Article

Multimodal Movie Recommendation System Using Deep Learning

Yongheng Mu [1,2] and Yun Wu [1,*]

1 Hubei Key Laboratory of Intelligent Geo-Information Processing, College of Computer Science, China University of Geosciences, Wuhan 430078, China
2 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
* Correspondence: wuyun@cug.edu.cn

Abstract: Recommendation systems, the best way to deal with information overload, are widely utilized to provide users with personalized content and services with high efficiency. Many recommendation algorithms have been researched and deployed extensively in various e-commerce applications, including the movie streaming services over the last decade. However, sparse data cold-start problems are often encountered in many movie recommendation systems. In this paper, we reported a personalized multimodal movie recommendation system based on multimodal data analysis and deep learning. The real-world MovieLens datasets were selected to test the effectiveness of our new recommendation algorithm. With the input information, the hidden features of the movies and the users were mined using deep learning to build a deep-learning network algorithm model for training to further predict movie scores. With a learning rate of 0.001, the root mean squared error (RMSE) scores achieved 0.9908 and 0.9096 for test sets of MovieLens 100 K and 1 M datasets, respectively. The scoring prediction results show improved accuracy after incorporating the potential features and connections in multimodal data with deep-learning technology. Compared with the traditional collaborative filtering algorithms, such as user-based collaborative filtering (User-CF), item-based content-based filtering (Item-CF), and singular-value decomposition (SVD) approaches, the multimodal movie recommendation system using deep learning could provide better personalized recommendation results. Meanwhile, the sparse data problem was alleviated to a certain degree. We suggest that the recommendation system can be improved through the combination of the deep-learning technology and the multimodal data analysis.

Keywords: recommendation system; deep learning; matrix factorization; multimodal technique

MSC: 68T07

1. Introduction

Nowadays, with the popularization of information society, we have easier to access more information than we have ever had in all of human history. However, according to the decision-making theory, too much information can lead to worse decisions, rather than making people happy and ensuring they get what they want [1–3]. Actually, information overload, an artifact of the digital and mobile revolution, becomes more and more serious. When confronted with massive network resources, users are facing difficult decisions with the problem of information overload, which can cause a real feeling of anxiety, mental fatigue, powerlessness, and overwhelm. Recommendation systems are information filtering systems that use artificial intelligence or AI algorithms that can greatly alleviate the problem of information overload by filtering Big Data with high efficiency to provide users with personalized content and services, which could greatly alleviate the problem of information overload [4,5]. A recommendation system is not only a fancy algorithm, but it is also about understanding the data and your users. Recommendation systems first arose in the mid-1990s and have been a focus of research ever since. Nowadays, many platforms and websites are witnessing the extensive application of different types of recommendation

systems. E-commerce websites have been using recommendation systems to suggest personalized products and services ranging from articles, books, music, and movies [6].

Typically, there are two types of data used in a typical recommendation system, i.e., the rating or buying behaviors that reflect the attributable information about the items, users, and keywords or textural profiles that record the user-item interactions [7]. According to the data used for inference, most recommendation systems can be divided into three categories: content-based filtering, collaborative filtering, and hybrid recommendation systems. Collaborative filtering models analyze commonalities and similarities among different users based on the ratings and then estimate new recommendations according to the relationships between users. By contrast, content-based filtering models use the contributions of an item to recommend other items similar to the user's preferences and are mainly concerned with the rating of one user instead of multiple users. Hybrid recommendation systems integrate the advantages of the above two types to improve more comprehensive algorithms [5,8,9]. Beyond these three types, many other recommendation techniques, including social network-based, knowledge-, utility-, demographic-, context awareness-, model-, and trust-based methods have been proposed in research and practice ever since.

However, each recommendation algorithm has its advantages and limitations. Nowadays, there are still many problems encountered in many recommendation systems, especially the data sparseness, cold start, lack of diversity in recommendation results, and the limitations of information expiration [9,10]. As a branch of machine learning, deep learning is essentially a neural network with multiple layers mimicking the human brain. Deep learning can mine deeper information between input features and, therefore, provide users with satisfactory recommendations even without complete information. In addition, in the last decade, deep-learning algorithms have been integrated in the field of recommendation systems and continue to prove to be effective in dealing with different recommendation tasks [11–16].

Here, we suggest a personalized multimodal movie recommendation system integrated with multimodal data analysis and deep learning. The MovieLens 100 K and 1 M datasets were selected to test and evaluate the proposed algorithm. In this model, we use deep-learning algorithms to mine and fuse multiple source data, including multi-users and polytype movies. Compared with traditional algorithms, the scoring prediction results show improved accuracy after incorporating the potential features and connections in multimodal data with deep-learning technology.

The main contributions of this paper are summarized as follows:

(1) Introducing a novel movie recommendation system through the combination of the deep-learning technology and the multimodal data analysis.
(2) Experimental results for the proposed recommendation system using the MovieLens 100 K and 1 M datasets have been reported, which have achieved good performance.
(3) We suggest that multimodal data, such as movie posters, could improve the performance of the recommendation system.

The structure of this paper is established as follows. The related works about the movie recommendation systems are reviewed in Section 2. In Section 3, the theoretical framework and the algorithm of our proposed movie recommendation system are introduced. In Section 4, experimental results, as well as the data analyses about the proposed methods, are demonstrated. Finally, Section 5 presents the discussion and conclusions of this paper.

2. Related Works

In this age of information, recommendation systems have changed the style of searching the things of our interest. The movie recommendation system is one of the most fascinating applications, and also one of the most lucrative. Many online video platforms such as YouTube, Netflix, TikTok, and Tencent Video deploy recommendation systems to help billions of users discover personalized contents from an ever-growing corpus of

movies and videos [15–18]. As shown in Figure 1, many kinds of algorithms have been tried and tested in various movie recommendation systems over the last two decades.
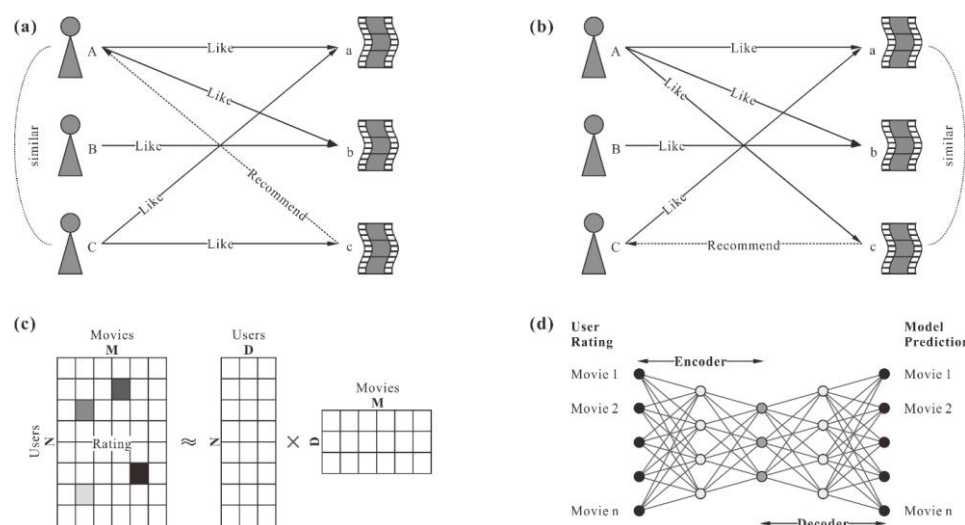

Figure 1. Typical recommendation methods. (**a**) User-based collaborative filtering, (**b**) Item-based collaborative filtering, (**c**) Model-based collaborative filtering, (**d**) Deep-learning neural networks.

When a typical content-based movie recommendation system is used, the user will receive some recommended movies that are very similar to those the user has watched before [8]. In this type of system, movies are generally grouped into different categories according to their similarities, including the type, actor, director, etc. On this basis, the systems can recommend some potential movies to the user based on the user's viewing preferences recorded in databases such as MovieLens. The movie recommendation systems, using content-based filtering, rely on user-annotated metadata, i.e., the text view of movies [19]. A large set of user-movie pairs are collected and labeled into different levels according to users' action on movies, which are trained to predict whether a user is interested in a movie or not. This kind of algorithm does not involve the similarities of the user community. It has some shortcomings, such as low accuracy, narrow applicability, and lack of novelty [20]. This recommendation algorithm will always recommend items that the user originally liked. In other words, it is difficult to recommend novel content to users. Furthermore, we noted that the user-annotated metadata is incomplete for many online movies. In order to increase the content representation of items in a content-based recommendation system, many different algorithms such as text analysis, leveled features, and social tags have been introduced. For example, to train an effective recommender system with a lower annotation cost, people applied some improved algorithms, such as a multi-view active learning framework for automatic video annotation [17]. A movie tag prediction algorithm was proposed to segment movies according to the predicted tags and to predict relevant tags for movies [21]. However, if the user has not rated a sufficient number of movies, the movie recommendation systems cannot really understand the user's preferences. In other words, it will not give satisfied personalized recommendations without sufficient useful features extracted from existing database. From this, it is suggested that a new user with limited ratings cannot receive satisfying recommendations [10].

Recommendations based on collaborative filtering, also known as CF recommendations, are the most widely used and most successful algorithms. The CF recommendations are simple and easy to use. In this kind of recommendation system, the user will receive some recommended items that are very similar to the user's previous tastes and aesthetics [8]. It not only considers the features of the user or the item itself, but it also considers the interactive information, such as ratings, reviews, purchases, and clicks, between the user and the item. The CF recommendations have many branches according to different methods of calculating similarity. Among them, item-based collaborative filtering (item-

based CF) and the user-based collaborative filtering (user-based CF) are the most famous two [6,20–23]. In a user-based CF movie recommendation system, it will calculate the degree of similarity between users through their preferences and then recommend movies to different categories of users based on their similarities (Figure 1). By comparation, an item-based CF movie recommendation system will calculate the degree of similarity between movies based their characteristics, including type, actor, and director, and then make some recommendations to users according to their record in the database (Figure 1b). Like many movie recommendation systems, the MovieLens application will record and store ratings of watched movies by different users. The CF recommendation systems collect and analyze the ratings for users and then present so called top-N-related movies to suit a user's personal requirements. The efficiency and accuracy of the CF recommendations depend on the sample size and the degree of difference among the users and the items. Therefore, such algorithms usually encounter data scarcity and cold-start problems [24] and are less suitable for online videos. Furthermore, according to the theory of decision-making, a user's process of decision-making involves many factors, including information quantity, information quality, environmental situations, and psychological states [1,3]. As a result, many traditional movie recommendation systems cannot produce satisfying results for users, especially the newcomers.

In order to solve the data sparsity and the cold-start problems, people have proposed a model-based collaborative filtering (model-based CF) method using deep learning, machine learning, and other technologies [6,20,23–25]. The most common model-based CF algorithms include matrix decomposition (MF), singular-value decomposition (SVD), and improved Funk-SVD algorithm, SVD++ algorithm, etc. In model-based CF methods, people have used matrix factorization to alleviate the data sparsity in collaborative filtering algorithms. For a movie recommendation system based on model-based CF, matrix factorization is conducted according to the user-movie rating matrix in which both movies and users are integrated into a joint factor space with lower dimensionality (Figure 1c). There are a number of advantages in model-based recommender systems such as space-efficiency, training speed and prediction speed, and avoiding overfitting [7]. However, it is suggested that matrix factorization still has problems in a sparse environment, the overfitting problem will become more severely with an increasing spare of the rating matrix [26]. In recent years, deep leaning has achieved rapid development in the field of recommendation systems. Convolutional neural networks (CNN) and deep-learning algorithms learn the underlying patterns in complex data during training and have exhibited outstanding performance in discovering hidden features in dataset [6,25–28]. There are two main parts in this kind of movie recommendation system (Figure 1d): the encoding part is the first half layer of the network, and the decoding part is the other half layer. At first, raw content information of all movies with user ratings are processed as vectors to obtain hidden features by multiple denoising autoencoders. Then, the feature representations of the noise-corrupted input information are learned in the encoding part and then made into a model prediction in the output. It has been proved that the cold start problem could be effectively alleviated in a recommendation system based on deep learning and collaborative filtering [24]. Actually, many researches indicated that the movie recommendation systems based on deep-learning techniques can work more efficiently with high accuracy [10,12,18,29]. Furthermore, it is important to notice that a movie contains multiple types of modalities, including text, image and audio features. Multimodal deep learning has a unique ability in processing and linking information based on various modalities. Therefore, multimodal deep learning is paving the way for better presentations to be extracted from unstructured multiple types of data [30]. Recently, some people have suggested that the incorporation of multimodal data such as audio, text, and image features will further improve the performance of movie recommendation systems [21,29–32].

3. Proposed Recommendation System

In this section, we propose here the system architecture of our movie recommendation system based on deep-learning technology and multimodal data analysis. This modal excavates the attribute characteristics of users and items in the dataset and integrates them into the recommendation system, combines the scoring data, trains the neural network built, and, finally, predicts the user's score of the movie more accurately, which is significantly improved compared with the traditional collaborative filtering algorithm.

3.1. Framework of the Proposed Model

The overall process model of our movie recommendation system with deep learning and multimodal data is shown in Figure 2. The input to the network is a dataset containing multimodal information of users and movies. The output is a top-N list of recommended movies for a user. First, the parameters of users and movies are transformed into single-value matrixes that contain non-zero singular values. Second, the CNN with multiple layers of convolving filters is trained to improve the level classification of data. Then, the recommendation system with a trained model utilizes the refined features to find the potential relationships between users and movies based on similarity criteria. The content similarities are further refined through several steps including removing redundancy, scores assignment, normalization, and filtering. Finally, top-N movies are recommended for a user based on similarity theory.
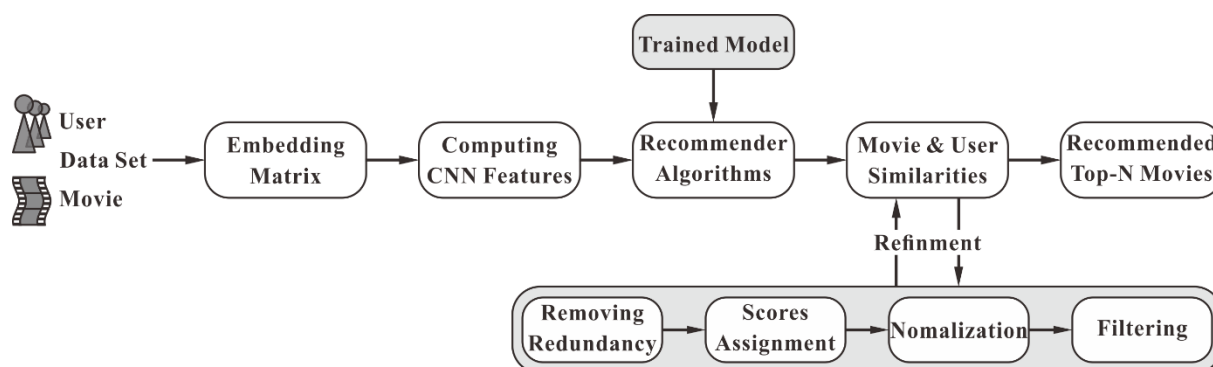

Figure 2. The framework of the movie recommendation system with deep learning.

3.2. Feature Extraction

In this present work, we utilize CNN to mine hidden features of users and movies from the MovieLens datasets. CNN is a variant of feed-forward neural networks with three main parts, i.e., a convolution layer that separates and identifies various features of input data, a pooling layer that condense data by selecting representative local features from previous convolution layer, and a fully connected layer. The fully connected layer multiplies the input by a weight matrix and then adds a bias vector. In comparison with traditional neural networks, CNNs may contain hundreds of hidden layers that are generally well-suited to discover intricate patterns or features in complex data without a given mathematical model [6,33].

The model CNN applied in our movie recommendation system is a slight variant of the CNN architecture proposed by Collobert et al. [33]. It consists of four parts, including the input layer, pooling layer, convolution layer, and output layer (Figure 3). The input layer transforms raw data into a dense numeric 32×32 matrix that represents the data for the next convolution. Three convolution layers are used to extract contextual features of input data from the MovieLens dataset, which are designed as six layers with 32×32 dimensions (6 @ 32×32), 16 layers with 10×10 dimensions (16 @ 10×10), and 120 layers, respectively. Two polling layers, set as 6 @ 14×14 and 16 @ 5×5, were used to extract representative features from the convolution layers. Finally, the output layer generates top-10 recommendations for the program. In practice, we selected four parameters (movie ID, type, title, and poster)

of movies and four parameters (user ID, gender, age, and profession) of users as input data to generate initial matrices for subsequent feature extraction processes (Figure 4).
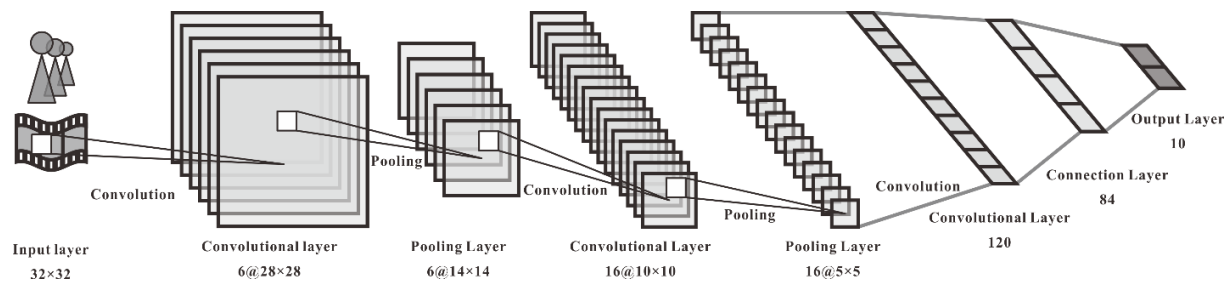

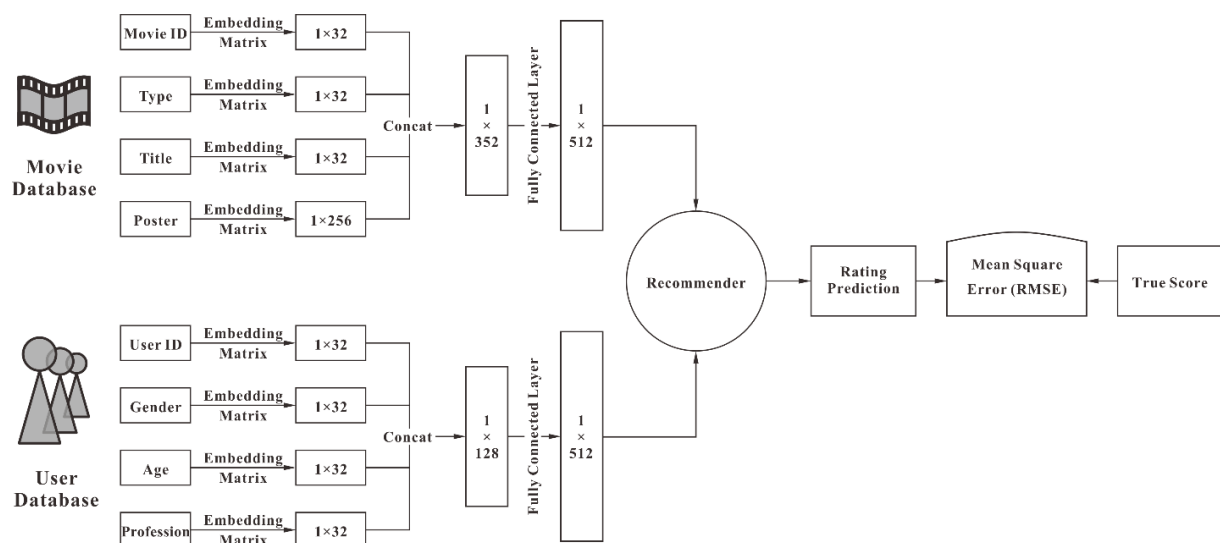**Figure 3.** The CNN architecture in the proposed movie recommendation system.

Figure 4. Score prediction model of the movie recommendation system based on neural network.

4. Experiments

The movie rating data from MovieLens were used to test the efficiency of our proposed algorithm. Three traditional classical algorithms, i.e., item-based and content-based filtering (Item-CF), user-based collaborative filtering (User-CF), and singular-value decomposition (SVD) algorithms, were compared with the proposed multimodal movie recommendation system based on deep learning and multimodal data analysis.

4.1. Dataset Introduction

The real-world MovieLens 100 K and 1 M datasets (https://movielens.org/, accessed on 18 May 2022) were used to test the effectiveness of the proposed recommendation method. The MovieLens datasets, describing people's expressed preferences for movies, were first released in 1998. The preferences use the form of tuples, in which a person expressing a rating (0–5 star) for a watched movie at a particular time. The MovieLens 100 K, released in 1998, is a large dataset with 100,000 ratings from 1682 movies and 943 users. The MovieLens 1 M dataset, released in 2003, contains about 1-million ratings from 6040 users on 3883 movies. It should be noted that there are plenty of movies in these datasets that have no ratings for. The data scarcities for the MovieLens 100 K and 1 M datasets are 93.695% and 95.359%, respectively (Table 1). The rating data in MovieLens include timestamps and explicit feedbacks [6], so we assume that the ratings are actually click events. In those cases, the labels in the dataset are assigned with 1, which represents an interaction between the user and movies.

Table 1. Summary of experimental data.

Dataset	Parameters	Value
MovieLens 100 K	#Users	943
	#Movies	1682
	#Ratings	100,000
	Sparsity	93.695%
MovieLens 1 M	#Users	6040
	#Items	3883
	#Ratings	1,000,209
	Sparsity	95.359%

4.2. Evaluation Indicators

Train/Test is a method to test the accuracy of the proposed model [6]. In this research, we split the MovieLens datasets into two parts, i.e., training sets and testing sets. The ratios of Train/Test were set as 6:4, 7:3, and 8:2, respectively. The learning rate, an important hyper parameter for tuning neural networks, was set as 0.001 and 0.0001, respectively, to evaluate the rate at which the proposed neural network learns. Following previous works [34,35], we calculate the RMSE scores based on the rating prediction and the true scores generated using different recommendation models (Figure 4). The smaller RMSE scores correspond to better prediction accuracies. To run the experiment, we used a computer with Intel core i5-8300H CPU @ 2.3 GHz, GeForce GTX 1050 GPU, 8 GB RAM, and Windows 10 as the operating system. The programming language is the Python 3.6 developed by Python Software Foundation and the CUDA 10.2 invented by NVIDIA corporation.

4.3. Comparing Deep Learning Algorithm with Others

The performances of our proposed algorithms based on deep learning and multimodal data were compared with three traditional classical algorithms, including basic Item-CF, User-CF, and SVD models. We use crawler technology to crawl movie posters as visual mode data of movies. In our proposed recommendation system, the "Multimodal" algorithm utilizes deep-learning technology to dig out the features and integrate into the text features. To compare the efficacy of the movie posters to the performance of our proposed model, we did not input the posters as visual mode data when using the "Deep Learning" algorithm. A comparation and analysis on different algorithms were made based on RMSE scores using the MovieLens 100 K dataset (Figure 5). From this figure, we see that our proposed models based on deep learning, whether introducing movie posters ("Deep Learning") or not ("Multimodal"), have much smaller RMSE scores compared to the classical Item-CF, User-CF, and SVD models. However, we noted that the scoring prediction results have not been improved obviously while introducing the crawled movie posters into our proposed recommendation system.

4.4. Influence of the Data Size and the Learning Rate

The learning rate in machine learning plays an important role in neural-network training. It is a hyperparameter used to control the pace of training of neural networks. The values of the learning rate vary between 0.0 and 1.0. Normally, a small value of the learning rate will lead to a long training process, whereas a large value will lead to an unstable training process [6]. In our models, we set up the system optimizer with a learning rate of 0.001 or 0.0001. The RMSE scores indicate that our proposed system achieved good performance with these reset learning rates values (Table 2). When using the MovieLens 100 K dataset, our proposed deep-learning recommendation algorithm achieves 0.9908 RMSE scores for the train set and 1.0263 for the test set, respectively. By comparison, the preset value of 0.0001 is slightly better (Figure 6). In addition, our proposed movie recommendation system could have improved performance with more denser data. That is, the prediction effect can be gradually enhanced with increasing the proportion of training set data. Actually, if we integrate deep-learning technology and multimodal data analysis

into the movie recommendation system, the input dataset could be used more productively, alleviating the problem of sparse data and single modality to some degree, and makes the prediction of movie scoring more accurate, and further improves the performance of the recommendation system.
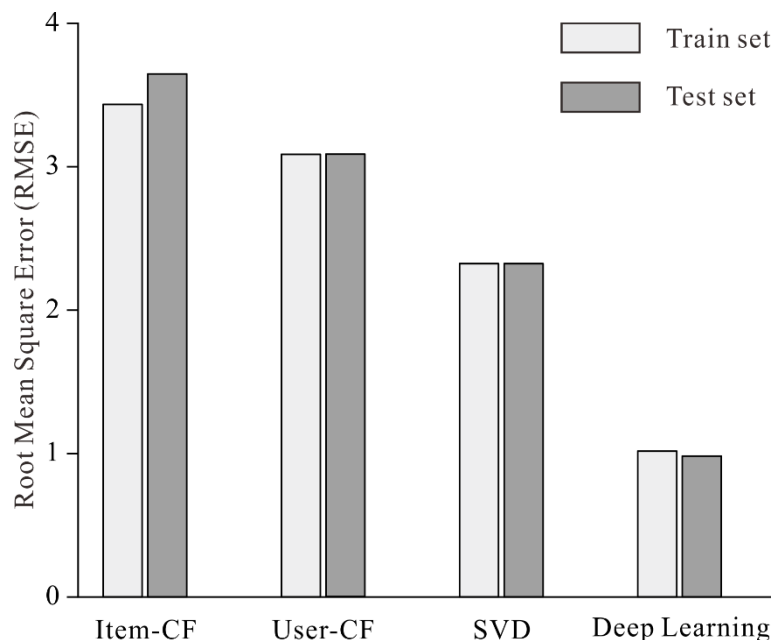

Figure 5. The RMSE scores for different algorithms based on the MovieLens 100 K dataset.

Table 2. RMSE scores of the multimodal movie recommendation algorithm based on deep learning.

	MovieLens 100 K		MovieLens 1 M	
Learning rate	0.001	0.0001	0.001	0.0001
RMSE of Train set	1.0263	1.1319	0.9264	0.9967
RMSE of Test set	0.9908	1.0877	0.9096	0.9807

4.5. Influence of Multimodal Data Analysis

The design goal for the movie recommendation system is to recommend specific user's favorite movies with high accuracy and efficiency, which helps users to alleviate the information overload. Therefore, the personalized reference service is one of the core concerns in the recommendation model. However, human decision-making is a complex dynamic process that is deeply influenced by a person's past experiences [3]. In a general scenario, a consumer behavior may be influenced by several factors, namely psychological, social, cultural, personal, and economic factors. As for movies, a user would prefer a certain movie based on his personal appetite, which was greatly influenced by the movie-watching experience. Movie posters are well-designed with visual signifiers according to the mood and genre of movies to attract an audience. Generally, movie posters are adjudged as successful only if they can grasp the right specific audience. For instance, if an audience enjoyed watching an action movie, he is more likely to select a movie that has similar poster designs as those he has watched before. The movie poster genres can be subdivided into different types based on visual similarities, such as comedy, blockbuster, cartoons, melodrama, science fiction, conclusion, and disaster movies.

In the proposed multimodal movie recommendation model using deep learning, we introduce the movie posters as extra input information to improve the performance of the recommendation system. A total of 1682 movie posters, in JPG format with resolution of 190×281, were crawled from the IMDB movie website (https://www.imdb.com/, accessed on 10 May 2022). These posters were firstly compressed to resolution of 128×128

and then embedded into the CNN to extract further hidden features of movies (Figure 3). The experiment was carried out on the MovieLens-100 k dataset and the 1682 movie posters, the number of training rounds was five rounds, the ratio of test set and training set was 8:2, the batch data size of each training was 256, the length of the embedding vector was 32, and the learning rate was set to 0.0001 and 0.001, respectively, for training. As shown in Figure 6, the horizontal coordinate in the figure represents the number of batches trained, and the ordinate coordinate represents the RMSE, which can be seen from the figure as the training batch increases; the RMSE of the predictive score continues to decline and eventually stabilizes, indicating that model training is effective. For examining the influence of the movie posters to the recommendation system, we compared two models for recommendation systems using deep-learning algorithms with or without movie posters based on the MovieLens 100 K dataset (Figure 7). The results show that after adding the movie posters, the recommendation algorithm based on multimodality has slightly improved the prediction effect of a movie score compared with the recommendation algorithm of single-mode deep learning. Therefore, we suggest that multimodal data, such as the visual movie posters, can greatly improve the performance of the recommendation system. In other words, if we want to make the recommendation system better, we can further collect more modal data to characterize users or movies and fuse the information of these multiple modes to jointly improve the accuracy and the performance of the recommendation system.
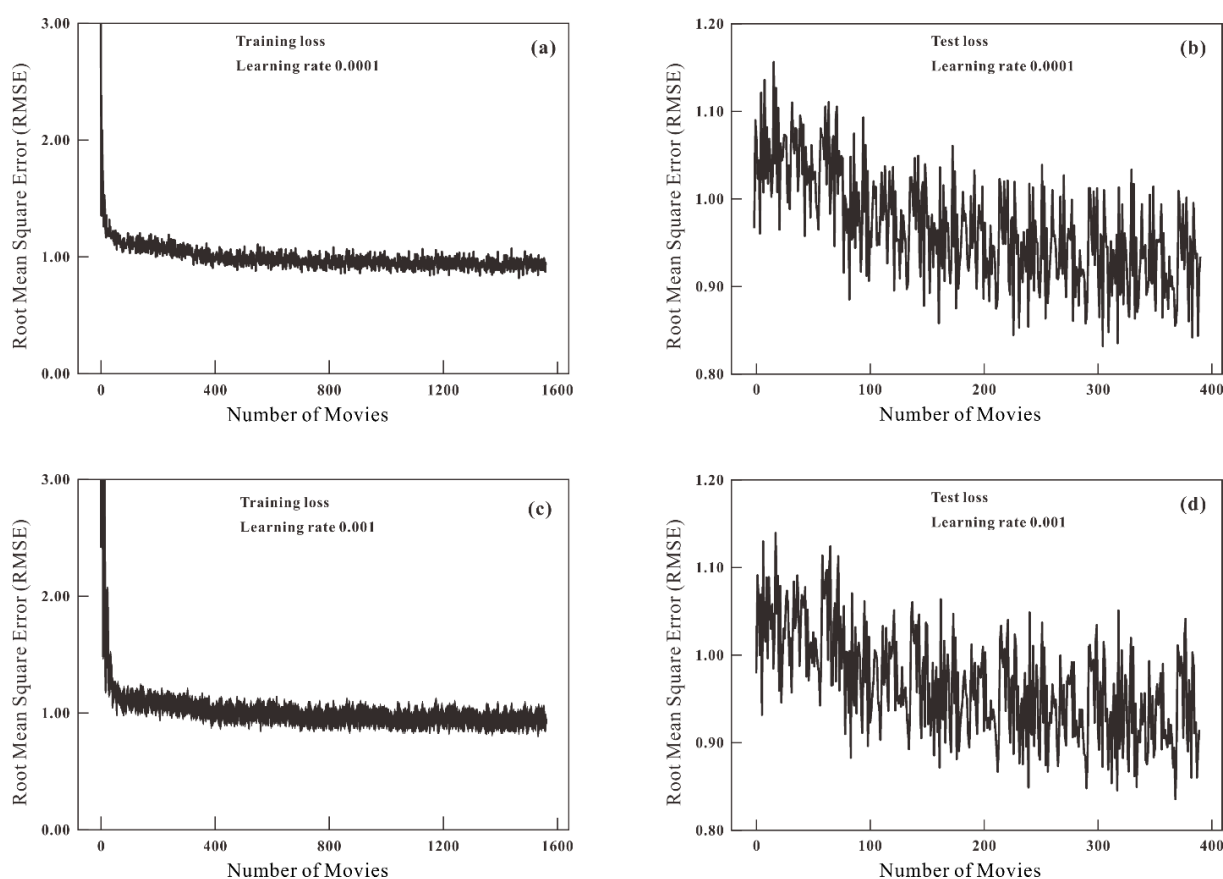

Figure 6. The RMSE scores of the training set (**a**,**c**) and test set (**b**,**d**) vs. number of movies for the multimodal recommendation systems using deep-learning algorithm with movie posters based on the MovieLens 100 K dataset.
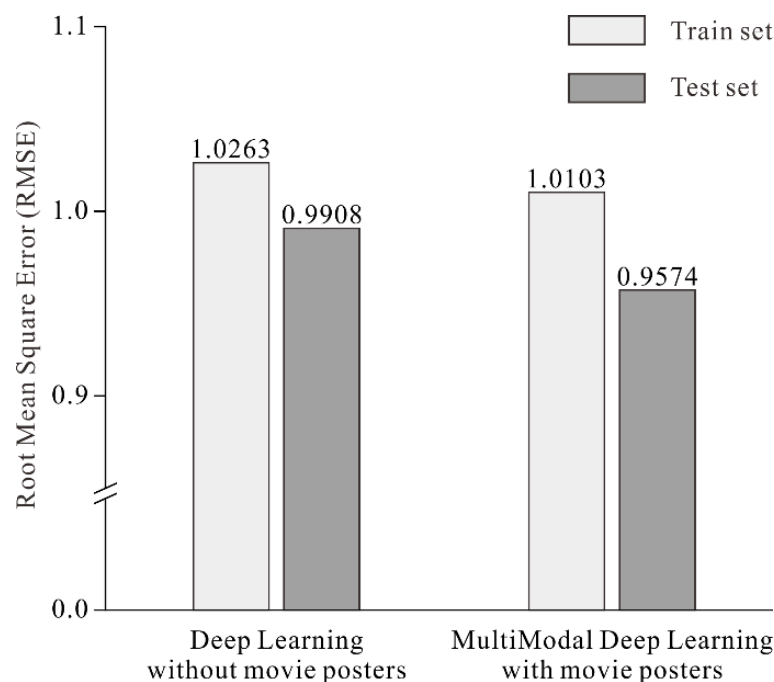
Figure 7. The RMSE scores for recommendation systems using deep-learning algorithm with or without movie posters based on the MovieLens 100 K dataset.

5. Discussion and Conclusions

Nowadays, with the widespread of use of the internet, people are facing the serious problem of information overload in their daily life. The recommendation systems can filter, prioritize, and deliver relevant information with high efficiency to provide users with personalized content and services, which could greatly alleviate the problem of information overload. In the past decade, we have seen the exponential growth of personalized recommendation systems for products and serves ranging from web pages, articles, books, music, and movies. One of the most fascinating applications is the movie recommendation system. However, many movie recommendation systems still have many problems, especially the data sparsity and the cold-start issue. Traditional algorithms such as collaborative filtering cannot mine the hidden information between input features. Therefore, it cannot provide users with satisfactory recommendations without complete information.

Deep learning is essentially an artificial neural network with multiple layers, having the ability to learn from large amounts of data. Deep learning, unlike traditional algorithms, can mine more hidden information from input dataset and, thus, provides users with better personalized recommendations. Here, we suggest a personalized multimodal movie recommendation system based on multimodal data analysis and deep learning. The datasets of MovieLens 100 K and 1 M were selected to test and evaluate the algorithm. The multimodal items of the movie (ID, type, title, and poster) and the user (ID, gender, age, and profession) were embedded in matrices. With the input information, the hidden features of the movies and the users were mined using deep-learning algorithms to build a deep-learning network algorithm model for training to further predict movie scores. When using the MovieLens 100 K and 1 M dataset, our proposed deep-learning recommendation algorithm achieves good performance based on RMSE scores. The results show improved accuracy after incorporating the potential features and connections in multimodal data with deep-learning technology. Furthermore, our system can alleviate the sparse data problem to some degree. Therefore, we suggest that incorporating deep-learning technology and multimodal data analysis with a movie recommendation system will produce better personalized customer service.

References

1. Jacoby, J.; Speller, D.E.; Berning, C.K. Brand Choice Behavior as a Function of Information Load: Replication and Extension. *J. Consum. Res.* **1974**, *1*, 33–42. [CrossRef]
2. Schwartz, B. *The Paradox of Choice: Why More is Less*; HarperCollins Publishers: New York, NY, USA, 2004; p. 304.
3. Chen, Y.; Shang, R.; Kao, C. The Effects of Information Overload on Consumers' Subjective State Towards Buying Decision in the Internet Shopping Environment. *Electron. Commer. Res. Appl.* **2009**, *8*, 48–58. [CrossRef]
4. Resnick, P.; Varian, H.R. Recommender Systems. *Commun. ACM* **1997**, *40*, 56–58. [CrossRef]
5. Isinkaye, F.O.; Folajimi, Y.O.; Ojokoh, B.A. Recommendation Systems: Principles, Methods and Evaluation. *Egypt. Inform. J.* **2015**, *16*, 261–273. [CrossRef]
6. Kane, F. *Building Recommender Systems with Machine Learning and AI*, 2nd ed.; Sundog Education: Winter Springs, FL, USA, 2021; p. 503.
7. Aggarwal, C.C. *Recommender Systems: The Textbook*; Springer: Berlin/Heidelberg, Germany, 2016; p. 498.
8. Adomavicius, G.; Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 734–749. [CrossRef]
9. Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G. Recommender System Application Developments: A Survey. *Decis. Support Syst.* **2015**, *74*, 12–32. [CrossRef]
10. Liu, J.; Choi, W.; Liu, J. Personalized Movie Recommendation Method Based on Deep Learning. *Math. Probl. Eng.* **2021**, *2021*, 6694237. [CrossRef]
11. Lund, J.; Ng, Y. Movie Recommendations Using the Deep Learning Approach. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 47–54.
12. Chavare, S.R.; Awati, C.J.; Shirgave, S.K. Smart Recommender System Using Deep Learning. In Proceedings of the 2021 IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 590–594.
13. Gopi, P.S.S.; Karthikeyan, M. Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model. *Intell. Autom. Soft Comput.* **2023**, *36*, 313–326. [CrossRef]
14. Blake, R.; Frajtova Michalikova, K. Deep Learning-Based Sensing Technologies, Artificial Intelligence-Based Decision-Making Algorithms, and Big Geospatial Data Analytics in Cognitive Internet of Things. *Anal. Metaphys.* **2021**, *20*, 159–173. [CrossRef]
15. Watson, R. The Virtual Economy of the Metaverse: Computer Vision and Deep Learning Algorithms, Customer Engagement Tools, and Behavioral Predictive Analytics. *Linguist. Philos. Investig.* **2022**, *21*, 41–56. [CrossRef]
16. Zauskova, A.; Miklencicova, R.; Popescu, G.H.; Abualhaj, M.M. Visual Imagery and Geospatial Mapping Tools, Virtual Simulation Algorithms, and Deep Learning-Based Sensing Technologies in the Metaverse Interactive Environment. *Rev. Contemp. Philos.* **2022**, *21*, 122–137. [CrossRef]
17. Cai, J.J.; Tang, J.; Chen, Q.G.; Hu, Y.; Wang, X.B.; Huang, S.J. Multi-View Active Learning for Video Recommendation. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, China, 10–16 August 2019; pp. 2053–2059.
18. Qin, Z.T.; Zhang, M.J. Towards a Personalized Movie Recommendation System: A Deep Learning Approach. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Information Systems (ICAIIS'21), Montreal, QC, Canada, 19–27 August 2021; pp. 1–5. [CrossRef]
19. Siersdorfer, S.; San Pedro, J.; Sanderson, M. Automatic Video Tagging Using Content Redundancy. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston MA, USA, 19–23 July 2009; pp. 395–402.
20. Fang, W.; Sha, Y.; Qi, M.; Sheng, V.S. Movie Recommendation Algorithm Based on Ensemble Learning. *Intell. Autom. Soft Comput.* **2022**, *34*, 609–622. [CrossRef]
21. Khan, U.A.; Martinez-Del-Amor, M.A.; Altowaijri, S.M.; Ahmed, A.; Rahman, A.U.; Sama, N.U.; Haseeb, K.; Islam, N. Movie Tags Prediction and Segmentation Using Deep Learning. *IEEE Access.* **2020**, *8*, 6071–6086. [CrossRef]
22. Zhang, F.; Lee, V.E.; Jin, R.; Garg, S.; Choo, K.R.; Maasberg, M.; Dong, L.; Cheng, C. Privacy-Aware Smart City: A Case Study in Collaborative Filtering Recommender Systems. *J. Parallel Distrib. Comput.* **2019**, *127*, 145–159. [CrossRef]

23. Sridhar, S.; Dhanasekaran, D.; Charlyn Pushpa Latha, G. Content-Based Movie Recommendation System Using Mbo with Dbn. *Intell. Autom. Soft Comput.* **2023**, *35*, 3241–3257. [CrossRef]
24. Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z. Collaborative Filtering and Deep Learning Based Recommendation System for Cold Start Items. *Expert Syst. Appl.* **2017**, *69*, 29–39. [CrossRef]
25. Hussein, A.H.; Kharma, Q.M.; Taweel, F.M.; Abualhaj, M.M.; Shambour, Q.Y. A Hybrid Multi-Criteria Collaborative Filtering Model for Effective Personalized Recommendations. *Intell. Autom. Soft Comput.* **2022**, *31*, 661–675. [CrossRef]
26. Duan, R.; Jiang, C.; Jain, H.K. Combining Review-Based Collaborative Filtering and Matrix Factorization: A Solution to Rating's Sparsity Problem. *Decis. Support Syst.* **2022**, *156*, 113748. [CrossRef]
27. Da"U, A.; Salim, N. Recommendation System Based on Deep Learning Methods: A Systematic Review and New Directions. *Artif. Intell. Rev.* **2020**, *53*, 2709–2748. [CrossRef]
28. Zuo, M.; Dai, G. P-Lsgof: A Parallel Learning-Selection-Based Global Optimization Framework. *J. Intell. Fuzzy Syst.* **2020**, *39*, 7333–7361. [CrossRef]
29. Covington, P.; Adams, J.; Sargin, E. Deep Neural Networks for Youtube Recommendations. In Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education, New York, NY, USA, 11–13 July 2016; pp. 191–198.
30. Ramachandram, D.; Taylor, G.W. Deep Multimodal Learning: A Survey on Recent Advances and Trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [CrossRef]
31. Yang, B.; Mei, T.; Hua, X.; Yang, L.; Yang, S.; Li, M. *Online Video Recommendation Based on Multimodal Fusion and Relevance Feedback*; ACM: New York, NY, USA, 2007; pp. 73–80.
32. Fan, Y.; Wang, Y.; Yu, H.; Liu, B. Movie Recommendation Based on Visual Features of Trailers. In *Innovative Mobile and Internet Services in Ubiquitous Computing, Advances in Intelligent Systems and Computing 612*; Springer International Publishing: Manhattan, NY, USA, 2017; pp. 242–253. [CrossRef]
33. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural Language Processing (Almost) From Scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
34. Rassweiler Filho, R.J.; Wehrmann, J.; Barros, R.C. Leveraging Deep Visual Features for Content-Based Movie Recommender Systems. In Proceedings of the 2017 IEEE International Ultrasonics Symposium, Washington, DC, USA, 6–9 September 2017; pp. 604–611.
35. Roy, D.; Ding, C. Movie Recommendation Using Youtube Movie Trailer Data as the Side Information. In Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), The Hague, The Netherlands, 7–10 December 2020; pp. 275–279.