

Movie Recommendation System Comparative Methods

Team Details

Guruprasad N Bamane PES1UG19CS174 Section - C
S Usha Priya PES1UG19CS407 Section - G
Shamitha S PES1UG19CS448 Section - G

Ever in this busy world amidst the work throttle there has always been a need for the source of entertainment. This might be habitual or rarely occurring which might bring in a state of pleasant mindedness. There are lot many founts of entertainment and methods followed upon like films, music, dance, and video games, etc... The springs arising depends on a particular individual and varies from time to time. One among these could also be a habit of a person based upon their interest. Even though there are lot many existing forms of killing time, the one which definitely has its place is movies or films.

With an ever-increasing diversity and technology one can see many kinds of motion pictures being available. It could be something like a short film based on a real-world story, an illusionary , a prediction made, or something purely based for entertainment which is quite long. The classification may also go in accordance with the timeline the movie was released . The grouping can also be based on genre like Western, Horror, Action, Drama and so on.

People might prefer flicks based on the time which they can spend upon it, those with a heck of schedule may tender towards ones which have shorter duration. Some particular genre or a language can be attractive which again depends on perspective.

Existing in this physical world, it is troublesome and time consuming to surf among people or sites to decide upon which show to spend time on. One would rather like someone to suggest things based on their interest. But this wouldn't be a realistic approach due to huge number of options and the scarcity of time.

Solution which doesn't require a constant manpower to rely upon would be suitable. This is exactly what most of the applications and websites like Netflix, Amazon prime and hot-star do. This would always give raise to an eagerness of how exactly these technologies come up with the right recommendations.

It is important to understand how the recommendations are made and the underlying principles behind it. Recommendations basically fall under the category of unsupervised learning in machine learning, wherein a model is trained upon a dataset which itself categorizes the available data into categories based on similar features . This model later can be made to predict further views.

Even under this unsupervised learning there are different ways to make recommendations. Content-based Filtering(CBF) can be used to make recommendations solely on an individual's interest like genre, Collaborative Filtering(CF) can be used to make recommendation based on similarity between users, Knowledge based and

hybrid approaches. In depth in each of these classifications one can use various metrics for calculating numeric.

Though there are wide range of models and works done under this field, our main area of interest is to make comparison among possible methods and to come out with a conclusion on which method would mostly comply to the given situation. Our research mainly focuses on making recommendations using different methods available, draw insights between variables and compare the methodologies used.

Literature Study:

CF comprises of user-based as well as item-based filtering [1]. These two approaches are used to make recommendations on the Movie Lens dataset, taking rating as the only parameter into consideration for making recommendations based on the similarity among various users. Users who have same kind of pictures as interest and almost have the similar ideology would have given same range of rating. Performance metrics like Pearson's correlation coefficient, cosine similarity, Euclidean distance and Jaccard similarity are used. According to the experimental results item-based(0.84) CF led to greater accuracy compared to user-based filtering(0.76). Also, for diverse datasets Pearson's correlation is one of the best similarity measures.

Instead of using only CF, content based should also be used to make comparisons.

Recommendations are made based on the history and choices that a user make. [2] Content-based recommendation is used for sentiment analysis and CF is used to give recommendations. The model works on user ratings finding the similar ones which the user might like. K nearest neighbor(KNN) and cosine similarity are used as metrics for evaluation. CF has a

higher F1 score(0.772) compared to the F1 score of content based(0.528). Specifically collaborative method based on user-user has slightly higher mean squared error(0.258) compared to that of item-item approach(0.248). Visualizations like bar graph and heat maps are used to analyze top rated movies. The recommendation framework uses the user's minimum rating to assess the type of film to recommend. This is a function that has not been used in prior programs of a similar nature. KNN might not work for large datasets.

Movie recommendation system might suffer from problems like long-tail problem, Sparsity, and cold-start problem[3]. The technique used is CF with Pearson's correlation coefficient as a metric. Experiment is conducted on MovieLens-100k dataset. All the movies which have ratings less than the average i.e., average between 0 and maximum, are ignored. Visualization shows that there is negligible difference between the predictions with the entire data versus movies with below average ratings removed and hence can be removed. If the provided ratings by users were wrong, then it can neglect good movie.

Due to day by day increase in the amount of information as well as the census there is a compromise seen in movie recommendation systems. [4] Hybrid approach comprising of both CBF and CF would rather help improvement in the accuracy, quality, computing time and scalability. Support vector classifier and Genetic algorithm are being used. Three different datasets are used as a combination. Accuracy, precision, recall and F1 score are comparatively higher for a hybrid approach. There is a need to work on the memory requirements of the proposed approach.

A machine learning approach to recommend movies to users using K-means clustering algorithm to separate similar users and creating a neural network for each cluster[5]. Machine learning approaches are used to guess what rating a particular user might give to a particular movie so that this information can be used to recommend movies to viewers.

Have used PCA, clustering, pattern recognition network for classification.

System showed on average 95% accuracy depending on the cluster.

User rating, user consumption ratio and user preference have been considered while building the system. K-means clustering has been used to separate users with similar taste in movies.

Movies mostly are rated based on the ratings provided and upon the similarity of interests. But these days users also add reviews to movies they watch which may serve as an additional parameter to recommend movies.[6] Sentiment analysis is used on TMDB dataset using Support vector machine by giving polarity scores on each supporting word and later summing them up together. An assumption is being made that the analysis of comments would largely affect the recommendations made along with the rating scores given. SVM was able to give accuracy around 85 % on a dynamic dataset. Also, the model received 65-70% positive feedback from the real-world users. Scalability and infrastructure are referred to as future scope for making the system even more efficient. In this model only text-based sentiment analysis is used for making predictions, history of the users is ignored which could be taken into consideration.

In this global world there is an adamant increase in the amount of data getting generated and users on the other hand have no energy and time to spend on searching. Though there are many movie recommender systems there is also a need to accommodate large data in this domain.

Distributed cloud computing using map reduce framework would suffice greatly[7]. Recommendations are made based on the past views of a user i.e., item-based CF technique is used. Assumption as in the same framework can be used also in ecommerce.

Neural networks can also be used in movie recommendation systems. The model comprised of visual movie contents, Recurrent neural network(RNN) and Convolution neural network(CNN)[8]. Introduced heterogeneous multimodal SMR network which utilize multimodal movie contents, social network and , users' feedback to generate movie recommendations. Assumptions was made as ,by learning a user ranking model relevant movies can be found which is trained based on a hand-crafted multimodal movie content representation & user feedback.

CF can be model based, or memory based [9]. The two approaches used are regression and clustering. Results show that FFM approach+ k means method shows less errors than other models. The FFM(field-aware factorization machine) overcomes the drawback of traditional approaches.

Proposed problem statement with the specific issue you intend to address

By and large the necessity of entertainment is must. There are relatively huge number of ways and means intended to do so. With the ever-increasing choices and tight schedules, there is a lack of time and energy one can spend upon deciding on facts . As said earlier one of the most creative way of entertainment is watching movies which might also be a passion for many indeed.

What can we say about the success of a movie before it is released? Given that major films costing over \$100 million to produce can still flop, this question is more important than ever to the industry. Can we predict which films will be highly rated, whether they are a commercial

success? So, movies recommender system is a system that seeks to predict or filter preferences according to the user's choices. Movie recommendation systems provide a mechanism to assist users in classifying users with similar interests. This makes recommender systems essentially a central part of websites and e-commerce applications.

Our main focus or issue in interest is to surf among various recommendation methodologies and tools, implement them with appropriate datasets, make comparisons among them and also as a major part draw insights through various visualizations and metrics.

So, the overall effort and goal lies in making the best recommendations taking into concern user's past reviews or history, or by taking into consideration ratings given by similar users.

How is your approach (or the type of problem you are looking at) different from what has already been done?

Though our research on high level view deals with normally used algorithms like CBF and CF along with common comparison metrics like cosine similarity and Pearson's correlation coefficient, it would seem different in certain ways.

Firstly, rather than focusing on the implementation details, the project is much oriented towards analytics of data i.e., making out which method to use when. Also deciding upon which factors affect the recommendations the most.

Basically, comparison between algorithms would give more insights than what we gather from the implementation part.

Many solutions already exist to this problem. But our implementation is different from others as we are using CBF filtering(uses a series of discrete characteristics of an item to recommend additional items with similar properties) and CF(approaches to build a model from user's past behaviour) and comparing in what way these models differ. And we will be using CF in two different ways

(Pearson's coefficient and Cosine similarity), and we do a comparison between these two models.

We have used both the methods possible on the same dataset to come to a conclusion on which method would yield better recommendations on the same dataset taking necessary factors into consideration. The same model of idea can be used in various ecommerce sites for the recommendation of products.

References

[1] Similarity Based Collaborative Filtering Model for Movie Recommendation Systems. Raghavendra, C K Srikantiah K.C -2021

<https://ieeexplore.ieee.org/abstract/document/9432354>

[2] User Centric and Collaborative Movie Recommendation System Under Customized Platforms. Souptik Saha, S.Ramamoorthy, Eisha Raghav – 2021

<https://ieeexplore.ieee.org/abstract/document/9451672>

[3] Analysis of Movie Recommendation Systems with and without considering the low rated movies. Muppana Mahesh Reddy, R. Sujithra Kanmani, Dr. B. Surendiran -2020

<https://ieeexplore.ieee.org/abstract/document/9077803>

[4] An Improved Approach for Movie Recommendation System . Shreya Agrawal, Pooja Jain – 2017

<https://ieeexplore.ieee.org/abstract/document/8058367>

[5] Movie Recommendation System Using Clustering and Pattern Recognition Network. Muyeed Ahmed, Mir Tahsin Imtiaz, Raiyan Khan – 2018

<https://ieeexplore.ieee.org/abstract/document/8301695>

[6] Movie Recommendation System Using NLP Tools. Nimish Kapoor, Saurav Vishal, Krishnaveni K S – 2020
<https://ieeexplore.ieee.org/abstract/document/9137993/>

[7] Movie Recommendation System Employing the User-based CF in Cloud Computing. Tianqi Zhou, Lina Chen, Jian Shen-2017
<https://ieeexplore.ieee.org/abstract/document/8005971>

[8] Social-Aware Movie Recommendation via Multimodal Network Learning. Zhou Zhao, Qifan Yang, Hanqing Lu, Tim Weninger, Deng Cai, Xiaofei He and Yueting Zhuang -2017

<https://ieeexplore.ieee.org/abstract/document/8010448>

[9] Movie Recommendation System for Educational Purposes Based on Field-Aware Factorization Machine. Fei Lang, Lili Liang, Kai Huang, Teng Chen & Suxia Zhu -2021
<https://link.springer.com/article/10.1007/s11036-021-01775-9>

.