

Design and Implementation of Recommendation System of Micro Video's Topic

Dongdong Jiang
School of Computer Science
Communication University of China
Beijing, China
jdd1004196157@163.com

Wenqian Shang
School of Computer Science
Communication University of China
Beijing, China
shangwenqian@163.com

Abstract—With the development of Internet technology and the arrival of the era of big data, it is necessary to analyze and excavate the micro video data. It can help micro video creators to create better to analysis micro video data. This paper mainly introduces the structure design, key technical points and specific implementation steps of the micro video topic recommendation system.

Keywords—*recommendation; micro video*

I. INTRODUCTION

In the internet era, more and more traditional industries have been affected by the Internet, of course, the video industry is no exception. For the evaluation of a micro video, the most direct criterion is the number of viewing. For example, the movie box office, TV ratings, and so on. Therefore, increasing the number of video viewing is a problem that video creators need to pay attention to from beginning to end. With the arrival of the era of big data, video production needs to pay more attention to the user's needs, so as to achieve the rapid increase in the number of video viewing.

There are many successful practices of the application of big data technology in the video industry. In the early stages of the production of the video, there are a lot of data mining and data analysis. First, obtain massive video data from the internet. Then, through the analysis of these data, obtain many video information that the user groups are interested in, like themes, actors, songs. In the end, it is possible for video producers to create videos that can get a large amount of video viewing times according to the information obtained from the analysis.

This system gets a large amount of video data from large social platforms, video sites, search engines and other ways. Then, combines massive video information and business requirements to provide more valuable creative information for companies' video creation, so as to make the video more in line with the needs of users, and enhance the number of video viewing.

II. THE DESIGN OF SYSTEM STRUCTURE

With the development of the new generation of information network technology represented by mobile Internet, big data, cloud storage, the traditional advertising micro video creation has a fundamental change. This paper studies the design of the IT architecture and the division of the system service level in

the era of big data[1], and designs the structure of the system, which is shown in Fig.1:

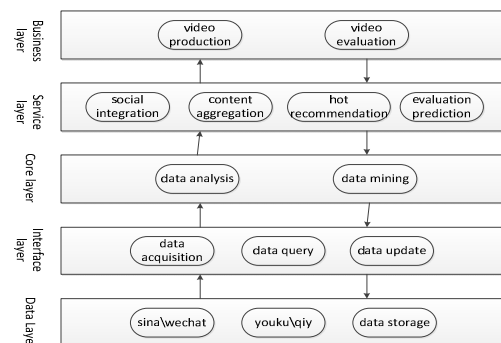


Fig. 1. the structure of system

In Fig 1, the data layer provides the basic data for the system, and stores the data. Data sources are mainly divided into three categories: data from social platforms, data from the video site, data from the search engine.

The Interface layer mainly provides the corresponding data acquisition, data import, data processing methods according to different data sources and data formats. The interface layer also provides data query and data update interface to the system through standard data transmission interface.

The core layer supports the core modules of the whole system: data analysis and mining. Data mining is the key technology of micro video topic recommendation.

The service layer is based on micro video content, providing the following services:

- Hot recommendation. The service can extract topics which are currently popular from the vast amount of information in the Internet. In the process of micro video creation, the creator can choose the most suitable hot words according to their own needs.
- Social integration. The service is a combination of many micro video publishing platform. For example, WeChat, micro-blog, YouKu and other micro video publishing platform. Through the integration of these publishing platform, we can realize the synchronization and update of information in multi-platform.

- Content aggregation. The service aggregates a number of similar content in the Internet, so as to provide users with more reference information.
- Evaluation and prediction. The service can evaluate the market performance of micro video. Based on the analysis of the characteristics and the market effect of the micro video in the past, this paper realizes the evaluation and prediction of the unpublished micro video.

The business layer includes micro video creation and evaluation. Micro-video authoring platform is for micro-video production team. The platform can select suitable themes according to the different objects of the user. Micro-video producers only need to create according to the relevant information, and can achieve a better market effect. Micro-video evaluation platform is developed for the company's decision makers. With the help of the evaluation system, the decision maker can understand the market impact of the release of micro video more timely, so as to guide decision-makers to make decisions.

III. RESEARCH ON THE KEY TECHNOLOGY

The core of the system is the data processing and analysis, so the data mining[2] technology is particularly important, is the technical support of the whole system.

A. Research on Clustering Algorithm

Content aggregation uses clustering algorithms[3]. Clustering is the process by which a physical or abstract set is divided into several classes of similar objects. Content aggregation is the clustering of data which is from different sources and different structures, and is similar. By clustering, we can get a unified description of micro video. This kind of external description is based on clustering algorithm, and it has a high degree of generality and wider coverage.

At present, the clustering algorithm can be divided into five categories: hierarchical clustering algorithm, partition clustering algorithm, clustering algorithm based on constraint, clustering algorithm in machine learning and clustering algorithm for high dimensional data. The performance of clustering algorithm is different in different application fields, that is to say, there are few algorithms which can be applied to several different application backgrounds.

Each algorithm has its own advantages and disadvantages, which decide the application scenario of the algorithm. For example, hierarchical clustering method can control the granularity of different levels flexibly, so it has strong clustering ability, but it greatly extends the execution time of the algorithm. The convergence speed of partition clustering algorithm is fast, and it can be extended to large scale data sets. But, the selection of the initial clustering centers and noise data have a greater impact on the clustering results.

B. Research on Regression Analysis Algorithm

Evaluation and prediction uses regression analysis[4]. Regression analysis is a statistical method for determining the quantitative relationship between two or more than two variables. The system can integrate a large amount of data

information and use regression analysis to get the logical relationship between user evaluation and micro video attributes. When certain properties of the micro video are determined, the future market trend of the micro video can be predicted by the logical relationship.

Regression analysis can be divided into simple regression analysis and multiple regression analysis according to the number of variables, and also can be divided into linear regression and nonlinear regression analysis according to the function expressions of dependent and independent variables.

IV. IMPLEMENTATION OF THE SYSTEM

The system uses the Python language to achieve the background, the front page is achieved by HTML5 and CSS3, but also the Bootstrap framework to achieve a responsive page[5] and Echartjs to display the visualization of data.

A. Crawl data

The system achieves web crawler to get URLs of video resources by using Python, then, manages URLs, parses URLs and creates index. For web site, this system crawls pages and creates index basing on the Apache Nutch framework and stores the results in the database for search. For WeChat, this system achieves crawl program by the use of WeChat's public platform API, so we can obtain XML or JSON format data packets, then parses these data packets and stores them in the database. For SinaWeibo, this system achieves crawl program by the use of SinaWeibo's public API to acquire micro-blog content, comments, topics, user information and other data efficiently. For video sites, this system uses web crawler technology based on Ajax technology to analysis dynamic web page in order to obtain the relevant video clicks, user comments, user ratings and other data.

B. Store Data

It is necessary to store crawling data in a large data storage platform. The crawling data has both structured data and unstructured data. So, it is necessary to use the research and development tools to realize the loading of crawling data to large data storage platform.

This system realize a data loading tool which is developed based on Sqoop open source framework[6]. It can load both of structured and unstructured data to Hadoop large data storage platform. At the same time, the tool can also achieve the transition of these data from the large data storage platform to relational database. For example, some of the data computed by the Hadoop data storage platform can be exported into a relational database by using the tools in order to achieve convenient data operation, so as to realize the flexible access to data.

C. Data Processing

The quality of data crawling from the Internet is uneven, so data preprocessing is needed before data storage. These preprocessing is also the cornerstone of data analysis. In this system, the preprocessing stage is mainly to filter data, merge data and perform other operations on data by using procedural means such as if-else after crawling data.

There are many data processing methods, mainly including: data cleaning, data integration, data transformation and data reduction, etc... Data cleaning is to "clean up" data by filling in missing values, smoothing noise data, identifying or deleting outliers and resolving inconsistencies. Data integration is to combine multiple data and store them in a unified way. The data transformation is to transform the data into a form suitable for data mining by smoothing aggregation, data generalization, and normalization. Data reduction can get the reduction of the data set and can reduce the size of data sets. However, it is still able to maintain the integrity of the original data, and the results after reduction based on the data mining is same with the results before reduction. That is, without changing the results, the efficiency of the system is improved.

After preprocessing the data, we analyze the processed data by using data mining algorithm. This system performs clustering analysis on videos and divides these videos into trailers, comedy, advertising, travel, love, records, animation, inspirational, spoof and other by using the classic k-means algorithm. Then, classify videos by using the Bias classification. After classification, this system gathers statistics of each category to find the hot topic of the micro videos. In the stage of evaluation and prediction, regression analysis is used to establish the appropriate regression model based on the existing data and then test this regression model. At last, this system predicts market trends of the micro video in the future according to this model.

D. Graphical Display of Analysis Results

After the analysis of the data, the system shows the results of the analysis with various types of graphs by using Echartjs. As shown in the following Fig.2 and Fig.3:

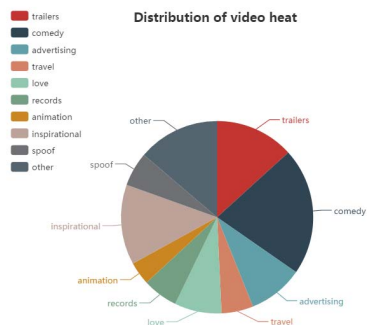


Fig. 2. Distribution of video heat

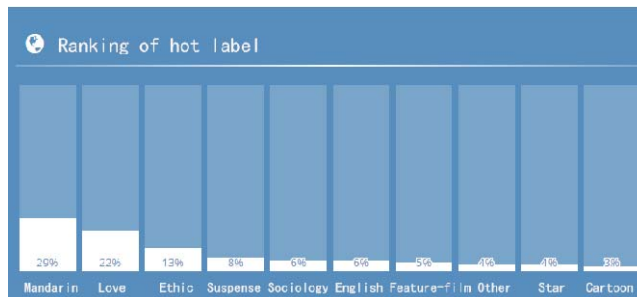


Fig. 3. Ranking of heat label

V. CONCLUSION

In this paper, we give a detailed description of the structure design of the micro video topic recommendation system, including data layer, interface layer, core layer, service layer, business layer. Then research on the key algorithms involved in the system, and pick up a suitable algorithm. Finally, introduce the implementation of the system.

REFERENCES

- [1] Y. B. Wang and X. K. Li, "Research on software architecture design method," J. Computer engineering and design, vol. 26, pp. 604-607, 2005.
- [2] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," J. Data Mining Concepts Models Methods & Algorithms Second Edition, vol. 5, pp. 1 - 18, 2010.
- [3] Hartigan and A. John, "Clustering Algorithms," J. Applied Statistics, vol. 23, pp. 38-41, 2014.
- [4] J. M. Bland and D. J. Altman, "Regression analysis," J. Lancet, vol. 327, pp. 908-909, 1986.
- [5] H. Shu, Y. Xiong and GEXuejiao, "Design and Implementation of Bootstrapbased Responsive Webpage," J. Journal of Beijing Institute of Graphic Communication, 2016.
- [6] M. S. S. Aravinth and A. H. Begam, "An efficient hadoop frameworks sqoop and ambari for big data processing," J. 2015.