AI in Entertainment – Movie Recommendation System

Karan Shah[1], Aliraza Shinde[2], Shreyas Vaghasia[3], Prof. Bhavna Arora[4]

[1,2,3]Student, Dept. of Computer Engineering, Atharva College of Engineering, Mumbai, India
[4]Prof, Dept. of Computer Engineering, Atharva College of Engineering, Mumbai, India

---

Abstract - *Artificial intelligence is now a critical component of a wide range of enterprises. Especially during a pandemic, when cinemas are closed and people working from home have to entertain themselves by viewing movies and listening to music, these people are turning to OTT platforms. On-demand streaming services provide nearly limitless content, and the introduction of JIO fibre, which provides high-speed data at low prices, as well as products like Amazon Fire Stick and Google Chrome-cast, which have transformed traditional television into a smart television, has fundamentally changed the landscape. These OTT services are a library of related content with unlimited debate subjects. One of the most important areas that have to be improved is personalization because it allows each member to have a unique perspective of material that adapts to their interests and can help them broaden their interests over time. Each experience is tailored in a variety of ways, including the films recommended and ranked, the way movies are structured into rows and pages, and even the artwork presented. To achieve this level of deep customisation, we must mix various computational approaches to satisfy the needs of each member. Personalization begins on the site and continues throughout the product and beyond, including selecting what messages to send consumers to keep them informed and engaged. Users should spend less time looking for material to watch and more time watching stuff that will provide actual value to their lives.*

Key Words: **Recommendation System, Artificial Intelligence, Personalization, OTT, Cosine Similarity, Content-based filtering, Collaborative filtering, KNN**

1. INTRODUCTION

The Recommendation System assists clients in finding the matters they preference from a big variety of options. These fashions have turned out to be extraordinarily famous in that they're being utilized in movies, books, television, restaurants, meals etc. A massive variety of groups are making the most of the advice gadget in enhancing purchaser pride and enhancing purchaser repeat rate. Many businesses, such as Netflix, Amazon, ZEE-5, and others, use recommendation algorithms to better serve their customers and increase profits. Still, it's an interesting research issue because determining what the user wants from the available resources is difficult, especially when our preferences change over time.

Nowadays what we purchase online is the recommendation. For example, if we want to buy a book, listen to music or watch a movie, there is a recommendation system in the background that makes recommendations based on the previous behaviour of the user. Many platforms like Netflix, Hotstar, ZEE5, which suggest movies, Amazon which suggest products, Spotify, Gaana, Hungama Music, which suggest music, Grammarly that suggests words and corrects grammar, LinkedIn that is used for recommending jobs or any social networking sites which suggest users, all these work on the recommendation system. Users may simply find out what they want based on their preferences by using these recommendation engines. The suggested system can store a huge quantity of data and produce accurate results. Our recommendation engine searches for the finest movies based on the genre that is comparable to the one we just viewed and returns the results.

2. Application and Scope

1. The proposed system will recommend movies based on ratings and genres given by the user itself. Content-based filtering is the term for this type of filtering.

2. The proposed system is capable of storing a large amount of data and giving efficient results. Our proposed recommendation system works with Bollywood and Hollywood movies for now.

3. Previous research lacks the accuracy of true recommendations due to their dependency on other users. It is vital to note that suggestions can be based on a specific person, and they guarantee to deliver accurate recommendations because they are based on the likes and dislikes of that user.

3. Basic Concepts

Proposals are based on ratings and genres, so it is difficult to provide reliable results. Collaborative filtering is the terminology. This is a common approach in recommender systems. To make it easier for users to overcome this shortcoming, content-based filtering mechanisms need to be implemented. The user's rating and genre are used. Recommendations are depending on the movie the user has watched before. This helps you make accurate suggestions, as ratings and genres are provided only by the user and are based solely on the user's preferences, not the preferences of others. Due to the inaccuracy of previous studies, no actual suggestions can be made. With content-based filtering, similarity can be calculated in three ways:

1. Cosine similarity

2. Euclidean Distance

3. Pearson's Correlation

Cosine similarity is better than both Euclidean distance and Pearson's correlation because two quantities represent different physical quantities. The Pearson correlation coefficient computes the correlation between two jointly distributed random samples, whereas the cosine similarity computes the similarity between two samples. The resultant angle determines the level of similarity and is inversely proportional (smaller angle higher similarity).

3.1 Collaborative Filtering

[2] In collaborative filtering, items based on the similarity between previously chosen items are recommended.

[3]. It works by collecting data from users in the form of ratings in a specific field, then determining their similarity to recommend an outcome [2].

3.1.1 Advantages:

1. Dependent on the ratings, thus making it content-independent.

2. It can suggest fortuitously recommendations based on the similarity of users.

3. It also considers the experience to create real-life assessment.

3.1.2 Disadvantages:

1. If the initial ratings are contradictory to the later ones then ambiguity arises.

2. Variations in review cases are difficult to the group in agree or disagree nature.

3. Difficult in tackling sparsity situations.

3.2 Content Based Filtering

[2] A content-based review program tries to guess the features or character of a user who has been given the features of an item, and responds positively to it. The matching metrics were generated from the item's material vectors and the user's preference vectors from his previous records at the time of suggestion.

3.2.1 Advantages:

1. [12] It does not require other users' data for recommendation for the given user.

2. The model can recognise a user's unique preferences and make recommendations for niche things that only a few other people are interested in.

3.2.2 Disadvantages:

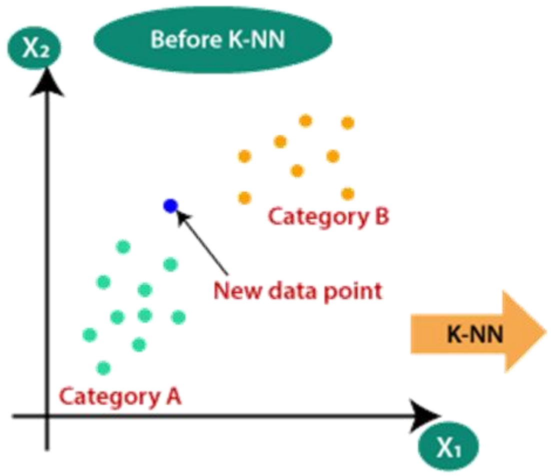
1. [12] Only depending on the user's current interests does the model make recommendations. To put it another way, the model's ability to capitalize on other customers' existing passions is limited.

2. Because it doesn't consider what others think of the item, it's possible that low-quality item recommendations will result.

4. Recommendation techniques:

4.1. KNN clustering:

KNN is a simple algorithm that uses the target function's local minimum to learn an unknown function with the appropriate precision and accuracy. The programme also determines the location of an unknown input, as well as its range and distance from it. It is based on the concept of "knowledge acquisition," in which the algorithm identifies the optimum way for predicting an unknown number.



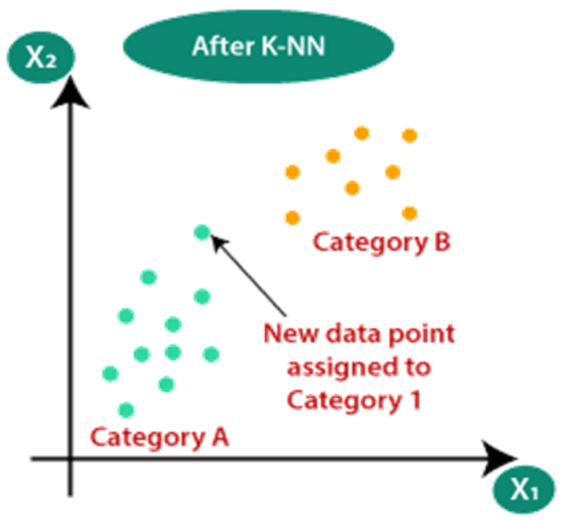


FIGURE -1: KNN CLUSTERING

4.2. Cosine Similarity:

Cosine distance is measured by taking cosine of the angle between them. It compares two documents on a normalized scale, by finding the dot product between them. If the angle between the two vectors is small, they are about the same, and if the angle between the two vectors is large, the vectors are very different.

Documents can be categorized based on thousands of attributes. Each attribute records the frequency of a single word (such as a keyword) or phrase in the text. As a result, each document is an object represented by a term frequency vector.

A metric is a unit of measurement that represents the distance between two items, or their separation. You can use another function called a similarity measure or similarity factor to estimate proximity in terms of similarity. [8].

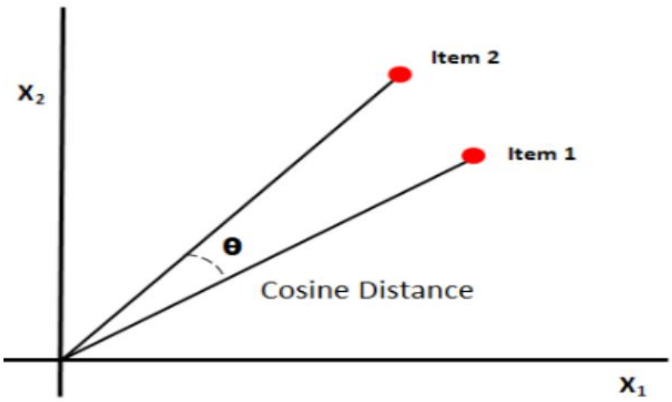


FIGURE -2: COSINE SIMILARITY

[9]We'll start with the set Rn and two vectors x,y∈Rn. The dot product $x \cdot y$ is an operation on the vectors that returns a single number.

It is the equation x·y=Σi=1nxiyi.

The norm of a vector x is ‖x‖=Σi=1nxi2:

The **cosine similarity** is defined by the equation CS=x·y‖x‖‖y‖

4.3. Advantage of Cosine Similarity over KNN:

1. KNN is also known as "Lazy Learner" and it is effective on a small dataset. This point itself is convincing to look for different alternatives to this approach.

2. Cosine Similarity is sensitive to accuracy but is computationally efficient.

3. KNN is more computationally memory consuming.

5. Proposed System:

In the proposed system, we will be developing a content based recommendation system using cosine similarity algorithms. The model training data will be taken from TMDB 5000 movies with different files for genres, casts, writers, etc.

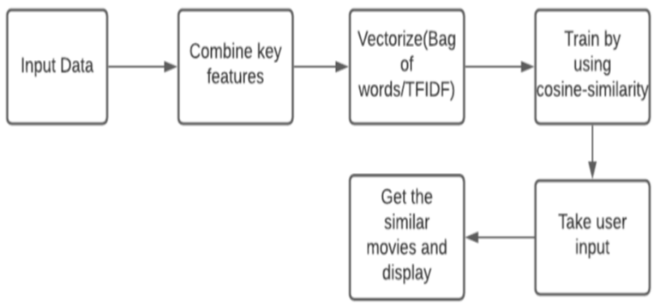


FIGURE -3 BLOCK DIAGRAM

1. First, we will check if the name entered by the user is a valid movie or not. For the model building process, the data will be pre-processed by cleaning all data entries and checking for missing values and removing duplicate values as well. During data cleaning we will remove stopwords (stopwords are those words which occur frequently in a sentence, for example articles) from the movie descriptions to provide relevant importance to similar words. This will help the system to identify similar descriptions easily.

2. The Dataset is in the form of comma separated files (csv) and there are different files for different attributes that the model will function on. The attributes are language of the movie, the cast, directors, writers and the genre of the movie.

After pre-processing all files will be vectorized by using TFIDF (Term Frequency-Inverse Document Frequency). Vectorization is a methodology to map words or phrases to corresponding vectors of real numbers.

3. After all pre-processing, all these attributes are combined in one dataset. Cosine similarity will be applied on this dataset to generate the similarity coefficient. Any movie title that the user will type based on the similarity coefficient we will get our recommended movie results.

6. CONCLUSION:

This paper discusses the various sorts of recommendation systems utilized in the industry. Its various use cases, benefits, and drawbacks are also discussed. A hybrid combination of several recommendation systems is necessary to develop a competent recommender system. It can be inferred that by employing a combination of similarity measures rather than a single similarity measure, a better user similarity can be achieved, and the system's efficiency is improved. One of the facts is that author-created similarity measures such as Cosine similarity have only been employed in movie recommendations up until now. In terms of efficiency parameters, the author also showed that one similarity metric outperforms the other. Any recommender system's accuracy can be questioned. Any recommender system's accuracy can be improved by including more movie features.

References

1. Suresh, Salini, et al. "Latent Approach in Entertainment Industry Using Machine Learning." International Research Journal on Advanced Science Hub, vol. 2, no. Special Issue ICARD, 2020, pp. 301–304., doi:10.47392/irjash.2020.106.

2.Desai, Harshali. "Movie Recommendation System through Movie Poster Using Deep Learning Technique." International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 4, 2021, pp. 1574–1581., https://doi.org/10.22214/ijraset.2021.33947.

3. (PDF) AI in the Media and Creative Industries - Researchgate. www.researchgate.net/publication/333041972_AI_in_the_media_and_creative_industries.

4.Sunilkumar, Chaurasia Neha. "A Review of Movie Recommendation System: Limitations, Survey and Challenges." ELCVIA Electronic Letters on Computer Vision and Image Analysis, vol. 19, no. 3, 2020, p. 18., doi:10.5565/rev/elcvia.1232.

5. Singh, Abhishek, et al. "A Research Paper on Machine Learning Based Movie Recommendation System." International Research Journal of Engineering and Technology, vol. 08, no. 3, Mar. 2021, doi:10.26782/jmcms.2020.07.00056.

6. Movie Recommendation System Using Cosine Similarity and KNN. www.researchgate.net/publication/344627182_Movie_Recommendation_System_using_Cosine_Similarity_and_KNN.

[7]. "Cosine Similarity." Wikipedia, Wikimedia Foundation, 11 Feb. 2022, en.wikipedia.org/wiki/Cosine_similarity.

[8]. Lüthe, Marvin. "Calculate Similarity-the Most Relevant Metrics in a Nutshell." Medium, Towards Data Science, 13 May 2021, https://towardsdatascience.com/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e.

https://towardsdatascience.com/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e

[9]. Lüthe, Marvin. "Calculate Similarity-the Most Relevant Metrics in a Nutshell." Medium, Towards Data Science, 13 May 2021, https://towardsdatascience.com/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e.

10. Liu, Jingdong, et al. "Personalized Movie Recommendation Method Based on Deep Learning." Mathematical Problems in Engineering, Hindawi, 19 Feb. 2021, www.hindawi.com/journals/mpe/2021/6694237/.

11. Movie Recommender System Using Critic Consensus. www.researchgate.net/publication/357268035_Movie_Recommender_System_using_critic_consensus/fulltext/61c3e8bc8bb20101842ed7dd/Movie-Recommender-System-using-critic-consensus.pdf.

[12] "Content-Based Filtering Advantages & Disadvantages | Recommendation Systems | Google Developers." Google, Google, developers.google.com/machine-learning/recommendation/content-based/summary.

BIOGRAPHIES

Karan Shah
(BE Student) Atharva College of Engineering, Department of Computer Engineering, Mumbai, India

Aliraza Shinde
(BE Student) Atharva College of Engineering, Department of Computer Engineering, Mumbai, India

Shreyas Vaghasia
(BE Student) Atharva College of Engineering, Department of Computer Engineering, Mumbai, India

Bhavna Arora
(Professor) Atharva College of Engineering, Department of Computer Engineering, Mumbai, India