

Construction and Optimization of Data Oriented Precise Information Mining System

Ying Xiong

Foshan Polytechnic Library, Foshan, China

Abstract: With the in-depth cross integration of disciplines and the enhancement of the integration of social development, economic development and scientific and technological development, the research of big data intelligence is developing from single domain analysis to full domain analysis. Data fine segmentation and presentation results highlight the characteristics of intelligence in data calculation, to meet the most accurate needs of users. Data integration can solve these problems and realize data sharing and interoperability. There are many reasons, such as different data models and data patterns, which are distributed, autonomous and heterogeneous, so it is difficult to perform interoperability. From the perspective of competitive intelligence acquisition and research analysis, there is still a certain gap in the domestic information market. However, with the urgent demand for competitive intelligence information, the quality improvement of domestic competitive intelligence system is an urgent development trend.

Keywords: *Data Oriented, Data Mining, Precise Information Mining System, Construction and Optimization*

I. INTRODUCTION

More and more industries rely on the ability to use digital scientific data and complex data mining, integration, analysis and visualization tools to transform it into information and knowledge. For big data intelligence research, the application of intelligent technology can automatically carry out advanced and complex information processing and analysis work, to a large extent, freeing researchers from the tedious physical labor. In the era of information environment changing rapidly, it is very important to collect information analysis and feedback in time, which needs the support of intelligent technology.

The study of competitive intelligence began with the commercial competition between the United States, Japan and the developed countries in Europe after the Second World War, and rose rapidly in the 1960s and 1970s when the Japanese manufacturing industry rose and the American manufacturing industry declined, and effectively helped the American manufacturing industry regain its momentum in the 1980s and 1990s. With the development of information technology and Internet technology, the research of competitive intelligence has been continuously improved, which has become a research field covering information science, library science, management science, mathematical statistics, economics, computer science and other disciplines, and has been widely used in national and industrial R & D strategies and applications. The complexity and integration of emerging technologies lead to the common characteristics of emerging technologies, as well as their relatively independent typical characteristics.

For data mining, on the whole, we need to find the process to support this decision through the integration of data. Therefore, for the information in data mining, it can not only be the data existing in the database, but also some data aggregation linked in various forms. Network information resources belong to the

information in data mining. Under the current technical ability, data mining can only deal with numerical data. How to deal with these data is also the most difficult problem for data mining.

The network mining technology applies the improved keyword engine weighting algorithm, so as to grasp the accuracy of the search target and index, improve the retrieval speed, and make users more and more like to use the network information mining technology to help them search personalized information needs.

II. THE PROPOSED METHODOLOGY

Data Oriented Analysis. Big data intelligence analysis contents of threats to information security: hackers, hacker sources, hacker attributes, hackers' use of weapons, hacker techniques, weapon genes, variants, traces of intrusion means, hiding means, ways of stealing secrets, means of transportation, malicious IP database, springboard, DNS suppliers, behavior purpose output, etc.

The network world and the physical world are not isolated. The network world is the reflection of the physical world level. Data is the DNA that seamlessly connects the Internet world with the physical world. Discovering data DNA and recombining data DNA are the continuous process of human understanding, exploring and practicing big data.

The information is formed by fully calculating the associated fragment big data, which is free to construct and present the most intuitive data mining content to users. Today's big data is the inevitable trend and demand point presented by information analysis. Due to the limitation of computing technology and application scenarios, many big data information resource providers can't meet the real needs or deep mining and presentation needs. They can only achieve big data computing and presentation. The ability of information guidance and indexing needs to be improved, and the ability of prediction and presentation and early warning is not high.

For network mining, there are a lot of heterogeneous and decentralized network documents. For data mining, the object is common database and common data warehouse. Taking database as the analysis object, the information on the Internet is equivalent to a more complex and huge database, and the site is equivalent to a data source with heterogeneous characteristics. These sites are composed of XML, HTML, or text and picture. Through these structured or semi-structured data sources, a large and heterogeneous network mining information is formed.

From the technical point of view, visualization, data mining and other computer technologies provide a powerful technical perspective for information research, which in turn gives guidance to the development of other technology fields. Logically speaking, the network belongs to a directed graph with document nodes and hyperlinks. Therefore, for the network mining, the content obtained through the network is not only

related to the content and structure of the network, but also can get the user's reading of the log during the browsing process.

The network is a kind of information source, which has strong dynamic characteristics. Both the number of web pages and the information content in web pages are growing and updating rapidly. For example, news page, stock page, company's advertising page and service center in the network all update their website content in real time, and the link information and access records in the page are also constantly updated.

From the technical point of view, visualization, data mining and other computer technologies provide a powerful technical perspective for information research, which in turn gives guidance to the development of other technology fields. Logically speaking, the network belongs to a directed graph with document nodes and hyperlinks. Therefore, for the network mining, the content obtained through the network is not only related to the content and structure of the network, but also can get the user's reading of the log during the browsing process.

The network is a kind of information source, which has strong dynamic characteristics. Both the number of web pages and the information content in web pages are growing and updating rapidly. For example, news page, stock page, company's advertising page and service center in the network all update their website content in real time, and the link information and access records in the page are also constantly updated.

The research on technological innovation in the process of scientific development can be traced back to or earlier than the Second World War. However, the "emerging technology" is exactly proposed by the "emerging technology management research program" launched by Wharton School of the University of Pennsylvania in 1994. They define "emerging technology" as "based on science, which may create a new industry or change." Change the innovation of an old industry ". Although, since then, emerging technology and its related research has been widely interested in politics, industry and academia around the world, and has formed an interdisciplinary research field with a unique perspective. However, the definition of "emerging technology" has always lacked a unified standard.

Intelligence Competition and Mining. From the aspect of information retrieval, the mining of network content is actually to improve the accuracy of information, while from the aspect of database, it is to integrate information by creating relevant models. In addition to a large number of public information that can be browsed by the vast majority of users on the network, there are also some information that can not be browsed by some users on the network. It is impossible to find these information by searching and browsing. Some of these information belongs to private information, and the other part is stored in DBMS or exists dynamically in user access. It is precisely because the resource classification of network content mining has diversified characteristics. So some scholars also call it the multimedia information mining of network.

In fact, web content mining is inseparable from text and link. In web content mining, use record is only its auxiliary function, but it will still be used. In addition, web content mining can also achieve the purpose of mining content by linking content and information such as anchor text. Many information of authority page and hub page can be obtained through related links. According to the scope of responsibilities, it can be divided into local scope and global scope. The former is a specific network

site; the latter represents the whole network. Therefore, for the mining of network content and network structure, from the aspect of information retrieval, both belong to the global. From the aspect of database, they are local. The main user requirements and operation service scope of the network intelligence system. The system can provide users with timely and accurate information on a regular basis according to their will.

Users can set their own needs according to their own needs, and get the information they want timely, effectively and comprehensively. These information are obtained after the system filters, integrates, edits and rearranges according to the user's needs. The main official Chinese and English websites at home and abroad are within the scope of detection. In addition, the major portal websites, news websites and industry websites in the network are also within the scope of detection. It is no exaggeration to say that the network intelligence monitoring has covered all the Chinese and English news agencies, newspapers, magazines, radio and television and other different sites in the world, including the sites of government agencies, information center stations and various associations.

With the rapid development of information technology and visualization technology, the highly intelligent quantitative analysis methods, such as artificial intelligence, data mining and pattern recognition, have come into the vision of researchers. The software tools for mining and displaying scientific and technological text information provide an effective tool for the technical roadmap based on quantitative analysis.

The network information information system mainly collects almost all the dynamic websites with a certain public reliability on the Internet in real time and continuously. In this way, the system collects the latest content of almost all the public trust information on the Internet, which can become the basic information base of the system. Through intelligent analysis technology, Internet harvesting will filter, arrange, remove, analyze, sort and classify these contents, and become the unique and effective information database of the system.

Intelligence collection management can also be called intelligence mining. In essence, it is used to collect all kinds of unprocessed data from the external environment. According to the input conditions, the system can feed back the corresponding output validity and priority in time, and can feed back the processing activities in time, and automatically adjust the collection environment and scope. In a word, through the established information and information subject, then collect and sort out all kinds of relevant information, and carry out the preliminary screening, recording, filing and other data or documents storage of information.

Intelligence integrated management can also be called intelligence data processing. It can gather intelligence information from different data sources from the external environment, and transform these information into information that can be easily browsed by enterprise users with XML enterprise heterogeneous data integration method.

Optimization of Information Mining System. Information service management is mainly to output the analysis results of information in a way. It is a functional module implemented for the information service personnel. First, according to the objectives of leaders at all levels of the enterprise, it analyzes the relevant information results, and transmits the information to the decision-maker's book through the system management method, usually in the ways of TXT, word, PDF, xlsx, etc. Assuming that the information of one-time service is not fully

satisfied, supplementary analysis services can be carried out to fulfill the requirements and objectives of decision makers as much as possible.

According to the non functional requirements of this design, the main network environment is divided into three parts: intranet, extranet, Internet. The main scope of network information mining is the Internet, because the Internet is a huge computer system composed of TCP / IP protocol, which has innumerable information data. The management of the Internet is relatively loose, so there are hidden dangers in terms of security. When introducing the Internet, the network information mining system needs to strictly control the security management and strict protection measures of the network.

The system involves three main networks: organization, information and interpersonal networks.

(1) organization network. The organizational network belongs to the internal organizational structure of an enterprise, which is specialized in the collection and analysis of intelligence data and the output of results and reports. The organizational structure of an enterprise generally includes: marketing department, research and development department and production department, which deal with the direct information coordination work of each department.

(2) information network. Information network is the main information processing means of the whole system, which mainly realizes the function of network information collection, analysis and dissemination of competitive intelligence, and is the core part of the whole system.

The competitive intelligence collection subsystem is based on the information subject arranged by the enterprise information supervisor. It uses computer and network technology to mine various data source information of the enterprise's environment, obtain the network information related to the information subject, and conduct the preliminary sorting of the information.

The information integration subsystem is to integrate the information obtained by the collection subsystem and store the information in the database.

The competitive intelligence analysis subsystem is to use the network mining analysis technology to analyze the network information collected from the information integration subsystem, and to obtain competitive intelligence products that meet the information subject. The information source can come from the surrounding environment or internal environment of the enterprise. The internal departments can realize the circulation and sharing of information, and collect internal intelligence information, such as: production product information, service information, customer information, etc. The internal staff of the enterprise can collect work with the help of the network, mine the interpersonal network in the system, and obtain useful enterprise competition information. The external environment belongs to the external network environment of the enterprise. Firstly, the staff input the set keywords, search the relevant web page information through the search engine technology and save it in the corresponding documents.

In the intelligence planning module, users set up key intelligence topics in the module according to the intelligence competition model for decomposition and analysis. The system will automatically push the latest information according to the subject for user reference. This module can effectively assist users to collect and analyze information competition for their own enterprises and decisions.

CONCLUSION

The collection subsystem of the network information mining system is responsible for the collection of information sources, which provides the basis for the initial data resources of information collection, and also determines the quality of information. Network information is an important source of market information. The simple network information mining system analyzed in this paper provides an important means for enterprises to quickly obtain favorable information.

References

- [1] Choi, Y. and Park, D.H., 2017. Development of Youke mining system with Youke's travel demand and insight based on web search traffic information. *Journal of Intelligence and Information Systems*, 23(3), pp.155-175.
- [2] Ding, R., Boutet, E., Lieberherr, D., Schneider, M., Tognolli, M., Wu, C.H., Vijay-Shanker, K. and Arighi, C.N., 2017. eGenPub, a text mining system for extending computationally mapped bibliography for UniProt Knowledgebase by capturing centrality. *Database*, 2017.
- [3] Lv, S., 2020. Construction of marine ship automatic identification system data mining platform based on big data. *Journal of Intelligent & Fuzzy Systems*, 38(2), pp.1249-1255.
- [4] Li, Y., Wang, D., He, H., Jiao, L. and Xue, Y., 2017. Mining intrinsic information by matrix factorization-based approaches for collaborative filtering in recommender systems. *Neurocomputing*, 249, pp.48-63.
- [5] Wang, Y., Li, J. and Wang, H.H., 2019. Cluster and cloud computing framework for scientific metrology in flow control. *Cluster Computing*, 22(1), pp.1189-1198.
- [6] Thulasiram, K. and Ramakrishna, S., 2017. An Automated Testing Approach in Data Mining System using Genetic Algorithm Framework. *International Journal*, 7(1).
- [7] Huang, Y.F., Liu, J.S. and Chen, P.H., 2019, December. Social Content Mining in Social Networks. In 2019 International Conference on Machine Learning and Data Engineering (iCMLDE) (pp. 54-59). IEEE.
- [8] Naidu, P.A. and Krishna, M.V., 2017. Comprehensive Review on Privacy Preserving Data Mining Techniques and Methods. *International Journal of Engineering and Management Research (IJEMR)*, 7(1), pp.121-126.
- [9] Yamaguchi, S., Terada, T., Manaskasemsak, B., Rungsawang, A. and Leelaprute, P., 2017, June. Tour Miner: Mining system of tour plans from SNS:—Smelting function from travel records to tour routes—. In 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW) (pp. 239-240). IEEE.