Review of Movie Recommender Systems Based on Deep Learning

ZiXi Yao[1,*]

[1]School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, Guangdong 511363, China

Abstract: With the development of the network, society has moved into the data era, and the amount of data is exploding, we need a tool to help users find corresponding data collections based on their interests, and recommender systems were born for this purpose. In the movie field, recommender systems suggest items that users may like, improving the efficiency of finding movies and optimizing the user experience thus driving the growth of the movie industry. Machine learning is a multi-disciplinary science that focuses on how to improve the performance of algorithms by continuously reorganizing existing knowledge structures in a way that mimics human learning. Deep learning is a research direction in the field of machine learning that has achieved results in many areas that far surpass previous related techniques. In order to better provide personalized services to users and improve the accuracy of the system's recommendations, it is necessary to integrate deep learning techniques into the recommender system to optimize the system's performance. In this paper, we review different approaches in deep learning based recommender systems.

1. Introduction

The 21st century is the era of data and information, a large amount of data brings many opportunities and the problem of information overload comes along with it. In the film industry, users are unable to find the movies they are interested in quickly and accurately in the presence of huge movie data.

Recommender system is an effective means to solve the information overload problem.

Based on the historical interaction data between users and movies, recommender systems can effectively capture hidden information such as users' personalized preferences and movie attributes based on these data, and filter and display to users a personalized list of movie interests from a large number of resources. Technically speaking, there are two main types of recommendation technologies, one based on collaborative filtering and the other on content[1]. The former relies on the influence relationship among users, while the latter mainly calculates the degree of matching from content feature attributes.

In recent times, deep learning has once again set off a wave of machine learning research and has become a hot spot for theoretical and applied research in the field of machine learning[2]. Compared with traditional shallow machine learning, deep learning has excellent ability to automatically extract abstract features and can learn more effective features from massive data. With continuous exploration and experimentation in the field of deep learning, deep learning has achieved better results than traditional machine learning in areas such as image recognition, speech recognition and natural language processing[3].

Recommender systems have been widely studied by academia and industry as an important means to alleviate information overload, and have achieved certain results. In essence, recommender systems extract the interest features of users from a collection of Ostensibly irrelevant data, and use the interest features to build user preference models. The application of deep learning to recommender systems is a new direction in the development of recommender systems in recent years. In fact, Spotify, Netflix, Google, Microsoft and other companies have conducted in-depth research on how to apply deep learning to recommender systems, and achieved good results in practical applications[4].

This paper reviews the research on recommendation systems and lists the deep learning methods that can be applied to movie recommendation systems.

2. Overview of traditional movie recommender system

Traditional recommendation algorithms are usually classified into four types: collaborative filtering recommendations, content-based recommendations, demographic-based recommendations, and hybrid recommendations[5].

2.1. Collaborative filtering based recommendation algorithm

Collaborative filtering is a very classic and popular information filtering technique, which finds similar users of a specified user in a user group, combining these

---

* Corresponding author: 1025312744@qq.com

similar users' evaluations of a a certain information forms the system's prediction of that given user's preference for this information.

There are many branches of collaborative filtering, which in general can be divided into user-based and item-based.

2.1.1. User-based collaborative filtering recommendation algorithm

In the user-based approach, we will find the user with a similar interest to the target user from the data set and recommend the other preferences of that user to the target user. For example, when recommending for Cary, both Cary and Johnny have watched the movie "Titanic" and "The Amazing Spiderman", while Cary and Mike do not have the same viewing records, so obviously Cary and Johnny have more similar viewing preferences, so it will be possible to recommend the movie "Big Fish", which Johnny has watched, to Cary.

2.1.2. Item-based collaborative filtering recommendation algorithm

This method recommends similar items by finding items that are related to the target user. Suppose the user likes two movies "Fast and Furious 8" and "Pearl Harbor", then the recommendation system will find out the three movies that are most similar to these two movies respectively, calculate the user's interest level to each movie according to the formula, and select the movie with the greatest interest level to recommend to the user.

2.2. Recommendation algorithms based on content

The principle of content-based recommendation algorithm is to find products with similar reviews, tags, categories and other relevant information to the products purchased by the target user and recommend them to the target user. This algorithm is developed on the basis of information retrieval and information filtering, and its principle is simple and interpretable. Since only item characteristics are needed to calculate similarity using content-based recommendation algorithms, there are no cold-start and new item recommendation problems. For example, when recommending a movie to user A, user A likes movie A. By comparing the genres and lead actors of movie A, movie B and movie C, the system finds that both movie A and movie C belong to the "comedy" genre and both star Jim Carrey, so movie A is considered more similar to movie C, and movie C can be recommended to user A.

2.3. Demographics-based recommendation algorithm

The recommendation algorithm based on demographics classifies users into several different categories based on their basic information, such as age, occupation, race, gender, home address, etc. Users belonging to the same category we consider to be similar, and users belonging to multiple same categories are more similar. For example, Lisa and Ketty are similar in age, same gender and both are students, so it is considered that Lisa and Ketty have high similarity in interests, and it can be assumed that Ketty is likely to like Lisa's favorite movie A, so it can be recommended to Ketty. the recommendation algorithm based on demographics only needs basic information of users, so it can be applied to recommendation systems in various fields.

2.4. Hybrid recommendation algorithms

All of the above algorithms have limitations. With the continuous development of the movie industry, the number and types of movies are accumulating, and users' tasks and demands for personalized movie recommendations are becoming more and more diverse, and the use of a single recommendation algorithm can no longer meet the diverse needs of people. The hybrid recommendation algorithm is born to solve the limitations of a single algorithm[6], and its essence is to integrate the advantages of different recommendation algorithms and compensate their shortcomings, so as to obtain the optimal recommendation[7][8].

3. Recommended approach based on deep learning

3.1. Restricted Boltzmann Machine (RBM)

Boltzmann machines (BM) are neural networks that store fully connected neurons, which have excellent learning ability but also lead to slow training speed. In addition, there are problems such as difficulty in accurately computing the distribution represented by the BM and difficulty in obtaining random samples. The Restricted Boltzmann Machine was created to address these shortcomings, and the RBM contains a visible layer, a hidden layer, and unconnected neurons that are different from the BM.

In 2006, Hinton et al. worked out a framework that can execute multiple RBM.

The trained model can be optimized according to the tuning of traditional learning algorithms. This learning algorithm is first trained in layers of RBM to optimize the parameters of the model, and then operated according to the traditional learning method. This saves training time and reduces computational effort.

The Restricted Boltzmann Machine is the first deep learning method used in recommender systems. It predicts missing scores by automatically extracting abstract features[9].

3.2. Convolutional Neural Network (CNN)

Convolutional neural network is one of the representative algorithms of deep learning, which contains feedforward neural network with convolutional computation and deep knots. Due to its representational learning capabilities and their ability to classify input information in a

translation invariant manner based on their hierarchical structure, convolutional neural networks are also called "translation invariant artificial neural networks". In the 1980s and 1990s, Yann LeCun first solved the defects in handwritten numbers with CNNs[9]. Since then, deep learning has progressed and convolutional neural networks have been used in a growing number of fields.

The functions of each of the five layers in a convolutional neural network are:

(1) Data input layer: remove the mean of the original image data, normalization, PCA dimensionality reduction.

(2) Convolutional layer: Study the feature information of the input data.

(3) Stimulus layer: Mapping the output of the convolutional layer using a nonlinear approach.

(4) Pooling layer: reduces data as well as parameters.

(5) Fully connected layer: Combining the obtained features by a nonlinear form.

3.3. Graph Neural Network (GNN)

Graph neural networks borrow ideas from convolutional neural networks, recurrent networks, and deep self-encoders to both use graphs to reflect complex relationships between objects and to extract and represent features on data generated in non-Euclidean spaces. It shows more powerful capabilities in learning graph-structured data. Because of the superior performance of GNN in graph learning, it is widely used in recommendation systems. Besides, it is also applied to many fields, such as text classification, natural language processing, disease prediction, and feature relationship extraction.

For recommendation systems, the general recommendation system includes three stages, which are matching, ranking and re-ranking. The principle of graph neural network is mainly: firstly, a graph structure model is constructed to reflect the relationship between entities and entities, and then a specific method is used to describe the nodes and get the nodes containing the final neighbor node information and topological structure feature state, which has been updated continuously for iteration, and finally the representation of these nodes will be output by some specific way and the required information will be obtained[10][11]. As for GNN recommendation systems, The main stages are similar to the graph neural network process, which is generally divided into only four steps:

(1) Building GNN models based on entity interrelationships.

(2) Deciding the information propagation and updating method of GNN model.

(3) Extracting the updated node features.

(4) Selection algorithm to implement recommendations

The approaches based on graph neural network recommendation systems are divided into five main categories:

3.3.1. Graph Convolutional Network Recommendation System (GCN)

The difference between a graph convolutional network and a traditional convolutional operation is the change from traditional data to graph data, it learns node embeddings of high quality graphs through a sampling technique and a graph convolutional operation, and is mainly applied to the recommendation problem of large-scale heterogeneous graph data, but only for two entities, i.e., users and items.

3.3.2. Graph Attention Network Recommendation System (GAT)

A graph attention network is a spatially based graph convolutional network that uses the attention mechanism to determine node neighborhood weights when aggregating feature information. It is able to amplify the impact of the most important parts of the data and adaptively learn neighboring node weights. Its drawback is that the computational cost and memory consumption increases rapidly with the computation of attention weights between each pair of neighbors.

3.3.3. Graph Autoencoder Recommendation System

Graph autoencoders are a class of graph embedding methods that are typically used to obtain node embeddings using multilayer perceptrons as encoders. The aim is to represent the vertices of the graph as low-dimensional vectors using a neural network structure. Kipf and Welling proposed the Variational Graph Autoencoder (VGAE) in 2016. Since then, graph self-encoders have been used in many fields, such as brain circuit diagrams, multi-views, social networks, etc.

3.3.4. Graph generation network recommendation system

The target effect of this method is to generate new graphs based on the observed set of graphs. The approaches to graph generation networks are based on specific domains, e.g., molecular graph generation, natural language processing, etc. Some approaches treat the generation process as alternating node and edge generation, while others use generative adversarial training. The drawback of this approach is the problem of new node connectivity and existence of cold start problem of the graph. For example, there is a large amount of missing information about user attributes, such as gender and education.

3.3.5. Graph Spatio-Temporal Network Recommendation System

The graph spatio-temporal network simultaneously captures the spatio-temporal correlation of the spatio-temporal graph, which is modeled using the spatio-temporal graph. The spatio-temporal graph has a global graph structure, and the input of each node

changes over time. The ideal effect of graph spatio-temporal networks is to be able to predict future node values or labels, or to predict spatio-temporal graph labels.

4. Conclusion

With the continuous maturation of technologies such as deep learning, data mining, and prediction algorithms, improving the accuracy, security, and privacy of recommendation systems will become a hot topic for future research. This paper first briefly outlines the traditional movie recommendation system, and then lists several deep learning-based recommendation methods and briefly explains them. These methods are well worth exploring in depth in the movie recommendation field, and are expected to be subsequently applied to movie recommendation systems or used to improve certain problems of current movie recommendation systems.

References

1. YU Meng, HE Wentao, ZHOU Xuchuan*, CUI Mengtian, WU Keqi, ZHOU Wenjie.Review of recommendation system[D]. Chengdu:Southwest University for Nationalities, 2022. -.

2. Erol Mrzic, Tarik Zaimovic.Data Science Methods and Machine Learning Algorithm Implementations for Customized Pratical Usage[D]. Alija Izetbegović 1, Sarajevo, Bosnia and Herzegovina:University of Sarajevo, School of Economics and Business, Department of Information Systems Development, 2020. -.

3. Yu Li, Du Qihan, Yue Boyan, Xiang Junyao, Xu Guanyu and Leng Youfang.Survey of Reinforcement Learning Based Recommender Systems[D]. Beijing:School of Information, Renmin University of China, 2021. -.

4. Wang Junshu, Zhang Guoming, Hu Bin.A Survey of Deep Learning Based Ｒecommendation Algorithms[D]. Nanjing, Jiangsu:Journal of Nanjing Normal University, 2018. -.

5. Zhang Zheng-feng, QIANG Cheng-kui, Duan Su-feng.Review of Personalized Movie Recommendation Algorithms[D]. :Xuzhou College of Biotechnology, 2021. -.

6. Robin Burke.Hybrid Web Recommender Systems[D]. 243 S. Wabash Ave. Chicago, Illinois, USA:School of Computer Science, Telecommunications and Information Systems DePaul University, 2007. -.

7. Saumya Bansal · Niyati Baliyan.Remembering past and predicting future: a hybrid recurrent neural network based recommender system[D]. :Springer-Verlag GmbH Germany, part of Springer Nature, 2022. -.

8. Xuanxuan Wang.Design and Implementation of Personalized Movie Recommendation System based on Hybrid Recommendation Algorithm[D]. :Huazhong University of Science and Technology, 2021. -.

9. Geng Yiwen.Research on Recommendation System Based on Deep Learning[D]. Beijing:Northern Polytechnic University, 2021. -.

10. WU Jing, XIE Hui, JIANG Huowen.Survey of Graph Neural Network in Recommendation System[D]. Nanchang:School of Mathematics and Computer Science, Jiangxi Science and Technology Normal University, 2022. -.

11. Chen Daoqiang.Design and Implementation of Movie Recommendation System Based on Graph Convolution Neural Network[D]. :University of Electronic Science and technology, 2022. -.