AI in Video Recommendation System

Eric Li

Beijing Luhe International Academic, China

Abstract. Short videos are very popular all over the world. Video recommendation system is an essential part in it. It can help people to watch the video that they are interested in. This paper is written for study the specific principle of the video recommendation system. The result was getting through relative literatures and actual test. Short video recommendation systems typically use collaborative filtering and deep learning techniques to achieve this. Collaborative filtering comes in two types: user-based and content-based. User-based collaborative filtering recommends videos to new users based on the viewing behavior of similar users. Content-based collaborative filtering uses video features and similarity to recommend similar videos. Finally, this paper shows how the video web set can learn what is the user's interest.

Keywords: Artificial Intelligence; Video Recommendation Systems; Neural Networks; Collaborative Filtering; Deep Learning; Datasets; Experimental Results; Research Directions.

1. Introduction

1.1 Overview of Video Recommendation Systems

Short video recommendation systems typically use collaborative filtering and deep learning techniques to achieve this. Collaborative filtering comes in two types: user-based and content-based. User-based collaborative filtering recommends videos to new users based on the viewing behavior of similar users. Content-based collaborative filtering uses video features and similarity to recommend similar videos.

Deep learning techniques are also widely used in short video recommendation systems. One common deep learning model is recurrent neural networks (RNNs), which are used to model a user's viewing history. Additionally, there are models based on graph neural networks (GNNs) and attention mechanisms, which can model a user's behavior and preferences.

In summary, the principle of short video recommendation systems is to model a user's viewing history, behavior, and preferences, and to use collaborative filtering and deep learning techniques to recommend video content that the user may be interested in.

1.2 Application of AI in Video Recommendation Systems

With the continuous improvement and upgrading of artificial intelligence technology, video recommendation systems are increasingly inclined to realize personalized recommendations, thus enhancing users' satisfaction and trust. In the recommendation system, artificial intelligence technology is widely used to analyze users' video preferences, mainly through deep learning algorithms, establish user behavior models, predict their video needs, and provide accurate recommendations. Furthermore, artificial intelligence technology will adjust the algorithms based on different user types, behavior patterns, and feedback to continuously optimize the recommendation effect. Therefore, based on the characteristics and advantages of applying artificial intelligence technology to video recommendation systems, we can foresee that it will continue to be a research hotspot and the main direction of application practice in the future.

2. Introduction of Datasets

2.1 Youtube 8M Dataset

The YouTube 8M dataset is a large-scale video dataset with labels for deep learning, released by Google in 2016. It includes millions of video segments, each about 2 seconds long, covering

thousands of different topic categories, such as movies, sports, music, animals, and games, among others. This dataset is mainly used for applications such as video classification, tagging, and search, and is important for academic research and industrial applications in the fields of video classification and content analysis.

The YouTube 8M dataset is a large-scale video dataset released by Google in 2016. It includes 8.5 million non-overlapping video segments selected from the YouTube platform and labelled to provide millions of effective video training samples for deep learning algorithms. Each video segment is identified by a unique video ID on YouTube, and each segment is roughly 2 seconds long, covering several thousand different subject categories. The dataset contains the largest number of public labels, over 4,800, covering a wide range of subjects and subcategories, such as movies, music, human activities, animals, and natural scenes.

Regarding video annotation, the dataset employs a combination of machine automatic labeling and human correction methods. Through deep learning models, videos are detected and analyzed, and each video segment is assigned a set of labels. Additionally, each label is assigned a confidence score to indicate how well it matches the video segment. The dataset also offers high-quality video thumbnails and video comments and other important metadata.[1]

The YouTube 8M dataset has wide-ranging applications, including video classification, video tagging, video search, video recommendation, and advertising. Its application fields encompass deep learning, machine learning, and natural language processing, among many others, providing precise and reliable experimental data for experts and researchers in related fields.

2.2 Netflix Prize Dataset

The Netflix Prize dataset was released by Netflix in 2006, aiming to improve its movie recommendation system through machine learning algorithms. The dataset consists of over 6,000 movie rating records from Netflix users from 1999 to 2005, covering over 18,000 users and over 4,000,000 ratings. The dataset is hosted on cloud network drives for academic and industry experts to download and analyze.

Netflix offered a $1 million prize to motivate researchers and data scientists to analyze the data and develop more accurate recommendation algorithms. The winners must establish algorithms to beat Netflix's previous Cinematch recommendation algorithm based on an RMSE of less than 0.9525 on Netflix's evaluation dataset. The Netflix Prize dataset has been one of the popular datasets used by researchers and academic institutions in the fields of machine learning, data science, and recommendation systems. It has had a significant impact on the development of recommendation algorithms and academic research.

The Netflix Prize dataset is a large-scale and high-quality practical dataset that provides massive user and movie rating data with strict data quality management and confidentiality requirements. The dataset includes multiple structured data, among which the most critical one is the movie rating data, and each rating is identified by a unique user ID and movie ID. Additionally, the dataset also contains rich movie metadata, such as movie title, actor lists, directors, movie genre, release date, etc.

To guide algorithm evaluation, the Netflix evaluation metric is based on root mean square error (RMSE). Participants need to use Netflix's provided training data to train their models and validate their results using the test set. All participants' models are ranked under the RMSE evaluation metric, with the aim of reducing the RMSE value as much as possible and improving prediction accuracy.

Due to its massive and sparse data characteristics, the data processing and modeling requirements of the dataset are challenging. For data scientists and machine learning practitioners, the dataset provides a good foundation and resource for recommendation algorithm and personalized recommendation research, providing strong support for academic research and practical applications.

2.3 TikTok Dataset

TikTok's recommendation system is fueled by a diverse range of datasets, each of which plays a crucial role in capturing and deciphering users' preferences and behaviors. The main datasets consist of user interaction data, content data, device data, and session data.

The user interaction dataset primarily captures user engagement through likes, comments, shares, and other forms of interaction. The algorithm uses this dataset to gauge a user's interests and make recommendations based on their viewing habits.

The content dataset contains critical video metadata, including the video's category, language, duration, and other attributes that can provide insight into a user's preferences.

The device dataset captures device specifications such as the operating system, device type, and screen size. This data is essential in optimizing recommendations for specific device settings to enhance the user experience.

Lastly, the session dataset tracks user activities, including the duration of each session, the number of consecutive sessions, and other parameters that help the system to create a personalized experience.

In conclusion, TikTok's recommendation system utilizes a complex and diverse range of datasets, reflecting the app's focus on delivering tailor-made video content to each individual user. By analyzing these datasets and deciphering patterns in user behaviors and preferences, the algorithm continually optimizes its recommendations, providing a personalized and immersive experience for each user.

3. Algorithms and Methods

3.1 Neural Collaborative Filtering

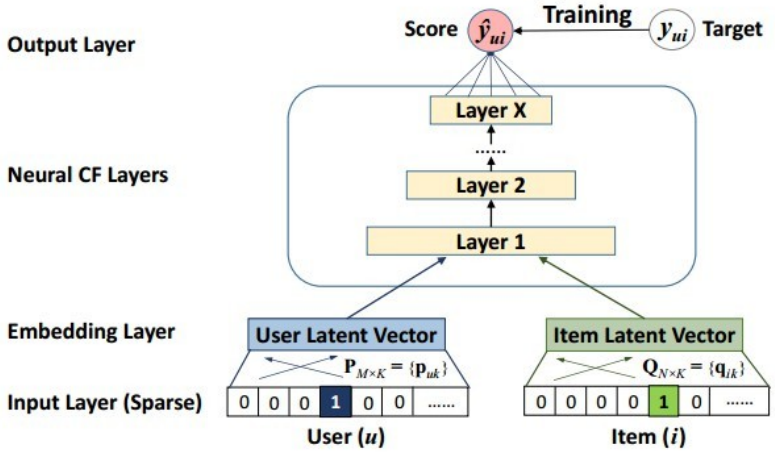


Fig 1. NCF model

Neural Collaborative Filtering (NCF) is a recommendation system models based on neural networks that can predict user ratings or recommend related items. In traditional collaborative filtering methods, shallow models such as matrix factorization are often used to predict on the interaction matrices between users and items. NCF combines the ideas of deep learning to construct neural network models that improve prediction accuracy.

In NCF, user and item information are represented as vectors. These vectors can be generated by embedding layers, which can convert sparse user-item interaction matrices into dense vector representations. The embedding layer maps each user and item to a low dimensional vector representation, and the dimension of these vectors can be adjusted as hyperparameters.

Next, these vectors are used as inputs to a multi-layer neural network, which consists of several fully connected layers and activation functions. The neural network learns to predict the degree of

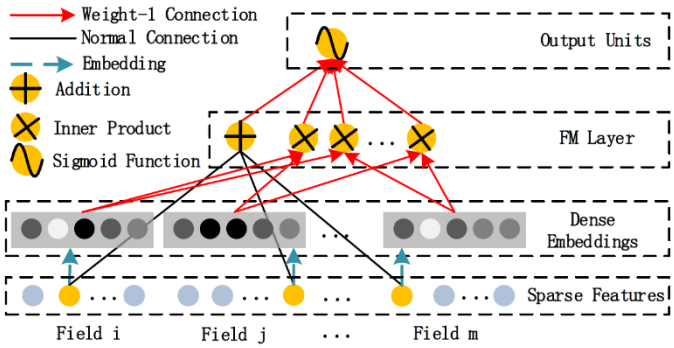
interaction between users and items, and finally outputs a predicted rating or a probability distribution of recommended items.

NCF model can be trained using different loss functions. The most common loss functions are cross-entropy loss and mean squared error loss, which are used to measure the difference between model output and actual values. When the model error decreases, the model can better predict the user's evaluation of different items and provide accurate recommendations for users.[2]

In summary, NCF is a recommendation system model that combines embedding layers and neural networks to achieve more accurate recommendations and predictions, suitable for large-scale recommendation tasks.

3.2 DeepFM

DeepFM (Deep Factorization Machine) is a hybrid model that combines factorization machine (FM) and deep neural network (DNN) and is used for classification or regression tasks. Unlike traditional FM, which uses only linear models for collaborative filtering, DeepFM uses DNN to perform nonlinear processing on features, thereby improving the model's prediction performance.



The architecture of FM.

Fig 2. The architecture of FM

The DeepFM model consists of three parts: the input layer, factorization machine (FM), and deep neural network. The input layer is used to process the input features, while FM and DNN are used to capture feature interactions and nonlinear features, respectively. The entire model consists of two parts: the factorization machine model, which maps the input feature vectors to a low dimensional vector space and then calculates feature interactions, and the DNN model, which attempts to learn more complex nonlinear feature representations from cross-feature mappings.

In the FM part, the model uses quadratic terms to model feature interaction, including pairwise combinations and bias terms. In the DNN part, the model uses one or more hidden layers to perform nonlinear feature transformations and feature interaction modeling. Like traditional deep neural networks, except for the output layer, each hidden layer of DeepFM uses activation functions such as ReLU and sigmoid.[3]

The advantage of DeepFM is that the combination of factorization machine and deep neural network enables it to capture sparse features and nonlinear feature interactions simultaneously, resulting in higher prediction performance. In addition, compared to DNN, FM has high training efficiency and is suitable for sparse data. It can be trained using methods such as stochastic gradient descent. DeepFM has been widely used in areas such as CTR prediction, click-through rate prediction, and recommendation systems, and has achieved good results.

In conclusion, DeepFM is a hybrid model that combines factorization machine and deep neural network. It is suitable for processing large-scale, sparse, and nonlinear data and has the advantages of efficiency and high prediction performance.

3.3 Deep Learning for YouTube Recommendations

Deep Learning for YouTube Recommendations (DL4YR) is a deep learning model used for video recommendations. This model was released by Google in 2016, and is therefore also known as the YouTube Deep Learning Recommendation System.

The goal of the DL4YR model is to recommend to users the videos they are most interested in on YouTube. In order to achieve this goal, the model needs to process a large amount of viewing history data and other user behavior data, including search history, comments, shares, likes, etc.

The DL4YR model consists of two main components: a candidate generation network and a ranking network. The candidate generation network uses deep learning techniques to generate a list of possible candidate videos in a large-scale video library. The ranking network sorts the list of candidate videos according to the user's level of interest and recommends them to the user. The input to the entire model is a user's video viewing history and other user behavior data, and the output is a personalized set of recommended videos.

DL4YR uses multiple deep learning techniques to generate and rank video candidates. For generating the list of candidate videos, convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) are used to process video embeddings, and negative sampling techniques are used to generate a batch of candidates for each user. For ranking, representation learning techniques and multi-layer perceptrons (MLPs) are used to rank the list of candidate videos. These techniques are able to learn video features and user interests from large-scale training data.[4]

The advantages of the DL4YR model are that it can tightly integrate user interests and video features to generate the most attractive recommended video list. In addition, the model is trained based on the interaction history data of users and videos, so it can continuously update the model parameters and recommendation results as data increases and changes.

4. Conclusion and Future Work

Artificial intelligence plays an important role in video recommendation systems, mainly in the following areas:

Data analysis

Video recommendation systems require massive user data for analysis and processing. AI can help the system accurately extract user preferences, analyze and mine user behaviors, and recommend video materials that better fit the user's tastes.

Content recognition

Through machine learning and visual recognition technology, AI can identify the features of video materials, such as scene, sound effects, color, style, etc. This helps the system accurately recommend relevant video materials to users, improving the accuracy and quality of recommendations.

Algorithm optimization

AI also has unique advantages in algorithm optimization. Through iterative learning and optimization of big data, AI can help video recommendation systems continually improve recommendation accuracy and make differentiated recommendations for different users or different scenarios, thereby enhancing user experience.

Personalized recommendation

In traditional video recommendation systems, the recommendation algorithm mainly relies on single indicators such as click-through rate and browsing length. AI can understand and analyze users' personalized needs more comprehensively and make more refined recommendations to meet different users' personalized needs.

In summary, AI plays an important role in video recommendation systems, providing users with more suitable video recommendation services, thus improving user experience and retention rate on video platforms.

References

[1] Wu, Liwei, et al. “Deep Neural Networks for YouTube Recommendations.” Proceedings of the 2016 ACM Conference on Multimedia Conference, ACM, 2016, pp. 165-174.

[2] He, Xiangnan, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural Collaborative Filtering." In Proceedings of the 26th International Conference on World Wide Web, pp. 173-182, 2017.

[3] Guo, H., Tang, R., Ye, Y. et al. (2017). DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining pp. 627-635.

[4] Lee, Joonseok, et al. "Deep Neural Networks for YouTube Recommendations." ACM Transactions on the Web, vol. 12, no. 1, 2018, pp. 1-22.