Application of Machine Learning in Recommendation System

Application of Machine Learning in Recommendation System

Shan Gao*

School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

shgao18@lzu.edu.cn

Abstract. China's science and technology has been basically maintained in a state of rapid development, among which machine learning has been widely used in many fields of society. For the economic and financial fields, if they want to make effective use of machine learning, the first thing to do is to comprehensively analyze the advantages and disadvantages of each algorithm, and then use machine learning to build the corresponding user recommendation system, so as to understand the actual adaptability of each algorithm. The following sections will analyze the application of machine learning in recommender systems.

1. Introduction

Now China has basically ushered in the information age. Although most of the data could be stored efficiently in this context, more data would be generated during the use process, which made it more difficult to use. Therefore, based on this, relevant personnel need to carry out a more in-depth mining of the data, while studying its use in various fields. One of the more widely used is machine learning technology. In addition, in recent years, in the information filtering and information retrieval, recommendation system occupies a more important position. Although the recommendation system has brought more economic benefits to the society, there are still a series of problems in the process of using it. For example, the accuracy of recommendation is low, information overload, etc. The effective way to solve these problems is to use machine learning in the recommendation system.

2. The process of machine learning

2.1. Data collection

In the actual operation of machine learning, it will first use some effective ways to achieve data acquisition. And this can show that machine learning is based on data, and in the process of its learning operation, the relatively least difficult step is to obtain data. Since the channels for generating data are relatively wide, the amount of data generated is also relatively high. The most obvious ones are the medical records of major hospitals and supermarket purchase records, etc., and the basis of machine learning is these data information[1].

2.2. Data cleaning

For most data, its biggest characteristic is strong irregularity. So there is a certain degree of irregularity when people are trying to get data. In this context, most of the data harvested in the end have no high utilization value, and even some of the data will affect the system operation to a certain extent. Therefore, the relevant personnel in the process of analyzing data, the first thing to do is to screen useless information, and carry out a comprehensive clean-up. In this way, the accuracy of the analysis

results can be strongly guaranteed. In addition, there are three different aspects that stakeholders need to pay attention to when processing data: distance measurement, sampling, and dimensionality reduction. For collaborative filtering recommendation system, it will use KNN classification. In most cases, this is determined by the distance metric. Among them, the measures that are put into practical use with relatively high probability are as follows, namely Pearson correlation coefficient and Euclidean distance. In addition, for sampling technology, its main function is to realize in-depth data mining, and several kinds of big data to effectively select the subset of relevant data. At the same time, the existence of this technology also plays a more important role in the interpretation step. Now the most common sampling method is the no-substitution sampling method. If the item is identified by the staff, it will be removed. But in this process, the staff can also use the substitution sampling method, the existence of this method can make the item after the selection is completed, do not have to remove it in the whole. In most cases, the purpose of dimensionality reduction is to remove some points with relatively low density and which will affect the results. After the removal is completed, the dimension is effectively reduced. In this way, dimensional disasters can be avoided.

2.3. Model construction and data analysis

For the entire process of machine learning, the core content is data analysis. In the process of analyzing data, there are many methods that can be used. Due to the different methods, there will be a certain degree of difference in the effect of the actual use process. This requires the staff to analyze the actual situation of machine learning and choose a more appropriate data analysis method before use. When using machine learning in a recommendation system, you can mainly use the following methods: The first is the nearest neighbor algorithm. For this method, the main principle is to predict the label category of the sample based on the stored thought record[2]. For this classifier, its main storage object is the training set. In this process, if it wants to ensure the effectiveness of the classification, what must be done is to ensure an effective match between the training set and the new record. Compared with other methods, this method is relatively simple in use. Because when using the nearest neighbor algorithm, the staff does not need to build a relevant model, and it also has a high degree of accuracy. The second is artificial neural network. For the artificial neural network, its algorithm simulates the biological network to a certain extent, and it is mainly composed of two parts, namely the internal connection point and the weighted chain. In most cases, artificial neural networks exist to solve regression and classification problems. In addition, in the process of machine learning, a relatively important branch is artificial neural network. The main feature of artificial neural network is that it can deal with the nonlinear classification task effectively, and it can realize parallel processing on this basis. Even if the real network damage environment, to also can maintain real-time operation. But in practice, artificial neural network system is difficult to provide a more ideal network topology for some problems.

3. Recommendation system

3.1. System architecture

Currently, there are three main modules that make up the recommendation system, namely the retrieval of related items, the ranking of items, and the extraction of user characteristics. For item retrieval, it is mainly based on the characteristics of each user, in a relatively short period of time, searching for items that users are interested in. In most cases, the module can mainly use the form of information retrieval to complete the retrieval of items, so that users can quickly find the items they need. For the sorting module, it is mainly based on the algorithm of machine learning, and then complete the work of improving the implementation of the indicators. In this way, the corresponding model can be derived to a certain extent. For example, it can conduct a comprehensive analysis of user characteristics and the characteristics of candidate items, and then budget the user's click-through rate of items. After the budget is completed, the module can combine the click-through rate of each item to effectively sort it. At the same time, in the actual sorting process, not only the novelty of the items

must be paid attention to, but the diversity of the results must also be considered. User body feature extraction refers to recording the user's daily behavior, and then analyzing the user characteristics based on the effective information. According to the extraction of user body characteristics, user interests can be accurately described.

3.2. User characteristic acquisition and clustering algorithm

Taking a book website as an example, each user can choose his own website to log in based on his own needs. In this process, the recommendation system can complete the extraction of user characteristics in a relatively short time. After the user characteristics are extracted, the system will recommend books that the user may be interested in. Among them, the characteristics of users can be mainly expressed through the following aspects. First, the identity that the user logs in during the process of purchasing books. Second, the user's own gender and age. Third, after the user has logged in to the relevant website, he has browsed some pages later. Fourth, the vocabulary that users have used and input on the website. Fifth, users collect and comment on some books on the website[3]. If the gender and age dimensions involved in the user characteristics are relatively low, while other characteristics are relatively high, then a series of problems will arise. One of the most prominent is the problem of sparsity. The existence of this problem will greatly increase the computational load of the retrieval module itself, and some functions of the system itself will also be reduced. During the operation of the retrieval module, the staff must ensure its operating efficiency, which can be achieved through corresponding dimensionality reduction. In this process, the relatively ideal dimensionality reduction effect is to effectively cluster users, such as comprehensively collecting user history records, and then classifying them clearly.

3.3. Retrieve items

After the system captures the characteristics of the items that the user is interested in, the recommendation system will recommend a series of items that the user may like based on this. For the recommendation system, it is mainly based on the indexing technology to realize the retrieval of items. The main function of this module is to use the inverted index for subsequent retrieval. The "item-item" index occupies the most important position in the recommendation system. In addition, when the recommender system explores the problem, the main calculation object is the degree of association between items. At the same time, in most cases, the item recommendation system is based on the user's appreciation of the item, and provides him with other items that he may be interested in. Currently, there are two main algorithms for calculating the degree of relevance of mainstream items, including content-based algorithms and behavioural algorithms. In addition, for collaborative filtering algorithms, it is mainly composed of behavioural item relevance algorithms, which are closely related to the relevance criteria of the machine learning industry.

4. Application challenges faced by machine learning in recommendation systems

For the recommendation system, its main role is to solve large-scale data, while recommending a large amount of information in a more accurate state. In the actual use process, not every item needs to be recommended. Only when the number of "subjects" is sufficient and the user does not select the "subjects" he browses, the recommendation system It is necessary to put it into actual use. For example, for Apple's official website, there are only dozens of types of mobile phones on sale, and the number of each type is also limited, so there is no need to use the recommendation system. After users have browsed all mobile phones on sale on Apple's official website, they can purchase products according to their own preferences. Although the problem that the recommendation system can solve is relatively simple on the surface, the construction of an efficient recommendation system is quite difficult. For the construction of recommendation system, it belongs to a relatively large system engineering. In the process of constructing this system, relevant enterprises need to invest a lot of human resources and material resources[4].

For example, “Toutiao”, its internal leaders have a more in-depth understanding of the value contained in the recommendation algorithm, so in the initial stage of the establishment of the company, the algorithm was regarded as the main core, and the ideal was achieved by building the system. product experience. In this context, Toutiao stood out among news clients within a few years and brought high economic benefits to the founder. However, both the construction of the recommendation system and its implementation are challenging to a certain extent. The main challenges faced by the application of machine learning in the recommendation system are as follows.

4.1. Data sparsity problem

In real application scenarios, the interaction information between users and projects is relatively sparse. In this context, not only user information and project information will be missing, there will even be some incorrect information. Therefore, in the process of constructing the recommended algorithm model, the technical staff must pay attention to this aspect.

4.2. Handling of unstructured in information

In the current process of constructing a recommendation system, technical personnel are more concerned about how to make full use of unstructured information related to users and projects, such as videos and pictures. So that more information can be input into the recommendation model, so that the recommendation effect can be improved. With the continuous development of deep learning technology, it has been used on a large scale in recommendation systems. The existence of this technology not only improves the data processing capacity, but also solves the problem to a certain extent[5].

4.3. The accuracy of recommendation results

At present, after using machine learning in the recommendation system, although the accuracy of its score prediction will be improved, at the same time the recommendation success rate will be reduced. For recommendation systems. It requires the rabbit’s data source to have sufficient user project interaction data, and a series of other available information, so that more accurate recommendation model parameters can be learned during use and better recommendation services can be developed.

4.4. Cold start problem

For the recommendation system, there will be data update problems during its use, and a large number of new users will join it. These new users do not have historical information, and this requires the recommendation system to have targeted solutions to recommend some new items to users so that users can get a better experience.

4.5. User interface issues

The recommendation system is mainly realized through the interaction between users and items in the process of providing recommendation services to users. In actual use, an excellent UI interface and user interaction experience will bring greater appeal to users. Therefore, technical personnel should also pay attention to the design of the UI, so that the user interaction experience can be improved, and at the same time, the recommendation system can play its own greater value[6].

4.6. Large-scale computing and storage problems

In the actual use process, the system's data processing and computing capacity will face high pressure due to the high number of users and projects. Therefore, in the recommendation system, technicians need to make use of the distributed technology to do a good job of data storage, transmission and calculation. Only after technicians can build a big data analysis and processing platform can machine learning be effectively used in the recommendation system[7].

5. Conclusion

In general, for machine learning, it involves a lot of algorithms, and when it is applied in the recommendation system, it can achieve excellent application effect. When designing the recommendation system, technicians should pay attention to the transformation of machine learning problem ability based on product requirements. At the same time, the technicians should select the appropriate algorithm according to the actual situation to ensure the operation effect of the system.

References

[1] Liu Zhongbao. Application of machine learning method in personalized recommendation system [J]. Intelligence Exploration, 2016(4):80-82(in Chinese).
[2] Zhou Qi. Recommendation System Based on Machine Learning[J]. Electronic Technology and Software Engineering, 2016(24):173(in Chinese).
[3] Qi Shuaibin, Hu Chenjun, Hu Kongfa, et al. Research and thinking on recommendation system based on machine learning and TCM theory[J]. Jiangsu Science and Technology Information, 2017(10):37-38(in Chinese).
[4] Lu Jun, Zhang Tianfan. Research on Recommendation System Based on Implicit User Behavior[J]. Journal of Hubei Institute of Technology, 2016 (3):22-27(in Chinese).
[5] Wang Lei. Intelligent optimization system design of LTE network based on machine learning technology[J]. Telecommunication Engineering Technology and Standardization, 2018 (1):39-42(in Chinese).
[6] Qiao Yu, Li Lingjuan. Collaborative filtering algorithm combining user similarity and rating information[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science,2017(3):100-105.
[7] Xiao Hong. Discussion on Machine Learning Algorithm Based on Big Data[J]. Communications World, 2017(6):265-266(in Chinese).