REVIEW ARTICLE **SPECIAL ISSUE**

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

Check for updates

RSP Science Hub

http://dx.doi.org/10.47392/irjash.2023.S051

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

A Review on the Movie Recommendation System Using Big Data

Chandan Kumar Sangewar [1], Chinmay Pagey [1], Sakshi Chauhan [1], Suganeshwari G [2]

[1] *School of Computer Science and Engineering, Vellore Institute of Technology Chennai, Tamil Nadu, India.*

[2] *Assistant Professor(Sr.),School of Computer Science and Engineering, Vellore Institute of Technology Chennai, Tamil Nadu, India.*

Emails: chandankumar.sangewar2020@vitstudent.ac.in, chinmay.pagey2020@vitstudent.ac.in, sakshi.chauhan2020@vitstudent.ac.in, suganeshwari.g@vit.ac.in

Article History

Received: 28 February 2023
Accepted: 21 March 2023

Keywords:

Movie Recommendation; Big Data; Movie Lens; Machine Learning; User Feedback; Data Sparsity; Accuracy

Abstract

The project will be built on a recommendation system, particularly for material pertaining to digital media such as movies or web series. A method known as collaborative filtering is going to be the approach that forms the basis of our research. In order to carry out the implementation of this model, we will make use of the ml-25m dataset, as well as spark (MLib), the ideas of matrix factorization, and the ALS algorithm. The distributed computing architecture that Spark offers will not only make it possible to analyse massive datasets quickly and effectively, but with the use of Deep Learning,it will also increase the scalability and performance of the system.

1. Introduction

The introduction of computers into our modern environment fifty years ago caused a shift in the way that data was understood in comparison to the way that data is understood in our current environment as a direct result of the introduction of computers into our modern environment. As a result of the exponential increase in the amount of data, the recommendation domain has also been subject to significant transformation. The astronomical increase in the quantity of data has been the driving force behind this paradigm shift. The success of major companies like Amazon, Flipkart, Instagram, Netflix, and Spotify, amongst others, is dependent on the efficacy of their respective recommendation systems and engines as a means of increasing user engagement on their respective platforms. This is the case because successful recommendation systems and engines are the key to increasing user engagement. The reason for this is that these businesses are in competition with one another to provide the best possible experience for their customers. It really shouldn't come as a surprise that recommendation algorithms are so widely used in popular media streaming websites like Netflix. These algorithms provide users with assistance in locating the most appropriate movies and television shows to watch based on their viewing history and the preferences they have selected. Within the framework of the J component project, one of our goals is to carry out additional research into the movie recommendation system in order to gain a deeper comprehension of it. Because the level of complexity that can be found in machine learning and big data continues to increase, investigations and research are continually being carried out into the types of systems that were previously mentioned.

As part of the service that is known as "recommendation," users are provided with recommendations, which may be in the form of content pieces or other types of goods. This is the fundamental idea that underlies the process of suggesting something to someone else. The user should have the ability to select their preferences and options, and the suggestions should be crafted in such a way that they are consistent with those choices. The user should have the ability to select their preferences and options from the drop-down menu. This system drew its inspiration from the human practise of making recommendations to one another; however, in this case, an algorithm will make such recommendations on your behalf instead of a human. This system was conceptualised after the human practise of offering suggestions and advice to one another. Even though the actions of users are a crucial part of the procedure for developing recommendation systems, these programmes can generally be broken down into two primary categories: those that are based on the presentation of content and those that are based on the participation of users in collaborative processes. The actions and behaviours of users are a crucial component.

A content-based movie recommendation system is a type of recommendation system that makes use of various components and qualities of movies in order to suggest other movies of a similar nature to a user on the basis of that user's prior preferences and actions. This type of recommendation system may also be referred to as a content-based movie recommendation service or a content-based movie recommendation platform. This type of technology generates suggestions by analysing movie content, such as genres, actors, and directors, and comparing it to movies that a user has previously rated favourably. Specific story details are also taken into consideration. The final product is a list of suggestions or recommendations. After that, the system will provide recommendations for movies that have similar qualities to those that the user has previously loved and found enjoyable. For example, the system may suggest movies that are similar to those that the user has found enjoyable. They are simple to set up and have the potential to deliver powerful suggestions to users who have rated a significant number of movies; however, they may struggle to create recommendations for new users who have limited movie preferences. This is because they have the potential to deliver powerful suggestions to users who have rated a significant number of movies.

The second type of technology is a recommendation programme that uses collaborative filtering as its foundation. These are a subset of recommendation systems that offer users customised recommendations based on the preferences and actions of people who are most similar to themselves as a whole. This method does not place any requirements on the suggested products with regards to their content or overall quality. These two factors are not taken into account. Instead, it makes recommendations to users based on the relationships between them and the interests they have in common with one another. The term ”collaboration” refers to more than just working together on a single project; it also encompasses working together on user-based and item-based projects respectively. It is possible to subdivide each of these groups into even more specific categories. It is possible that this will help people who already have well-established preferences get recommendations that they like, but it may be difficult for new users with little history to get recommendations that they like using this method. They may also have difficulty in contexts where there is a dearth of trustworthy ratings or a great deal of background noise in the data. This indicates that they may have difficulty succeeding in environments in which there are few opportunities to earn ratings.

2. Motivation

The idea for the project came about as a result of the fact that researchers have just scratched the surface of what may be accomplished in the area of recommendation engines. The reason for our decision to work in the film industry was so that we would have the opportunity to discuss the following topics: -

2.1. Enhancing the User Experience

Our software is able to enhance the user experience by offering individualised and pertinent movie suggestions. These recommendations make it simpler for users to find new and intriguing films.

2.2. Handling Large Datasets

Spark is an ideal platform for movie recommendation systems that need to analyse large amounts of data, such as movie ratings and reviews, because it is designed to handle large and complex datasets. This makes Spark an ideal platform for movie recommendation systems. Because of this, spark is an excellent option for software that makes movie recommendations. We would have the opportunity to learn everything there is to learn about the management of large amounts of data if we utilised spark. This would be the case because spark would provide us with the opportunity to l earn everything.

2.3. Integration with other big data technologies

The system is capable of being integrated with other big data technologies, such as Hadoop and No SQL databases, through the utilisation of Spark to produce a comprehensive big data analytics framework for movie recommendation systems.

3. Literature Review

In this section of the article, we are going to provide an overview of the many different types of movie recommendation systems, some of which include collaborative filtering, content-based filtering, and hybrid recommendation systems, amongst others. In addition to this, we go over a variety of machine learning algorithms, such as clustering, deep learning, and matrix factorization, which are utilised in recommendation systems that were utilised by the writers who came before us. For example, we discuss how clustering can be used to categorise items in a database. Clustering, deep learning, and matrix factorization are all examples of these types of algorithms. The authors who came before us made use of these methods to write their works. (R. Wang et al.)

After conducting research on the subject, we came up with the idea of a hybrid recommendation system as a potential solution to the problem of providing customers with a selection of movies to choose from. Within the framework of this system, various filtering methodologies, including content-based filtering and collaborative filtering, would be utilised. (Y. Wang, M. Wang, Xu, et al.) Finding users who are similar to one another based on their previous activities and then recommending items that those users have shown an interest in is what collaborative filtering entails. The first step in this process involves locating users who are comparable to one another, and the next step involves suggesting products. On the other hand, content-based filtering involves recommending to a user things that are comparable to things in which the user has previously shown an interest. In this particular instance, the user may have indicated an interest in a movie, a book, or an article. (Roy, Dutta, et al.)

Incorporating sentiment analysis of user ratings into the content-based filtering methodology yields an improved hybrid recommendation system as a result of the research. (Singh et al.) A method known as sentiment analysis is one that can be implemented in order to determine the tone of user feedback. After the sentiment has been identified, the method can be utilised to make recommendations regarding films that contain a sentiment that is analogous to the sentiment that was initially identified. (Agrawal, Jain, et al.)

Recent studies have demonstrated that hybrid recommendation systems do better than either content-based or collaborative filtering on its own. (Zhang et al.) Hybrid systems are able to capitalise on the benefits afforded by both techniques in order to provide customers with suggestions that are even more specific. This allows hybrid systems to provide customers with recommendations that are more likely to meet their needs. It has been shown that an improvement in the accuracy of movie recommendations occurs when sentiment analysis is included into the process of content- based filtering. (Furtado, H, et al.)

During the course of the research project, a big data analytics platform was utilised so that the substantial amount of data that was generated throughout the motion picture recommendation procedure could be effectively managed. A distributed file system was utilised by the framework for the purposes of storing the data as well as processing it. In addition to that, the framework made use of the MapReduce algorithm in order to perform concurrent and distributed analyses of the data. (Gan, Cui, et al.)

In one of the techniques, we provide a sentiment-enhanced hybrid recommendation system for suggesting movies by making use of a framework for big data analytics. This way is only one of the many that we offer. The hybrid system employs methodologies for collaborative filtering and content-based filtering, which allows it to provide users with trustworthy recommendations. (M, N, et al.) These methodologies are described in more detail here. The addition of sentiment analysis into the content-based filtering process contributes significantly to an increase in the usefulness of recommendations. (Halder, Sarkar, Y.-K. Lee, et al.) The findings of this study also highlight how essential it is to make use of a big data analytics framework in order to effectively handle the substantial quantity of data that is involved in the process of making recommendations. (Nanou, Lekakos, and Fouskas) The approach that was suggested will have significant repercussions for the movie recommendation systems that are already in place, and it has the potential to be utilised in such a way that it will provide users with recommendations that are accurate as well as tailored to their particular requirements. (K. Lee et al.)

4. Methodology

The movie recommendation system that makes use of the MovieLens dataset typically takes the form of a collaborative filtering method, which is a common way that is used. This is because the MovieLens dataset contains a large number of movie ratings and reviews. This is due to the fact that the dataset provided by MovieLens contains a significant number of ratings and comments provided by users. It is called "collaborative filtering," and it is a method that takes the actions of users (such as rating movies), analyses those actions to find patterns in the users' behaviour, and then employs those patterns to create recommendations for other users. This process involves taking the actions of users (such as rating movies), analysing those actions to find patterns in the users' behaviour, and then employing t hose patterns to create recommendations for other users (such as rating movies). It is predicated on the hypothesis that individuals who have previously demonstrated a preference for a specific film genre are likely to continue to have preferences that are comparable to those genres in the future if they have previously demonstrated a preference for that film genre in the past. This is because individuals who have previously demonstrated a preference for a particular film genre tend to have preferences that are comparable to those genres. The purpose of this survey is to put this hypothesis to the test by gathering responses from individuals who have indicated in the past that they have a preference for a particular film genre.

The overarching concept of collaborative filtering may be broken down into many distinct subcategories, the two most important of which are user-based filtering and item- based filtering. The technique that is involved in user-based collaborative filtering involves finding other users on the internet who have interests in movies that are similar to those of the target user and then recommending movies that these other users have loved. Item-based collaborative filtering, on the other hand,includes locating movies that are similar to the ones that the target user has loved and proposing them. This may be done by comparing the two sets of movies.

Content-based filtering is yet another method that is often used in recommendation systems for motion films. [Case in point:] Using content-based filtering to identify which movies should be suggested to customers requires looking at characteristics of movies (such as the genre, the actors, and a summary of the plot) in order to figure out which movies should be recommended. This method is built on the concept that customers who liked one movie are likely to appreciate other movies that have traits in common with the first movie they liked, and that these similarities will increase the likelihood of the customer purchasing further movies.

Over the course of the past few years, there has been a meteoric rise in the number of individuals opting to make use of hybrid recommendation systems. When it comes to providing recommendations, these systems combine content- based filtering with collaborative filtering to produce a more encompassing and well-rounded experience for users. These kinds of systems are able to make the most of the benefits provided by both strategies in order to provide customers with recommendations that are more accurate. Matrix factorization, deep learning, and ensemble techniques are just a few examples of the many different approaches that have been utilised in the process of developing movie recommendation systems. There are a great number of other approaches as well.

5. Proposed Idea

Within the bounds of this investigation, we propose making use of deepFM with the intention of creating automated movie recommendation systems. DeepFM is a recommendation system that is powered by deep learning and is based on the combination of factorization machines (FM) and deep neural networks. Deep learning is the engine that drives DeepFM. The combination of these two technologies is what gives DeepFM its moniker, "DeepFM" (DNN). Deep neural networks are able to learn intricate features, whereas factorization machines can only express the relationships between features themselves. Additionally, deep neural networks are able to learn the connections between individuals and films.

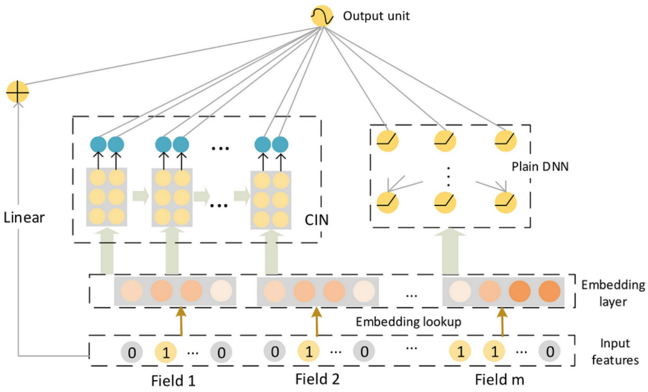

FIGURE 1. **DeepFM Architecture**

The system that is being proposed would provide recommendations by taking into consideration the past viewing patterns, ratings, and preferences of users in addition to the characteristics of movies, such as the film's genre, director, and cast members. At the beginning, the system will make use of FM to model the interactions between the user features and the movie features. After that, it will transmit the output to a deep neural network, which will learn complex user features as well as the linkages between users and movies. Following that, the system will provide watching recommendations based on the information that it has acquired about the qualities and ratings of various movies.

(Note: The image above 'FIGURE 1' was taken from **https://towardsdatascience.com/extreme-deep-factorization-machine- xdeepfm-1ba180a6de78)**

6. Conclusion

In this study, we made the suggestion that movie recommendation systems should make use of deep FM. The combination of factorisation machines and deep neural networks that is used in DeepFM gives the system the ability to understand intricate characteristics and connections between users and movies. The user's watching history, ratings, and preferences, in addition to the characteristics of movies, would all be taken into consideration by the suggested system when making suggestions. The suggested method has the potential to overcome the constraints of previously implemented recommendation systems while also improving the accuracy of suggestions.

7. Acknowledgement

We would like to recognise the research papers, publications, and other materials that were utilised to build the concepts that are given in this overview work on the use of deepFM for movie recommendation systems. Throughout the process of developing and delivering this proposal, the information and insights obtained from these sources proved to be quite helpful. I would also like to take this opportunity to extend my appreciation to the research community for the contributions that they have made to the field of recommendation systems. These contributions have made it possible for users to access movie recommendation systems that are both more accurate and more tailored to their individual preferences.

8. Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the pa-per was free of plagiarism.

References

Agrawal, Shreya, Pooja Jain, et al. "An improved approach for movie recommendation system". *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (2017): 336–342.

Furtado, F, A S I N G H, et al. "Movie recommendation system using machine learning". *International Journal of Research in Industrial Engineering* 9.1 (2020): 84–98.

Gan, Mingxin, Hongfei Cui, et al. "Exploring user movie interest space: A deep learning based dynamic recommendation model". *Expert Systems with Applications* 173 (2021): 114695–114695.

Halder, Sajal, A M Jehad Sarkar, Young-Koo Lee, et al. "Movie Recommendation System Based on Movie Swarm". *2012 Second International Conference on Cloud and Green Computing* (2012): 804–809.

Lee, Kyuhan, et al. "Predicting movie success with machine learning techniques: ways to improve accuracy". *Information Systems Frontiers* 20.3 (2018): 577–588.

M, Goyani, Chaurasiya N, et al. "A Review of Movie Recommendation System: Limitations, Survey and Challenges ". *Electronic Letters on Computer Vision and Image Analysis* 19 (2020): 18–37.

Nanou, Theodora, George Lekakos, and Konstantinos Fouskas. "The effects of recommendations' presentation on persuasion and satisfaction in a movie recommender system". *Multimedia Systems* 16.4-5 (2010): 219–230.

Roy, Deepjyoti, Mala Dutta, et al. "A systematic review and research perspective on recommender systems". *Journal of Big Data* 9.1 (2022): 59–59.

Singh, Pradeep Kumar, et al. "Recommender systems: an overview, research trends, and future directions". *International Journal of Business and Systems Research* 15.1 (2021): 14–14.

Wang, R, et al. "Deep & cross network for ad click predictions". 2017.

Wang, Yibo, Mingming Wang, Wei Xu, et al. "A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework". *Wireless Communications and Mobile Computing* 2018 (2018): 1–9.

Zhang, Jiang, et al. "Personalized real-time movie recommendation system: Practical prototype and evaluation". *Tsinghua Science and Technology* 25.2 (2020): 180–191.

Embargo period: The article has no embargo period.

To cite this Article: , Chandan Kumar Sangewar, Chinmay Pagey , Sakshi Chauhan , and Suganeshwari G . "A Review on the Movie Recommendation System Using Big Data." International Research Journal on Advanced Science Hub 05.05S May (2023): 376–381. http://dx.doi.org/10.47392/irjash.2023.S051