Research of the Social Media Data Analyzing Platform Based on Cloud Mining

Yi-Tang ZENG, Yu-Feng ZHANG, Sheng CAO, Li LI, Cheng-Wei ZHANG

School of Management, Wuhan Donghu University, Wuhan, China

346316766@qq.com
*Corresponding author

Keywords: Cloud mining; Social media data; Analyzing.

Abstract. With the development of social media technology, the social network data increasing greatly. The social media data is of volume, complex and dynamic. How to mine the value of the social media data has become the focus and difficulty of the current data mining research. The combination of cloud computing and data mining forms cloud mining technology which is very perfect. The applying of cloud mining technology in social media data analyzing has great advantages. In this paper, the cloud mining is introduced, and the principles of cloud mining technology are explored, a logical and a physical framework of social media data analyzing platform based on cloud mining are constructed, and the advantages of the platform are analyzed. This research provides a feasible way of thinking and method for the mining and analysis of social media data.

Introduction

The data is increasing at an unprecedented speed with the advent of the Internet, which gradually forms big data. “The big data” with the feature of volume, variety, velocity and value is affecting and changing people's life. In recent years, the social media technology is developed quickly, and blogs, forums, content sharing communities, micro-blog, wikis, video sharing sites, social networking sites and communities for comment continuously produce various types of social network data. Compared with the traditional data, social media data has two prominent characteristics: Firstly, the data is very large. The cost of acquisition, storage and use of the data is high. It is a kind of typical big data. Secondly, the data structure is complex. The social media data includes all kinds of information that created or shared by users in social media, such as comments, video, photos, location, personal information, social relations. It is the collection of self organized relationships between actors and it reflects multi-level social entity’s relationship [1].

According to “The 37th Statistical Report on Internet Development in China” which is issued by China Internet Network Information Center (CNNIC), China had 668 million Internet users, and the Internet penetration reached 50.3% by the end of 2015[2]. Ordinary users are not only the consumers of network data, but also the makers of network data. The users’ interactions form a huge social network. The social media data is collected by the micro behavior of each netizen and contains huge value. It will cause changes in the government’s public administration, market research and marketing. Therefore, to realize the mining and analysis of social media data has become the hot research issues focused by the scholars of several fields, such as social business, marketing, data mining, evaluation of public opinion, knowledge management.
In the analysis of social media data, the most important task is to understand the data meaning and explore the nature of data, so as to find the relations between the data and users’ behavior. In this way, the thing hidden in the data can be mined. The research of econsultancy.com points out that many businesses face a problem: They don’t know the meaning of the social media data for their business, and they don’t even know how to analyze [3]. This means that the social media data mining and analysis is difficult, and it is hard to get desired results by using the traditional data mining methods.

Cloud computing is a public participating computing model based on Internet. Its computing resource is dynamic, contractible and virtualized, and it is provided by means of service. It has massive storage capacity & powerful computing and processing ability. Cloud computing has become an effective way to realize massive data mining [4]. The combination of cloud computing and data mining forms the fifth generation of data mining technology that is cloud mining technology. It provides premise and foundation for the deep development and utilization of big data [5].

**Cloud Mining**

Cloud mining refers to the combination of distributed, real-time dynamic data management technology and data mining technology to achieve dynamic, distributed, real-time, efficient processing , mining and analysis of massive data. The success of cloud mining depends on the following key technologies: massive data storage mode, data preprocessing method based on the cloud platform, parallel algorithm for massive data mining based on cloud platform[6].

The cloud mining is the perfect combination of cloud computing and data mining. Users propose data mining tasks to the cloud terminal, and then the cloud terminal mobilizes the corresponding service resources to meet the needs of users. It also provides convenient operation platform, and it helps expand the range of user groups. In this way, the data mining ability is improved, and the cost of mining is reduced. The framework of cloud mining system is shown in Fig. 1.

![Figure 1. The framework of cloud mining system.](image-url)
The realization principles of cloud mining are as follows:

1. Users use computers, pad, mobile phones to log on to the cloud mining system, and then impute mining request. Meanwhile, the users can set the corresponding algorithm parameters, and impute the basic data.

2. After receiving the user's request, the cloud mining system responds immediately. Firstly, it analyzes the idle state of the working nodes, and then, it assigns the mining task to the idle working nodes.

3. According to the request and algorithm parameters imputed, the cloud mining system starts to deal with the data imputed by users in the distributed storage system, such as calculating missing data, completing data type conversion, filtering noise data, eliminating duplicate records.

4. Cloud mining system working nodes automatically select the corresponding data mining algorithm to carry out parallel data mining on preprocessed data. After the model evaluation and interpretation, useful information and knowledge will be obtained.

5. Cloud mining system collects the results of mining from each working node. The results of mining will be passed to the users by appropriate visualization tool.

The Logical Framework of Social Media Data Analyzing Platform Based on Cloud Mining

From logic view, the social media data analyzing platform based on cloud mining can be divided into six layers: user interaction layer, platform application layer, platform management layer, data analyzing layer, virtualization resource layer, infrastructure layer, as shown in Fig. 2.

![Figure 2. The logical framework of social media data analyzing platform based on cloud mining.](image)
request to the platform but also obtain the visualization results provided by the platform. The functional module in this layer includes customized analysis task module and knowledge display module.

(2) Platform application layer
This layer provides strategy application to meet the demands of the users. Such as marketing planning, customer behavior analysis, related sales, customer potential demand analysis, customer preference analysis, competitor analysis, marker environment analysis and so on. The functional module in this layer includes task response module, task scheduling module, task correlation algorithm module, task output module.

(3) Platform management layer
This layer provides management and service for social media data analyzing platform. The management includes user management, resource management and platform safety management. The functional module in this layer includes user management module, resource management module, security management module, network management module.

(4) Data analyzing layer
This layer is the core of the whole platform. MapReduce is adopted as the distributed parallel computing model to call the underlying resources according to the demands of users. Through processing the obtained social media data by storage managing, data cleaning, algorithm calling, data mining, results evaluating and results outputting, the results of the analysis of social media data can be provided to service application layer. The functional module in this layer includes data loading module, parallel ETL module, mining algorithm module, pattern evaluation module and parallel output module.

(5) Virtualization resource layer
This layer is the basis of the whole platform. The virtualization of application and hardware can be achieved by virtualization technology to integrate distributed physical resources, so as to build corresponding resource pool where the similar resources form clusters. In this way, the resource can be directly operated and optimized. The functional module in this layer includes data storage module, parallel computing module.

(6) Infrastructure layer
This layer is the cornerstone of the whole platform. It provides the necessary facilities and equipment. A powerful computing cluster is formed by resource virtualization technology. And it can provide physical support for the upper layers to complete normal computing and storage task in each layer. The functional module in this layer only includes the resource virtualization module.

The Physical Framework of Social Media Data Analyzing Platform Based on Cloud Mining
In the framework of social media data analyzing platform based on cloud mining, nodes are established in the following layers, such as large enterprises, the subordinate branches of the large enterprises, suppliers, distributors. Meanwhile, clusters of resources are formed by virtualization technology in each node, which will form physical resource pool. The physical resources pool locates in cloud computing data center which is formed by the distributed server cluster, which can realize the data storage, the algorithm design, the data analysis and information services. The social media data analyzing platform based on cloud mining can make full use of the cheap network resources to carry out parallel processing of data preprocessing and data mining algorithms in the Hadoop environment and the preprocessing
data will be stored in disks of each node by the HDFS. MapReduce programming model is used to complete the data mining tasks. The parallel programming mode of the mining algorithm is carried out by the computers which distribute in the nodes. The physical framework of social media data analyzing platform based on cloud mining is shown in Fig.3.

Figure 3. Physical framework of social media data analyzing platform based on cloud mining.

The Advantages of Social Media Data Analyzing Platform Based on Cloud Mining

The social media data is of great volume, complex, dynamic. It is good for realizing the intelligent analysis of social media data by using the cloud mining technology. Thus, the effectiveness of data analyzing will be enhanced and the cost of data analyzing will be reduced. The social media data analyzing platform based on cloud mining has following advantages:

(1) The internal structure of each logic layer of the social media data analyzing platform is separate. Every layer can provide service for its upper layer transparently and the tasks of each layer are very clear. Each layer can call all the functions directly or indirectly from the lower layers through the interface between the layers and functional module in the same layer can share data and call data effectively.

(2) The social media data analyzing platform has super processing ability of massive data and the very strong scalability. The users can expand the function of the platform by using the good user interface according to their own needs, so as to obtain the social media data and information they need.

(3) The users do not need to develop application software themselves. The platform allows users to propose service request at leisure. And users can customize the social media data mining objects, social media data mining tasks, social media data mining algorithms by making use of the functional modules in the interaction layer.

(4) The social media data analyzing platform can mine and analyze the social media data by using distributed parallel data mining method. It can effectively use and deal with the data which naturally distributes in the nodes. It also can conveniently use the computing devices to unify the message format to realize the data storage and data computing, and these data are multi modal, multi-source, multi format, multimedia. In this way, real-time and efficient social media data analysis will be realized.

(5) Social media data analyzing platform can provide users who do not have the data
mining knowledge with "one-stop service". The users do not need to understand the internal work of the platform, and also need not worry about the capacity of storage and the ability of calculation of the platform. The platform has good demonstration ability, and it can output the results with visual form by comparing the results of different algorithms.

(6) Users do not have to set up computer room and data center. They can obtain the storage space, calculating ability and information analysis services according to their need in social media data analysis platform. Thus, the users can save cost and enhance effectiveness.

Conclusion
This paper introduces the cloud mining technology, and a logical and a physical framework of social media data analyzing platform based on cloud mining are constructed, and the advantages of the platform are analyzed. It is hoped that this research provides a feasible way of thinking and method for the mining and analysis of social media data which is of great volume, dynamic and complex. How to do a good job of acquisition and integration of social media data according to the types and characteristics of social media data, how to choose suitable mining algorithm to get the results of visual analysis and how to promoting the efficiency and effectiveness of analysis take further research.

Acknowledgement
This research was financially supported by the grants from Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce (under Construction) (Wuhan Donghu university research [2016] 15 words) (Research on Agricultural Product Logistics Analysis Based on Cloud Mining under the Big Data).

References


