoder Integrated Deep Neural Network for effective analysis of malware in distributed Internet of Things (IoT) Devices. The International journal of analytical and experimental modal analysis; 9(11): 226-232.