A MOVIE DATA RECOMMENDATION SYSTEM USING A HYBRID MODEL

Amaefule I.A[1], Ibebuogu C.C[2], Chilaka U.L[3]

[1&2]*Department of Computer Science, Imo State University, Owerri. Imo State, Nigeria*
[3]*Department of Computer Science, Kingsley Ozumba Mbadiwe University, Ideato. ImoState, Nigeria*

ABSTRACT

In recent years, the rate of expansion in the amount of information available online has led to both a surge in Internet users and massive amounts of data. The use of the recommendation system has been made possible by the enormous amount of information available to users. This study, proposed a hybrid recommender engine that could combines recommendations from content-based and collaborative filtering. This aims to investigate how existing collaborative filtering frameworks can improve prediction accuracy. We examine whether a recommendation system that combines content-based and collaborative filtering, employing a Mahout Structure and developed on Hadoop, will enhance accuracy of the recommendation and as well resolve adaptability problems presently encountered in handling huge data sizes for users recommendation of items. The enhancement of features was used. The hybrid feature augmentation technique was used, in which the output from collaborative filtering was used as an input to content-based recommendation. To extract user and item content features, the well-known Movie-Lens data was linked with the Internet Movie Database (IMDB). The text files created as a result of integrating the two databases were then used as inputs for Mahout's collaborative filtering architecture. The ideal Mahout Components parameter optimization for our model was found through a number of tests. By comparing our model's Root Mean Square Error to that of the most recent model, we further investigated these models. When compared to the pure collaborative model, the study demonstrated a considerable improvement. Our investigation showed that the derived user and item content attributes may result in more accurate prediction

Keywords: *collaborative Filtering, content-based, movie data, optimization, hybrid recommender, data, Recommendation system.*

INTRODUCTION

Huge volumes of data and a rise in internet users are the results of the internet's rapid expansion of information. Users have been inundated with a vast amount of information as a result of this enormous data explosion, which presents a significant information overload concern. Consequently, this has made it highly challenging for individuals to digest such data manually and also quite challenging for individuals to identify the appropriate information. Users' inability to make wise selections given the sheer volume of information frequently leads to great confusion.

The management of this information boom has proven challenging for big internet corporations like Amazon, Google, and Facebook. This problem has been cleverly transformed thanks to the usage of recommendation systems, which demonstrates how system recommendations have stepped up to save people from such perplexity.

The massive increase in internet data and its users has led to rise in big data. The Big Data sector has shown the most concern in the System Recommendation. Bid Data has improved our capacity to offer broad recommendations. The recommendation system has grown more important for consumers because of its capacity to anticipate the accurate information from a vast volume of information. The system is a specific kind of information filtering that makes use of prior user behavior or the behavior of users who are similar to the user in question to create a list of information items that is clearly suited to the preferences of the end user.

Recommendation Systems (RSs) are currently widely utilized in e-commerce to give customized information by anticipating users' preferences for certain items; aiming to recommend content (movies, music, books, news, websites, etc.) that is most likely to attract users' interests. RSs are widely used by portals like Amazon, Netflix, and others to propose material to their viewers. By delivering the most appealing and pertinent content, RSs seek to reduce the problems associated with information overload. Every e-commerce portal now considers RSs to be essential.

RECOMMENDER SYSTEM

In order to help individuals cope with the massive growth of information on the internet, systems recommendation have developed in response to a clear need. It was created, to put it simply, to help with the information overload issue. It has also become clear that it is able to link users with shared interests, rather than just providing them with pertinent information [1].

An algorithm called the Recommender System predicts how users will react to certain options. On the basis of their prior behaviors, for instance a history of transactions and/or items queries, clicks, and rankings, it provides internet users with ideas for what their interests may be. The crucial goal of a Recommender System is to make a recommendation intended to assist users in several decision-making processes.

Using data from the user's prior preferences and actions, Amazon employs this strategy to present to a specific user a list of suggested things which may be of consideration. Recommender systems are programs that operate in the background to record user behavior and provide recommendations for particular products depending on that behavior. Facebook creates the "people you may know list" by using the same recommender algorithm to decide which friends to suggest.

ENTITIES IN RECOMMENDATIONS SYSTEMS

Typical types of entities [1]

1. **Item:** The system suggestion to users is described using item.
2. **Users**: Recommender systems either ask users directly about their choices or deduce them from the interpretation of user behavior, for instance, when a user clicks on a product. Users have choices for particular items, and these choices need to be removed from the equation.
3. **Utility Matrix**: Data representation is done using utility matrix. Utility matrix is used to, users score objects in a user-item using a scale pair. This rating reveals the amount of the user's liking for the product. Generally on 1 to 5 scale.
   - A system's recommender objective is to foresee the gaps in the Utility Matrix.
   - It is crucial to populate the Utility Matrix since without it; it is very difficult to recommend products.
4. **Transactions:** A user's engagement with the RS that has been recorded is stated as a transaction. Transactions are log-like data that are used by the system's recommendation generating algorithm to record significant information created during human-computer interaction. An instance of user-designated item and a elucidation of the framework (such as the user objective or inquiry) for that specific recommendation, for example, could both be found in the transaction log. If viewed, that transaction may also contain a direct comment from the user, such as the item's rating. The most common type of transaction data that an RS gathers is ratings, literally.

[2], States that he classified five recommendation techniques based on the nature of the contextual and data input along with the algorithm employed to generate the recommendations, as illustrated below.

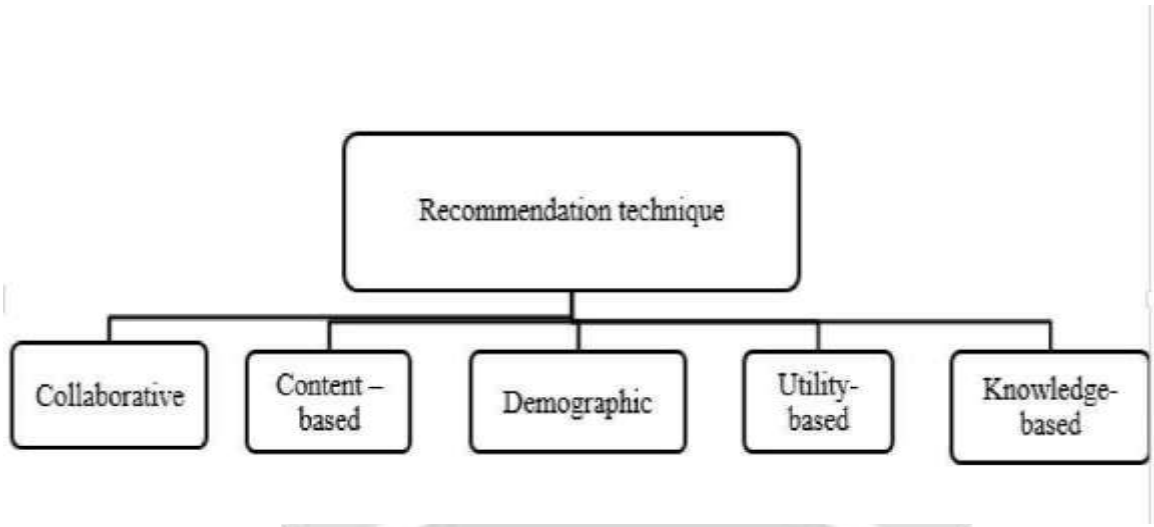

Figure 1: Instance of recommender systems taxonomy [2]

LITERATURE REVIEW

[3]. Uses a combined methodology to determine user choices for television viewing. It employs CBR based on text-based plot summaries of TV programmes. The collaborative technique is then employed to discover more about the choices of different users. The concluding suggested program comprises recommendations made jointly and based on content. PTV priorities content-based recommendation over collaborative replies.

[4] Utilized the feature combination hybrid. It details trials in which the inductive rule learner Ripper was used to make movie recommendations using both user ratings and content information. This method significantly improved prediction exactness above entire collaborative method. Though, this benefit was obtained only by manually selecting content attribute. [4], found that using every piece of content feature increased recall; however not accuracy.

[2] A collaborative, knowledge-based recommender with cascading steps, The Restaurant Recommender EntreeC [2]. It needed restaurant knowledge to provide recommendations based on the user's stated choices. After being placed in equal-preference jars, the recommendations are further ranked in each jar using collaborative filtering to break any ties.

[5]. Collaborative filtering and knowledge-based algorithms were applied to Usenet news by the GroupLens research team. Feature augmentation was used [5]. They created a number of knowledge-based "filterbots" that took into account various factors, like the size of the contained messages and the quantity of typos. As if they were real users, these bots add their ratings to the database of ratings used by the Collaborative part of the system. Email filtering was made better with the deployment of relatively basic agents.

[6] Suggested a system recommender that will combine unrated movie evaluations with movie ratings on the Internet. Additionally, sentiment analysis was used to analyze user preferences in cases when explicit ratings were not directly linked to movie reviews. The IMDb dataset, which had 53112 reviews and 50% unrated movies, was used to test the Singular Value Decomposition (SVD) matrix factorization approach that he presented for collaborative filtering. [6] were able to overcome data sparsity, however unlabeled data limited their ability to provide accurate recommendations.

[7]. An experiment using the k-Nearest Neighbor (KNN) and Bayesian methods suggests that such a integration can get high accuracy values to suggest movies to users. The model is based on the integration of the correlation technique with the matrix completion technique. Their goal is to create a distinctive system recommender that can

recognize a user's tastes and automatically generate a list of products that are appropriate for that user. The item-based collaborative filtering method used to compare two items using user preference data uses the social recommender known as the Pearson correlation coefficient. Utilizing decision templates, classification and regression base recommenders are combined with the bayesian closest neighbor to predict user preferences. The system was able to provide high accuracy numbers, but it still had a significant drawback because it couldn't handle enormous datasets

[8]. Took a somewhat different strategy to creating a recommender system by combining Fuzzy C Means (FCM) clustering and the Bat algorithm. Users are grouped jointly using Fuzzy C Means, and the Bat method is utilized to determine the beginning point of the clusters for high-performance recommendation standard. MAE, recall, and precision were used to compare the experiment. Limitation: Sometimes, bat optimization results may not be accurate.

[9]. Had conducted research and suggested a hybrid system recommender that integrates multi-layer ANN with k-Means and x-Means clustering. He had organized the film genres into clusters and tabulated the results using clusters according to the training cycle, learning rate, momentum, precision, recall, and recommendation accuracy. Although it had modest performance ratings, the system was capable of handling the cold-start issue.

Various hybrid system configurations have been utilized to enhance the foundational recommender systems, which, depending on the algorithm used, have a number of flaws. To address the issues with the traditional recommender systems, each of them includes specific algorithms and methods.

METHODOLOGY

A website with a database of many movies will make up the system. Any new user must create an account through the website's user interface. Users will be questioned about specific films and movie genres. Based on the input provided, the user will categorize and given a list of recommendations. Real-time investigation will allow the system to change dependent on user behavior. Users with account can access features like reading plot summaries and making watch lists.

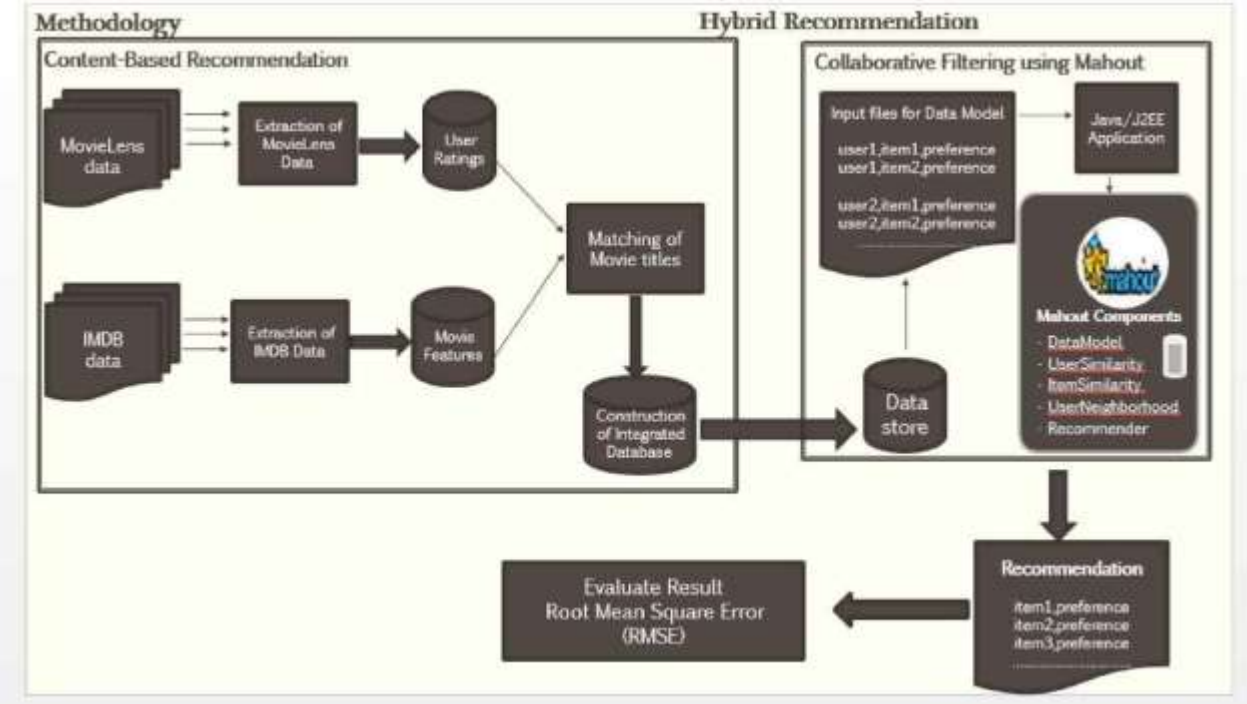

Figure 2 . Methodology for improving Recommendation Prediction Accuracy

This study presented a hybrid recommender system that offers movies recommendations based on content, collaborative filtering, and framework for more effective recommendations. The phases in the method for creating recommendations for movie are as follows with descriptions:

1. The essential features for movie recommendation must first be obtained.
2. Once the necessary information has been gathered, including rating data and contextual data, an advanced algorithm is used to generate pertinent recommendations.

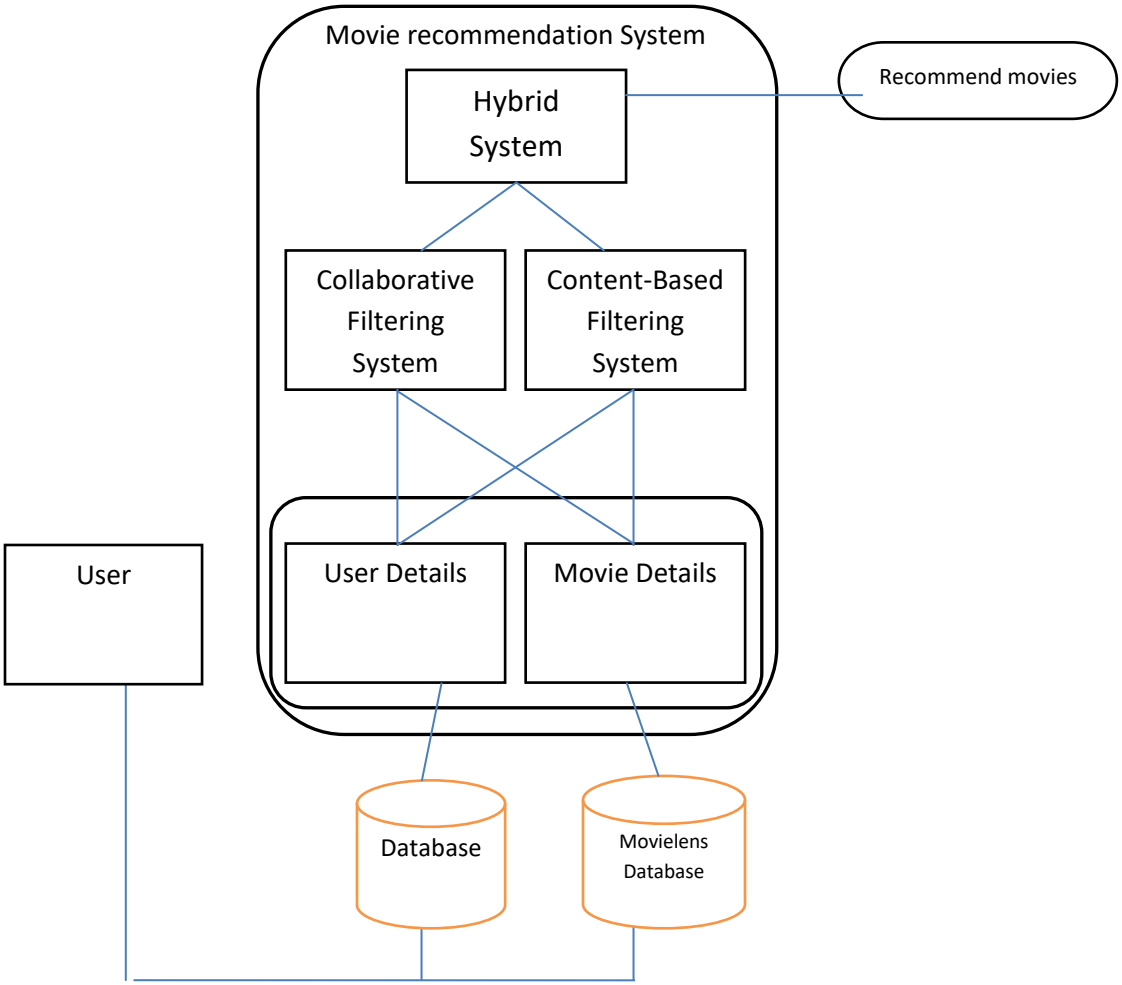

Figure 3. System Architecture a hybrid recommender system

Step 1: Gather user data

The system requests the new user to open an account in order to collect both his personal data and the ratings of a preference of films.

Step 2: Create a database of movies

Collect details about the movie, like the name, director, cast, premiere date, reviews, etc., to offer recommendations.

Step 3: Movie Recommender

At this level, movie recommendations are provided for the user.

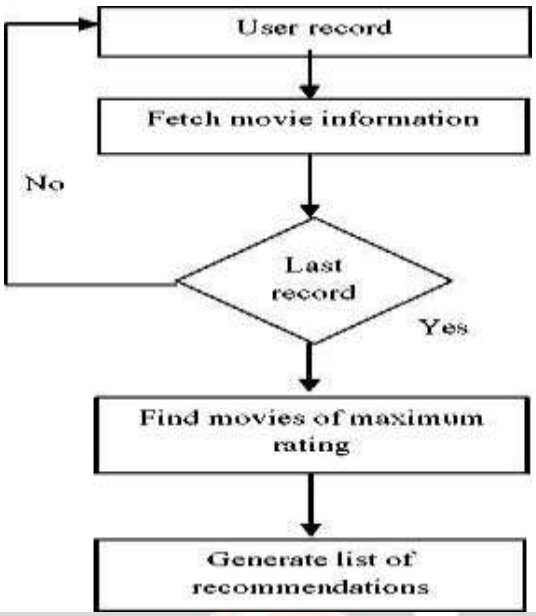

Figure 4. The flowchart for updating movie recommendations.

It can be very challenging to sift through the vast amount of web material to find the pertinent information. Recommender systems, which offer pertinent recommendations, are a solution to this issue. The system under consideration uses a hybrid filtering technique to make movie recommendations, and we can add more online-available information in the future. Additionally, social tagging can be used to improve the system. Users frequently rely on recommendations from others, whether through word-of-mouth, reviews of movies and books in newspapers, etc. Utilizing a straightforward GUI, these systems assist users by suggesting appropriate movies based on user feedback. By providing them with better recommendations, this technology will assist users in managing information overload.

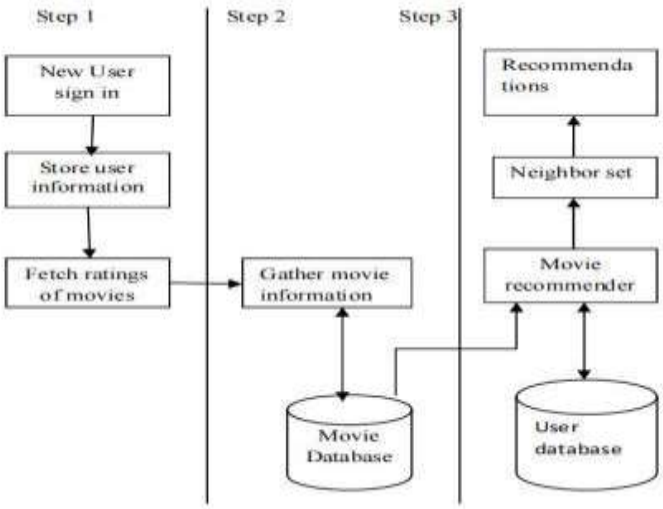

Figure 5: The flowchart of a hybrid recommender system

Conclusion

The goal of the paper was to improve recommendation results by introducing a novel technique to increase suggestion accuracy. A hybrid recommender engine was created by combining the concepts of collaborative and content-based filtering. The model is distinctive; a straightforward but original method was used to merge the rating and content details from the Movie Lens data and the IMDB data into a single system. The hybrid recommender system derived user poll characteristics, like user content characteristics and certain movie item qualities as item attributes. Less parameters and more accurate prediction outcomes were the main benefits of this unified model. Using Apache Mahout, the hybrid recommender was put into practice. The recommender parts were adjusted to get the best parameter for recommendations. It also showed, through a variety of trials, how the extracted content attributes improved the hybrid recommendation engine's ability to anticipate outcomes accurately. Furthermore, the hybrid recommender could attest to the superior performance of the tested item features over user features. Instead of relying just on collaborative filtering, a hybrid recommendation engine has been developed that combines both that framework and content-based recommendations. This advancement can be ascribed to numerous items and user attributes that were taken from the Movie Lens data and the IMDB data.

REFERENCES

[1] Beyond Recommender Systems: Helping People Help Each Other Loren Terveen and Will Hill AT&T Labs – Research

[2] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, 12(4):331-370.

[3] Smyth, B. and Cotter, P. 2000, 'A Personalized TV Listings Service for the Digital TV Age'. Knowledge-Based Systems 13: 53-59.

[4] Basu, C., Hirsh, H. & Cohen, W. (1998). Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In *Proceedings of AAAI'98*, 714-720.

[5] Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J. Miller, B. and Riedl, J.: 1998, 'Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System'. In: Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, WA, pp. 345-354.

[6] Filipa et al. (2013) A Recommender System For The Tv On The Web:

[7] Bonjori and Moradi (2012). A Novel Collaborative Filtering Model Based On Combination Of Correlation Method with Matrix Completion Technique:

[8] Kalimuthu and Vellaichamy (2017) Hybrid Collaborative Movie Recommender System Using Clustering and Bat Optimization:

[9] Yucebas (2019) Movieann: A Hybrid Approach to Movie Recommender Systems Using Multi-Layer Artificial Neural Networks: