International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Comparative Analysis of Movie Recommendation Systems

Shubhi Srivastava
Shubhisrivastava474@gamil.com

student
National P.G College
Lucknow, Uttar Pradesh, India

Vaishnavi Srivastava
vaishnavilavi@gmail.com

Student
National P.G College
Lucknow, Uttar Pradesh, India

Mahesh Kr. Tiwari
Masheshyogi26@gmail.com

Asst. professor
National P.G College
Lucknow, Uttar Pradesh, India

Abstract

Recommendation systems are an important part of modern live-streaming services as they deliver personalized and relevant content to users. Machine learning algorithms have become integral to these systems because they can automatically learn from user behavior and provide accurate recommendations. Evaluating the machine learning algorithms used for recommendation systems in live streaming services is important to ensure that recommendations are accurate and effective. The recommendation system is derived from collaborative filtering, content-based, and hybrid-based approaches. These approaches will be used in the paper for finding the sentiments in movie recommendations and the further methods used are genetic algorithm and k means grouping.

INTRODUCTION

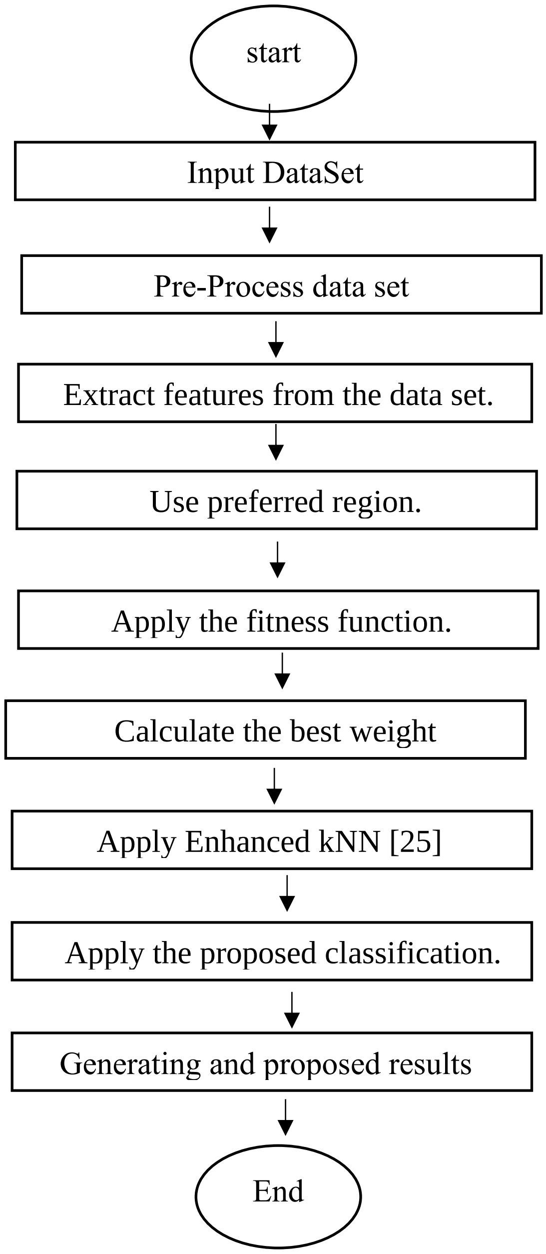
Recommendation systems have become a vital part of our daily lives, assisting us in discovering new materials and products that are relevant to our interests. In live streaming, when there is a vast amount of content available, recommendation algorithms play a vital role in directing consumers to relevant and entertaining streams. Machine learning algorithms are used to assess enormous volumes of data, such as user behavior, feed features, and social media interactions, to deliver personalized suggestions. Recommendation systems make use of a classifier ensemble built with Naive Bayes and the Genetic algorithm. They automate the system by using polarity or sentiment to determine whether or not to recommend a specific item. A review is described as having either a positive or negative polarity. If sentiment-based automated classification functioned well, it would have numerous applications. Second, categorization by customer feedback can be utilized to automatically tabulate the data. It is the process of opinion mining, which involves extracting the viewers' sentiments. Sentiment analysis is a type of text categorization that categorizes material according to the sentiment orientation of the opinions it contains. The remainder of this paper has a section on recommendations.

KEYWORDS: *Naive Bayes, Genetic algorithm, mining, accuracy*

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Related Works

Evaluating machine learning algorithms for live-streaming recommendation systems is a challenging undertaking. This necessitates assessing the accuracy and efficacy of these algorithms in forecasting user preferences and feed relevancy. These algorithms are evaluated using a variety of measures, including precision, recall, coverage, and diversity. Furthermore, scalability and real-time performance are required for a recommendation system to function in a live-streaming setting. The processing of the data set for recommendations works as follows: - Start -> Input -> Pre-process -> Extracting -> Using Preference Region -> Applying Fitness Function -> Calculate Best Weight ->Apply enhanced KNN classification -> Apply proposed classification -> Generate proposed result -> Finish.



step 1: Initialization: run pre-determined generation of KNN to generate initial data set;

step 2: Parameter input: ask the DM to input data according to the information;

step 3: Determining reference region:

step 4: Updating: update the data in accordance;

step 5: Interaction: If the result is unsatisfied with reference information; go to step 4; otherwise, output the results.

$$a = 1/N \sum \frac{f_i - fi_{-1}}{f_{t-1}}$$

where N is data about the region and values of f are expiremented.

Methodology

The methodology section describes the methods for obtaining information for the research paper. It has various algorithms that are used to find the sentiments using a variety of data sets which will help to find and recommend a movie to a user.

Movie Recommendation Systems

Movie recommendations function by filtering out data that is relevant or irrelevant and determining if it meets the traits or attributes. The System operates by modifying data to guarantee that it efficiently generates data-driven decisions. Amid the available product information, the systems must determine what fits a certain client and what does not. The solutions go beyond target and retargeting marketing to enhance product viewership and hence boost the likelihood of customers purchasing. To secure product sales or movie viewing, developers must create systems with greater performance features and efficiency that match customer preferences. There are four basic types of filtering methods: collaborative filtering, content-based filtering, context-based filtering, and hybrid filtering.

- ***Filtering by Collaborative***

  It is a filtering method that works by comparing the similarities between users and products. It collects attributes for both the person and the products he or she uses. In general, filtering is done using specific attributes from rating matrices. The recommender system makes recommendations based on the information the user supplies or what other individuals with similar information are watching or seeing. For example, in collaborative filtering in movie recommender systems pick the users who have the same kind of information, such as age, gender, or ethnicity, and these facts are used for other users as well. People or users with the same demographic attributes may have the same likes. The accuracy of the suggestion is often limited since the data used to make suggestions for the same demographic variables may not have similar preferences.

- ***Filtering By Content***

  Content-based approaches enable people to provide suggestions based on feature vectors, as an alternative to collaborative filtering. To recommend movies, the content-based technique only looks at your past viewing habits or history; in contrast, the collaborative filtering technique looks at user behavior and serendipity, regardless of domain knowledge, to recommend content. This is the main distinction between the two approaches. The main characters and the film's genre are two items that are included in movie suggestions.

- ***Filtering Based on Context***

  An advancement over the collaborative filtering method is this filtering technology. This filter indicates that if two individuals share a concern about a certain subject or a preference for some topic, then there's a chance that they'll choose the same option in most situations. This is the operation of context-based filtering. It features Context-aware recommender systems (CARS), which enable the system to provide better results because the concept of context is properly defined. The collaborative filtering approach has been enhanced by this filtering technology.

- Filtering with a Hybrid Approach

  This filtering method achieves higher performance and faster processing times by combining all the filtering methods that are utilized to overcome the most superior method. To address issues such as a lack of knowledge about domain dependencies, it integrates collaborative, content-based, and context-based filtering.

MACHINE LEARNING ALGORITHM FOR MOVIE RECOMMENDATION SYSTEMS

These algorithms are employed in data mining and information filtering to provide the intended results.

K- Means Grouping

This method of grouping individuals according to their interests is among the most straightforward collaborative filtering techniques. It is customary for someone wishing to buy anything to inquire about the opinion of someone who has already purchased the product or item. There is a greater likelihood that the opinion or preference of the owner who has previously purchased the product will affect the current owner who is about to purchase it, as he or she may attempt to compare it with the recommendation of the owner who has already used it. In the same way, the algorithm contrasts the intriguing aspects.

When making recommendations, K-means clustering makes use of the user's or viewers' shared interests, which include age, gender, movie, time, history, location, and many more. It's comparable to when we ask friends for movie reviews or their thoughts on what they think we should show people who have already seen the film. K-Means clustering attempts to organize the features into clusters that symbolize the group's attributes. Assume that if the classification is done according to age, the approaches will be similar to those used for children, teens, youth, and adults. Movie recommendations will be made if the user is within the age range of these areas based on what other users or clients of that age have viewed or typically watch.

Analysis Of K – Means

Dataset:

Analyzing the gathered datasets to achieve the desired results involves taking 100 datasets and following the results to use the k-means algorithm.

Steps involved:

f = ['R', 'RY']
standard Scaler() = scaler
scaled is equal to scaler. transform movie data[f] with fit
k = number of clusters
K means = K Cluster Means (k), Random (random state) Means (k). suit (scaled)
cinema data['Cluster'] = k means. movie data labels print

Here, we are taking 'f' as the movie's features, with 'R' as the rating and 'RY' as the release year, this is how data is included from the dataset, and it is utilized for feature selection. Scaler and scale are, on the other hand, used to scale features. In the other line, we apply clustering to the data, setting the random state to 42 to regulate the random number generation for centroid initialization. The value of k is responsible for determining the number of clusters that will be present, then cluster labels will be added to the original data frames, and in the end will display the data frames with clustering labels.

METAHEURISTIC ALGORITHM

Genetic Algorithm

The genetic algorithm is a machine learning technique that uses a metaphor to represent some of nature's evolutionary dynamics. This is performed by generating a population of individuals represented by chromosomes, which are effectively a collection of character strings. Individuals propose solutions to the optimization problem being solved. In genetic algorithms, individuals are frequently represented as n-bit binary vectors. The resulting search space is n-dimensional boolean space. It is anticipated that the quality of each potential solution will be evaluated using a fitness function. Genetic algorithms choose people from the current population based on fitness to create offspring for future generations. Mutation and crossover are two of the most often employed processes in genetic algorithms that represent individuals as binary strings. Mutation modifies a single string at random, whereas crossover modifies two parent strings to produce two children. Other genetic representations necessitate the use of relevant genetic operators.

The process of fitness-dependent selection and the use of genetic operators to generate successive generations of individuals is repeated until a satisfactory solution is discovered. A genetic algorithm's success is influenced by several factors, including the genetic representations and operators used, the fitness function, the fitness-dependent selection mechanism, and user-defined parameters like population size and operator likelihood. The fundamental operation of the genetic algorithm is described as follows.

Procedure: Begin.

t <- 0
Initialize P(t)
while (not termination condition) t
<- t + 1
select P(t) from p(t - 1)
crossover P(t)
mutate P(t)
evaluate P(t)
end.

Firefly Algorithm

- The program is also bio-inspired by fireflies and uses a fuzzy C-means clustering technique. In the natural world, fireflies are drawn to the brightest firefly by the light signal. Each firefly attracts the other, but the brightest is the most attractive, and other fireflies swarm around it. Similarly, the algorithm bases its recommendations on the attributes of the most appealing persons (highest user ratings). If a movie has received the top rating from a large number of users, the movie recommender system will generate the following recommendations based on movies with comparable features. The algorithm for the recommender system is shown below:

- Fireflies are unisexual and attract each other based on their brightness (feature reduction using the firefly algorithm).

- This brightness is linked to a primary function of the FCM. FCM assigns memberships and uses them to display data components from one cluster to another.

- The FCM divides a finite set of items (X = X1.... Xn) into fuzzy clusters, resulting in a list of cluster centers (C = C1... C2). The partition matrix W = Wi [0,1]; i = 1... n; j = 1...c denotes the degree to which each element Xi is assigned to a cluster Cj.

- Then the fuzzy C-means clustering happens

The recommender system efficiency and performance are generally higher than the traditional K-means clustering.

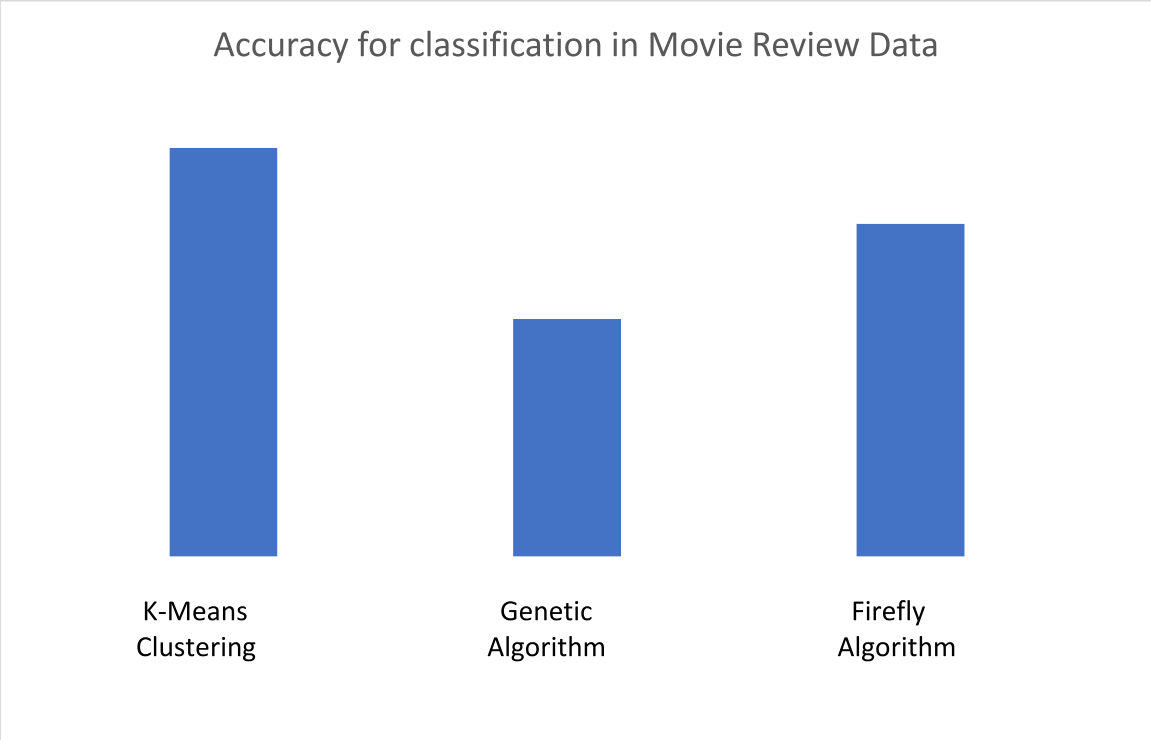
EXPERIMENT

Dataset Description :

The baseline data set consists of 2000 movie reviews, 1000 labeled positive and 1000 labeled negative, in a uniform distribution, and is downloaded from Bo Pang's website, where we also have an additional data set of 100 more reviews with mixed reviews collected by the questioner.

Results and Discussion :

Dataset	Classifier	Accuracy
Movie Review Data	K – means Clustering	93.25%
	Genetic Algorithm	91.02%
	Firefly Algorithm	92%



The above data set shows how accurate all three algorithms are.

CONCLUSION

The conclusion says that the best filtering among all is hybrid filtering as it provides accuracy while among algorithms k means is the most accurate to recommend in comparison with the genetic algorithm and firefly algorithm.

REFERENCES

1. International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-4 Issue-13 December 2013.
Sentiment Analysis of Movie Reviews Using Hybrid Method of Naive Bayes and Genetic Algorithm
M . Govindarajan

2. Movie Recommender Systems: Concepts, Methods, Challenges, and Future Directions, Sambandam Jayalakshmi [1], Narayanan Ganesh [1], Robert ˇ Cep [2],* and Janakiraman Senthil Murugan [3]

3. Journal of Interdisciplinary
Cycle Research ISSN NO: 0022-1945
Movie Recommendation System
2 Author(s): Yogesh Kumar, Ms. Naveen Kumari
Guide Name: Ms. Naveen Kumari (Astt. Prof.)
yogi.peshlo@gmail.com , naveencse2k4@gmail.com
Punjabi University Regional Centre for IT & Management, Mohali

4. A Abbasi, HC Chen, and A Salem, (2008), “Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums”.
ACM Transactions On Information Systems, Volume 26, issue 3, No 12.

5. L. Brieman. (1996), “Bias, Variance, and Arcing Classifiers”, Technical Report 460, Department of Statistics, University of California at Berkeley, CA 94720.

6. S Chandrakala and C Sindhu, (2012), “Opinion Mining and sentiment classification a survey”, ICTACT journal on soft computing.

7. Efron, M. (2004), “Cultural orientations: Classifying subjective documents by cocitation analysis”, In Proceedings of the AAAI Fall Symposium Series on Style and Meaning in Language, Art, Music, and Design, pp. 41-48.

8. Freund, Y. and Schapire, R. (1995), “A decision-theoretic generalization of online learning and an application to boosting”, In proceedings of the second European Conference on Computational Learning Theory, pp. 23-37.

9. Freund, Y. and Schapire, R. (1996), “Experiments with a new boosting algorithm", In Proceedings of the Thirteenth International Conference on Machine Learning, Bari, Italy, pp.148-156.

10. Hesham Arafat, Rasheed M. Elawady, Sherif Barakat, Nora M. Elrashidy., (2014), "Different Feature Selection for Sentiment Classification", International Journal of Information Science and Intelligent System, vol 3issue 1, pp. 137-150.