International Journal of Database Theory and Application
Vol.9, No.6 (2016), pp.171-180
http://dx.doi.org/10.14257/ijdta.2016.9.6.17

Framework and Key Technology Review of Big Data Analysis in the Social Network Background

Zhendong Ji

Shandong Women's University, Shandong Jinan, 250300, China
Namejzd@163.com

Abstract

As the network promoting in people's lives, more and more people use the social networks and other network platform, the concepts of big data have been gradually reference to the network data analysis, researching big data analysis framework and technology can improve enterprise management level. This paper analyzes the framework and its key technology of big data analysis in the context of social network, in order to be able to provide the corresponding theoretical support and reference for analysis of social network data.

Keyword: *Big data analysis; social networks; key technology*

1. Introduction

With the continuous development of the Internet, a variety of social networks emerging, there are more and more data on the social network and gather into big, analyzing in depth of these data can effectively resolve various issues of social and enterprises, and big data analysis provides a convenient for enterprise to analyze network data, there are more and more institutions adding into the research of big data analysis, the development of big data analysis techniques has brought enormous value for social management and enterprise management.

2. Research Background

With the rapid development of information network in the 21st Century, existence of large data and establishment of analysis model is increasingly influencing and even changing the overall operation management and developmental pattern of an enterprise or even the nation. However, hundreds of millions of existence of network data is involved with widespread levels, in all aspects of the individual, enterprise and the nation. The universality and complexity makes the analysis and data reductions become a big problem of the development in 21st Century. It means that under the circumstance of today's social network, anyone who takes lead in mastering the scientific and standardized analysis of big data and establishing the high effective data analysis model will seize the market opportunities and win the initiative in competition. In the current economic development, our country attaches great importance to the sustainable scientific development of economy, taking transformation of economic development mode as the direction and focusing on the improvement of the structure of economic development. However, the emergence of big data provides breakthrough point for transformation of China's economic development mode and adjustment of structure of economic development. But at the same time, because China's technology of information network falls behind western developed countries and starts late, related personnel need to do corresponding theoretical study and specific model analysis and study of the key technology based on the existing characteristics of big data in China according to the analytical model of big data in

ISSN: 2005-4270 IJDTA
Copyright ⓒ 2016 SERSC

advanced countries, in order to promote the analysis of large data formation in China in the framework of the society under the background of the network.

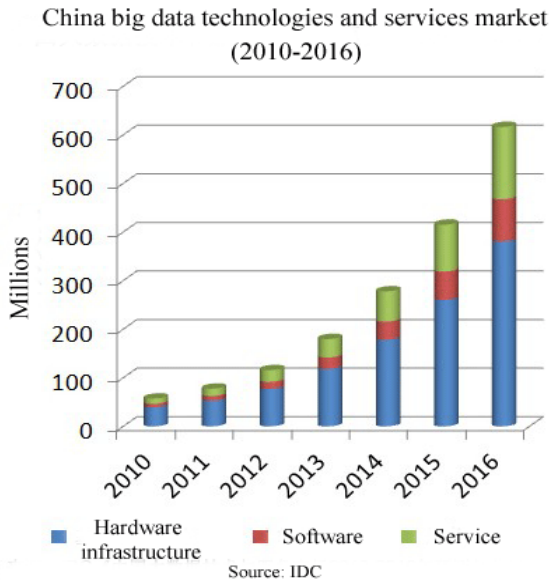
2. Overview of Big Data Analysis

The concept of Big Data is appeared by companied with the development the Internet, which is refers data sets with a particularly large volume and data categories, which making it impossible to use traditional databases to capturing, processing and analysis the data, The volume of big data are generally in 10TB, and in the practical application, many companies applied several data sets together, which makes the body more substantial. In the category of data, since the data from a multiple data sources, the category is no longer the conventional areas of structure, even including unstructured and semi-structured data, which make it different in data categories than before. Specifically, big data includes the following features:

2.1. The Availability of Data

With the development of social network, more and more individual, enterprise and even national institutions have joined the social network. The individual joins the social network because he wants to understand himself through social network, or other people's behaviors or is interest, such as Micro-blog landing, zone, circle of friends, renren and other social network. The enterprise wants to expand the business mode of production and broaden the production chain through social network, do data collection, hire professionals for large data analysis, and establish the analysis model of large data in enterprise. On one hand, it is beneficial to seize the market. On the other hand, market segments according to the results of the analysis data model and characteristics of consumers' behavior is beneficial to the targeted marketing management, thus increasing the value of enterprises and establishing a corporate image. But the country gets the public voice and focus of public attention through keyword search, suggestion collection and other methods in social network, so as to grasp the dynamics of social development, understand the people's livelihood and more effectively carry out management. It can be seen that social network data has wide serviceability in every aspect of social development. Figure 1 is the prediction and analysis of large data technology and service market in 2012 to 2016. With the expansion of the social network application scope, the social network data is changing our personal behavior habit, life style, life concept and life attitude, even the mode of production, enterprise management strategy, corporate planning and the development of the entire business model, the development of the entire business model and the overall national development strategy and development mode. It can be seen that the social network data has become an indispensable part in modern society and its widespread use has been fully demonstrated.

 Copyright ⓒ 2016 SERSC



<Analysis and forecast of China big data technology and services market 2012-2016>

Figure 1. Analysis and Forecast of China Big Data Technology and Services Market 2012-2016

2.2 The Fast Growth Data

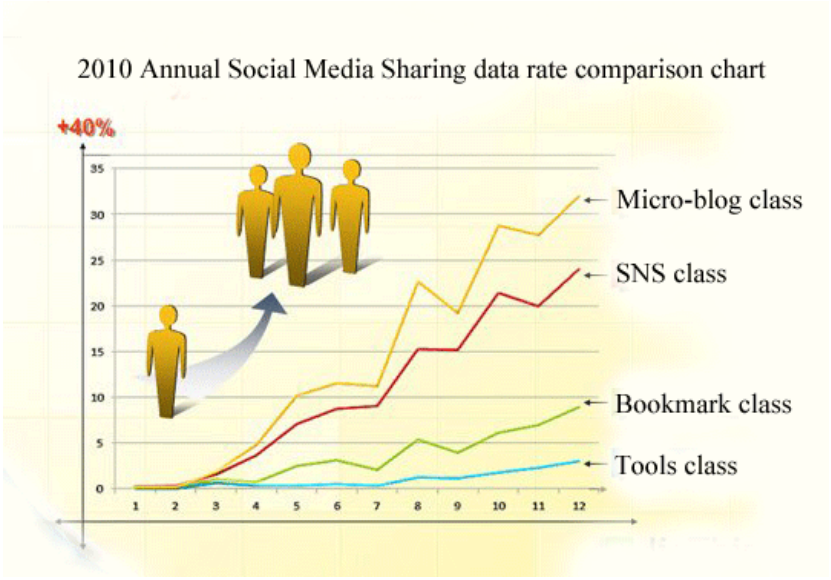


Figure 2. The 2010 Annual Social Media Sharing Data Rate

With the more and more applied in our life of the Internet, faster and faster development of Internet technology, social networks are growing too, users is growing and wide, user groups with more and more wide scale, and the social networks data also showed a trend of rapid development. As shown in Figure 2, Social network data is also showing a trend of rapid development. Take the most widely used Sina microblogging, ended in March 2013, Sina microblogging users has increased by 6.8% compared to the end of 2012, just in four months, Sina microblogging user data grows in millions, The speed of social networks data is very fast. Another example is in the world's most widely used Facebook, The user number in June 2013 years increase of 21% comparing with in

Copyright ⓒ 2016 SERSC

June 2012 year, the number of Twitter users grew by 16.7%, from the growth of these social networks users, we can get the in the growth trend of social network data in foreseeable future. At the same time, social network data include not only the user data, but also includes the message data, such as the micro-Bo, a microblogging users can reach one thousand messages within a month, this is just one user's data, and there are millions of microblogging users, the number of messages can be tens of billions in just one month, the growth is faster, and the huge data has brought great challenges to the data analysis and processing.

2.3 The Large and Complex Data Type and Scope

Big Data obtained huge number and complex type from literally, the amount of data is very large in social networks, just microblogging, Sina microblogging has more than 540 million users, the amount of Tencent microblogging has more than 580 million users, which resulting the data of user is very large, and the message number of per user is increasing every day, Four social networks including Facebook, Sina microblogging, Twitter, Tencent microblogging, the daily volume of messages has exceeded 1 billion, 2 billion and 2.5 billion, these huge data has laid a good foundation for big data analysis. Furthermore, the type of big data is very rich and diverse, just as mentioned earlier, and now the data type has get rid of structural model, the unstructured and semi-structured data types are also included. Each user has the basic information including user name, Email, address, the created time, the number of friends and so on.

And there are some relationships between each user, which are the relation data, and the more users probably constitute a group because of common concern or interest, so the group appearing, the views and comments in the process of the user taking a message constitute the unstructured information data

Thus, a lot of data types and structures included in a social network, these data types have different structure and storage methods, constitute the diverse social networks together, and how to analyze these diverse and complex data structure and dig out the valuable information for individuals, businesses, and national service, is the main task of big data analysis tools.

3. The Framework of Big Data Analysis

In the context of social networks, big data analysis technology has become a hot technology in network data analysis, social network data analysis covers anthropology, sociology, psychology, geography, communications technology, organizational study, biology and other aspects of knowledge, it is a multi-disciplinary cross multifaceted technology, the social network also contains a large information data, the relationship data and user data, *etc.* to analyze the different types of data requiring complex analytical techniques to accomplish, the corresponding data analytical framework can structure for all aspects of a comprehensive data analysis. The framework of big data analysis is data layer, the support layer, the analysis layer and the result showing layer, the basic of the whole data analysis is data layer, including capture and storage data for the social network, the analysis layer and the support layer constituted the core of the data analysis framework which analyzing and processing data in-depth, the result showing layer can show data in a clearly and friendly way and apply it to the associated program which, specifically, the functionality and structure of each layer are as follows:

 Copyright ⓒ 2016 SERSC

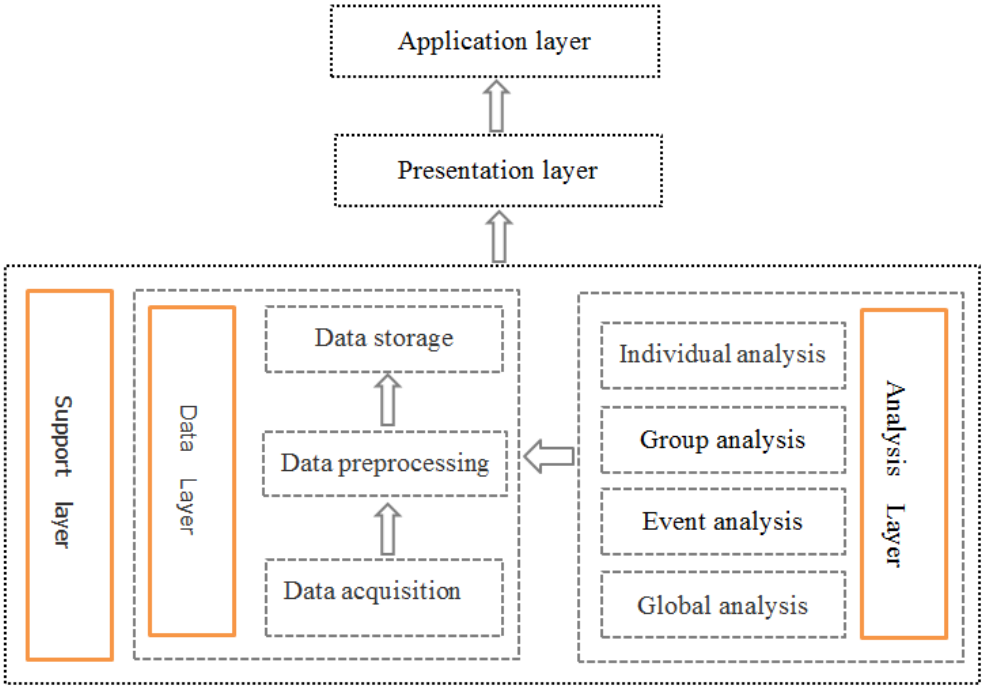


Figure 3. Large Data Analysis Framework

3.1. The Data Layer

The data layer is the basis of the data analysis framework, which includes data acquisition, there are three main ways of the data acquisition by framework, the first way is through the network flow, to get the data by identify and resolve network flow, the second way is by API or not-API, including web analytic method, distributed crawler method and API overrun determination methods, the third way is to get the data by other ways, such as meta-search method, incremental method or the acquisition of specific groups, *etc.* To collect the various data in social networks through these methods, and then pretreated. The data processing is the initial big data processing in the data layer, including data cleaning, marking, association, *etc.* data cleansing mainly do data normalization, deduplication and compression, and then marking for the data, mainly for data relationship weights, extracted keyword from the information data, and marking to the group data, mainly on the micro-blog or tweets, on the events and characters. The final preprocessing step is processing data association, including data storage, log storage and characteristics storage.

3.2. The Analysis Layer

The analysis layer is one of the core layer of the data analysis framework, which is further analysis for the preprocessed of data layer, specifically to analyze by the four dimensions, the first dimension is the individual analysis, including analyzing the individual circles, emotions, interest, behavior and characteristics, tightness and area networks, *etc.*, the user's personal interests and behavior habits can by able to understand from the analysis of this information. The second dimension is the group analysis, including analysis of specific groups, analyzing opinions of the group leaders, analyzing the relationship, analyzing groups interest, analyzing the evolution of the group, analyzing the potential group members recommendation, it is possible to understand the group by analyzing the information, and which can manage the group of targeted. The third dimension is the event analysis, including analysis of the events discovery, to restore the

Copyright ⓒ 2016 SERSC

path of the incident, to hunt for the source of the event, to analyze the propagation of events, to analyze the opinion leader of the events, to predict the trend of events spread *etc.* through these analysis, The occurrence and the development of the event can be grasped, which have an important role for the master in the developments of the event. The fourth dimension is the overall analysis, including sorting of the popular characters or events, doing statistics of the overall platform, analyzing the regional hot events, analyzing the global topology, The whole social network can be grasped by analyzing these data, in order to understand the status and the role of social networking in the network groups.

3.3. The Support Layer

The support layer is another important core of data analytical framework, including distributed parallel processing, machine learning, data mining, visualization techniques and natural language processing technology, these key technologies support the operation of data analysis of the entire social network, analyzing and processing by a intelligent and efficient by these techniques for the data in social network, and getting the relevant analytical data.

3.4. The Showing Layer of Results

The showing layer of results is showing the final results from the analysis layer of which that through by the platform, including individual showing, event showing, groups showing and the whole showing, these results showing ultimately be applied to the relevant application, according to the results of data analysis can be fitted to a variety of applications to help enterprises or social networks do more application which will be more in line with the user experience into social networks. The specific applications include advertising precise delivery, community management, personalized information, the poor user.

4. The Research of Key Technologies in Big Data Analysis

The all levels of data analysis framework need to various technology to support to complete the data collection, processing and showing, these key technologies occupies an important position in the data analysis, the key technology in each level is different, these different key technologies support the entire data analysis framework together.

4.1 The Data Layer

The key techniques or methods used in data layer consists of three parts, the first one is data acquisition, data acquisition methods mentioned above is contained three methods, network flow is mainly on the way to obtain data which is not encryption. This technique can get the network flow data quickly by resolving and identifying, the system firstly needs to analyze the corresponding network flow during acquisition, and then extracts the significant network flow characteristics, identifying the target flow from the background flow. Then, resolving the target flow, extracting the relationship data, basic data, forwarding data, *etc.*, to realize data acquisition. The second way is to get the API or not API method to obtain data, API is an application programming interface, there are many social networking services provide API, data analysis framework can obtain the relevant data from that, OAuth1.0 and 2.0 are two common ways, the basis information and relationship data of social network can be obtained by connecting the user interface on, there are a certain number and speed restrictions to protect user data in API providing by social network, to avoid being illegal or violent acquisition, when getting data by API it is necessary to control the frequency and speed, in order to ensure the data acquisition is normal. In order to break through the limitations of API, and now there are non-API

 Copyright ⓒ 2016 SERSC

acquisition mode like the web resolving, using the technology of Web crawler to simulate user logs to obtain the appropriate data.

The second one is data preprocessing, including cleaning data, marking and data association. Wherein the data cleaning technology mainly is processing the data accuracy, effectiveness and integrity, filling in the missing data, to eliminate the abnormal data, smoothing the noise data, *etc.*, to complete the data cleaning process. At the same time data cleaning technology can also compress the data to reduce the data storage space, improve storage capacity. The data marking technology is marking to data with a variety of types, according to the current social network data analysis, which can be divided into groups marking, the characters marking, the relationship weight marking, events marking and microblogging marking and so on. Finally is the data relationship technology, there are a lot of information and data are interrelated in the social network, which need to rely on data analysis techniques to correlate data relationships, such as a user using the same basic information to register Microblogging, Renren, Facebook and other social networks, these networks information will be linked to the only one user, so the data association technique is necessary to associate these social networks number. Furthermore, if the same event broadcast in different social networks, data association technique also requires to make information associate and integrate.

The third one is the data storage technology, to analyze huge data from social networks needs to have enough space to store the data together, The storage technologies of big data analysis framework includes data storage, log storage, characteristics storage and history storage library, *etc.*, different types of data can be stored by these storage technologies, so the data layer can analyze. Big Data Analytics frame data analysis and data mining basic form of expression is an object - attribute matrix or table, shown in Figure 4.

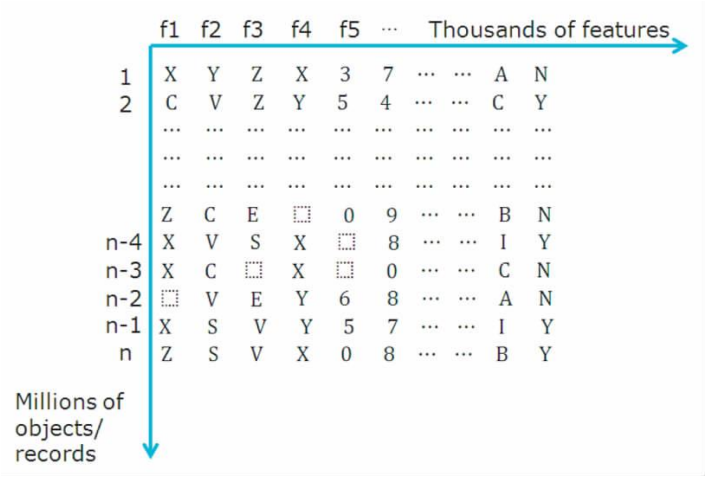


Figure 4. Object-Attribute Matrix

4.2. The Analysis Layer

The key technical in the analysis layer mainly include individual analysis techniques, group analysis, event analysis and the overall analysis, individual analysis techniques is mainly analysis on users' personal information data, including knowledge of individual emotion, behavior, such as analyzing the individual circles is analyzing the relationship, and then summarized the impact of the individual in the circles, and also be able to predict the evolution of the entire circle through individual circle analysis. Sentiment Analysis is to analyze the user's emotional tendencies, including individual negative emotions, neutral emotions and positive emotions, *etc.*, in order to understand the individual's emotional

Copyright ⓒ 2016 SERSC

changes, or the attitude of the individual to the events. Group analysis is analyzing the group which is composed of multiple users, including cohesion among this group, the relationship, the relationship between the groups, interests, behavior, psychological and impact, to understand the organization's position and characteristics in the social network.

Event analysis is a technique which analyzing impact of the events occur in the social network or in life, including development of the process of the entire event, attitude in the different groups and the impact of the events generated. The overall analysis is analyzing the entire social network, including the popular figure, popular events, regional events, and so hot, for example the popular microblogging rankings in the Sina microblogging interface, which is the show of the overall analysis.

4.3. The Support Layer and the Showing Layer of Results

The key technology designed in the support layer is many , including machine learning, data mining technology, natural language processing technology and visualization technology, *etc.*, these technologies perform their duties jointly support the entire data analysis framework, these technologies have played supporting role in each module, and there are some similarities, the support layer is composed of the extracted common technologies.

The showing layer of result is a platform which display the data in a direct way, the technology which it is applied mainly is user experience, which have the characteristics of visibility, multi-dimensional and interactive, *etc.*, to show the analysis results more friendly, simple and visually, allowing users to get more user experience, thus the user data can increase continuous and promote the development of social networks.

Thus, a variety of key technology be applied in the framework of the four-level can enrich the content of the data analysis framework continuous, which provides more powerful data analysis capabilities for the data framework, to analyze the different objects effectively in social network to provide the appropriate reference data for enterprises, social networks and country.

5. Conclusions

With the development of the Internet, The scale of social networks have been expanding, and contains more valuable information, in order to explore this information more effectively, it is necessary to do some research and explore on big data analysis techniques. The paper focuses on big data analysis framework and the related key technologies to discuss, hoping to play a valuable role in helping mining the potential value in the social network, to be able to build more social network applications humane and fit people's lives needed based on data analysis.

Acknowledgement

5. No. : ZR2013FL026 2014 Natural Science Fund Program of Shandong: A Study on the Mining of Mass Traffic Moving Data Based on Cloud Technologies.

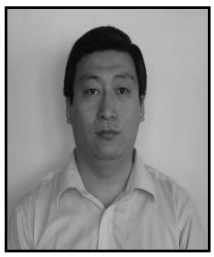
References

[1] D. B. Lin, L. S. Song and Z. S. Yong, “The social networks analysis by the structure”, The Computer Journal, no. 4, **(2012)**.
[2] Z. Sai, X. Ke and L. H. Tao, “The measurement and analysis of information dissemination of Microblogging class social network”, Xi'an Jiaotong University Journal, no. 2, **(2013)**.
[3] Y. C. Qi, B. Y. Yuan and X. Y. Bo, “Social networking big data analysis framework and its key technologies”, ZTE technology, no. 1, **(2014)**.
[4] M. Tao and Z. Wen, “The ponder for big data analysis technology”, Information and communication technology, no. 6, **(2013)**.
[5] L. C. Ming, W. Tao and Z. H. Wen, “Big Data help China Unicom to enhance the quality of information consumption”, Information and communication technology, no. 6, **(2013)**.

 Copyright ⓒ 2016 SERSC

[6] Z. Qing and T. Jing, "The Research of telecommunications data warehouse architecture and key technologies", Information and communication technology, no. 6, **(2013)**.
[7] S. G. Zhong, X. Xing and W. An, "The technologies and applications of Removal Anonymous in the era of big data", Information and communication technology, no. 6, **(2013)**.
[8] J. N. Yun, Z. H. Song and X. Xia, "Big Data analysis techniques in resource of culture management", Chinese basic science, no. 1, **(2014)**.
[9] L. Rui, "The big data analysis and application systems on radio and television operators", Cable TV technology, no. 9, **(2013)**.
[10] Z. Y. Zhuang and L. J. Feng, "Information security opportunities and challenges in Big Data era: The example as disclose information and intelligence", Defense Technology, no. 3, **(2013)**.
[11] R. M. Bond, C. J. Fariss and J. J. Jones, "A61-million-person experiment in social influence and political mobilization", Nature, vol. 489, **(2012)**, pp. 295-298.
[12] P. K. Atir, "Supervised rank aggregation approach for link prediction of the 21st international conference companion on World Wide Web", 2012, ACM, New York, NY, USA: ACM, **(2012)**, pp. 1189-1196.
[13] G. Hacidh, "A predictive model for the temporal dynamics of information diffusion in online social networks", Proceedings of the 21st international conference companion on World Wide Web, 2012, ACM, New York, NY, USA: ACM, **(2012)**, pp. 1145-1152.
[14] H. Yongqiang, L. Rubao and H. Yin, "RCFile: a fast and space-efficient data placement structure in MapReduce based ware-house systems", Proceedings of the 27th international conference on data engineering. Hannover, Germany: IEEE, **(2011)**, pp. 1199-1208.
[15] M. Stonebraker, D. J. Abadi and D. J. DeWit, "MapReduce and parallel DBMSs: friends or foes", Communications of the ACM, vol. 53, no. 1, **(2010)**, pp. 64-71.

Author



Zhengdong Ji, received the B. Sc in Applied Mathematics from Shandong Normal University (SDNU) and M. Eng degree in Software Engineering from Shandong University (SDU) in 2000 and 2011respectively. He is currently researching on Data Mining (DM) and Mathematical Modeling.

Copyright ⓒ 2016 SERSC

 Copyright ⓒ 2016 SERSC