

Research on an Enterprise Profiling Technique Based on Government Big Data

Yongbo Li, Suping Liu *

Guangdong University of Science and Technology, Dongguan, Guangdong, 523083, China

* Corresponding author: Suping Liu (Email: 17039636@qq.com)

Abstract: With the development of modern society and global integration of the economy, as well as the domestic commercial system reform continues to promote the depth of the number of enterprises is growing rapidly, playing an important role in promoting the development of the national economy growth and stabilization of employment, while the number of small and medium-sized enterprise write-offs each year is also increasing. In order to better understand and grasp the comprehensive development quality of enterprises, taking Dongguan City as an example, by obtaining the operational data reported to the governmental system by enterprises in the process of operation, we put forward a method based on the processing of governmental big data to conduct research and application of the health of enterprise operation and development using the enterprise profiling technology, so as to provide data reference for the government to guide and cultivate enterprises.

Keywords: Government Big Data; Healthy Business Development; Enterprise Profiling.

1. Introduction

In recent years, with the rapid development of modern society and economy, the deepening of the reform of the commercial system, the continuous optimization of the business environment in Dongguan, the continuous stimulation of market vitality, the increasing convenience of enterprise registration, and the rapid growth of the number of enterprises. According to the data of Dongguan Statistical Data Public Service platform released by Dongguan Bureau of Statistics, as of 2021, the total number of registered industrial and commercial households in Dongguan will be 1,463,400, with a total of 688,700 enterprises, including 36,500 domestic (non-private) enterprises and 635,800 domestic private enterprises. As well as 15,200 foreign-funded enterprises, enterprises have accounted for 47.07% of the city's market players, playing a significant role in promoting economic development and stable employment. With the rapid increase of the number of enterprises registered in the government system and the data generated in the business process, the quality of business development of enterprises has brought certain difficulties to the supervision, guidance and cultivation of government departments. In order to solve this problem, this paper studies a kind of enterprise profiling technology based on the processing and mining of enterprise government big data, comprehensively evaluates the quality of enterprise development, helps the government and the public better understand and supervise the health of enterprise operation, and provides reference for the innovation of enterprise profiling in the application of enterprise big data.

2. Research on Enterprise Profiling Technology

The enterprise profiling technology based on government big data first needs to take the government big data collection involving enterprises as the data basis of enterprise profiling, and then according to the implementation path of big data key technologies, through data collection and pre-processing, data

storage and management, data processing and analysis and data visualization, and finally realize the application of enterprise profiling technology. Based on the original enterprise data obtained from the big data set of government affairs, big data technology is used to extract enterprise characteristics and label them, and an enterprise profiling indicator system is established, so as to sketch the business activities of enterprises.

2.1. Data Sources

The enterprise data in the government big data collection mainly comes from the data registered and regularly updated by the relevant functional departments of the government in which enterprises carry out production and operation activities. As the main body of administrative control of service enterprises, government functional departments need to describe the management attributes of administrative items of enterprises. First, manage and register enterprise affairs through various management systems of government affairs, so as to better supervise and serve enterprises; The second is to regularly publicize publishable enterprise administrative supervision data to make enterprise information more authoritative, transparent and credible, such as the enterprise credit provided by the national enterprise credit information publicity system, which describes the various dimensions of enterprise attribute indicators from basic information, administrative licenses, administrative penalties, abnormal business lists and blacklists, etc. Some provincial and municipal departments have also formed corporate public credit information reports based on such data.

Enterprise data are classified according to the theme of government data, mainly from the themes of economic construction, people's livelihood services, education technology, energy and environment. From the perspective of the management of the municipal functional departments, the enterprise data mainly belongs to the market supervision Bureau, the tax bureau, the Human Resources and Social Security Bureau, the Commerce Bureau, the Water Bureau, the science and Technology Bureau, the ecological environment Bureau, the finance Bureau, the Industry and

information Technology Bureau and the Development and Reform Commission and other departments, involving industry and commerce, taxation, social security, economic and trade, energy, science and technology, environmental protection and other enterprise data. At present, the municipal government network can directly obtain or apply for enterprise-related information on a monthly or annual basis through the municipal government sharing open platform.

2.2. Big Data Critical Technology Realization Path

Key technologies of big data were used to design the data architecture of the enterprise profiling studied, as shown in Figure 1.

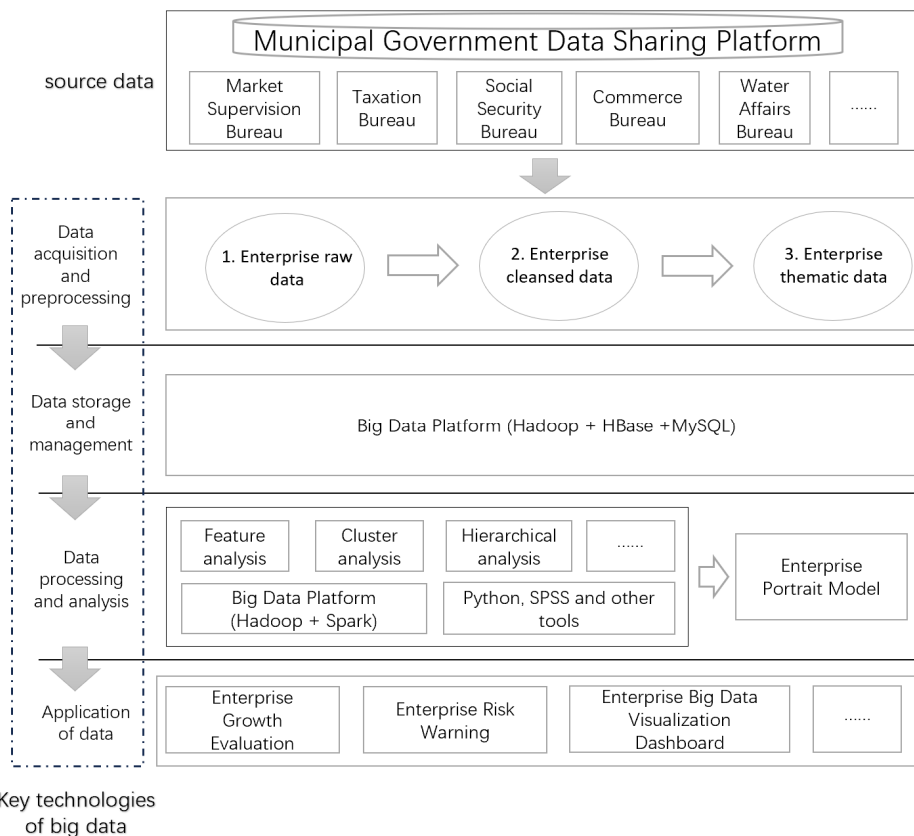


Figure 1. Data Architecture Diagram of Enterprise Profiling Technology Based on Government Big Data

2.2.1. Data Acquisition and Preprocessing

During data collection, according to the rules, regulations and methods of government data management, the enterprise data belonging to the functional departments of the enterprise are regularly exported through the government affairs sharing platform, such as enterprise registration information, shareholder information, annual report information and administrative penalty information of the market supervision bureau, enterprise tax information of the tax bureau, and enterprise employment and social security payment information of the Social Security Bureau. Bureau of Commerce enterprise import and export data, etc.

There are two ways to collect data, one is to export data in media such as CSV, JSON and XML, and the other is to access data through authorized tokens in the way of data interface. The collected enterprise data is structured data, all of which are original data uploaded to the platform by functional departments according to the requirements of government affairs sharing. In this paper, enterprise data exported by conventional CSV is used.

In the pre-processing of enterprise data, it is mainly to clean the missing values and abnormal values of enterprise data, integrate, transform and standardize the data, and then divide and generate enterprise theme data according to functional departments, such as market supervision bureau, tax bureau,

social security Bureau, tax bureau and water bureau.

2.2.2. Data Storage and Management

In this stage, big data platforms Hadoop and HBase are built to store and manage the pre-processed enterprise theme data, update the enterprise theme data regularly, flexibly design the storage of massive enterprise data, and store various types of enterprise data after system expansion. In addition, in the data processing and analysis stage, the enterprise label and configuration information knowledge base of the enterprise profiling can be stored and managed in this stage, such as the enterprise label stored in the MySQL database.

2.2.3. Data Processing and Analysis

In enterprise profiling, the extraction of corporate features is one of the core contents of enterprise profiling technology. In order to describe the health of enterprise operation and establish the label of this enterprise profiling model, it is necessary to extract the relevant enterprise data features and classify them from the enterprise theme data that has been stored in the big data platform. The economic operation of an enterprise includes the basic information of the enterprise, finance, taxation and finance, people's livelihood services, energy utilization, scientific and technological research and development and credit report, etc. The basic information includes the enterprise name, registered capital, date of

establishment, company type, legal person, business term, company business scope, company address, registration place and time and other attributes; Fiscal and tax finance includes personal tax arrears, annual tax payments and receipts and expenditures invoicing. As for the method of enterprise feature extraction, the attributes of enterprise data types are directly extracted, hidden attributes are extracted by statistical analysis, and attributes are extracted by distributed parallel programming model and computing framework Spark combined with machine learning and data mining algorithms (K-means, LDA, CNN, etc.) to realize the processing and analysis of enterprise massive data and feature extraction.

Based on the extracted enterprise features, considering the same classification situation, in order to reduce the overfitting problem caused by linear factors between features, the principal component analysis (PCA) and other algorithms are used to reduce the feature dimension, and the enterprise features are labeled to establish the label system of enterprise profiling.

Finally, on the basis of the label system of enterprise profiling, according to the specific business objectives, the enterprise profiling model is constructed, such as the enterprise business health indicator model, to achieve the profiling of the enterprise business health. In the evaluation of business health, it is necessary to determine the weight coefficient of health indicators at all levels, conduct non-dimensional data under the enterprise label, and finally calculate the health evaluation value according to the model and data. This paper uses the comprehensive evaluation method, based on the combination of entropy method and hierarchical analysis method, to analyze and determine the index weight coefficient.

2.2.4. Application of Data

Based on the enterprise profiling model, the enterprise data can be processed, analyzed and visualized, which can present the valuable information related to the enterprise operation and development. To realize the profiling of the business health status of the enterprise with the business health indicator model, various applications can be established to reflect the tax delinquency of the enterprise through the tax credit of the enterprise and whether there is credit risk; Through the enterprise development index to understand whether the enterprise is stable in the production process; Through the enterprise health index, we find out whether the enterprise's output value, employment and social security expenditure are healthy. Through the enterprise innovation index to know whether the annual production and research and development investment expenditure is at a low level; These applications can greatly help functional departments to understand and supervise the health status of the quality of enterprise development, formulate relevant enterprise policies, and then stabilize the economy and employment, and continuously improve the level of government management and service efficiency.

3. Literature References

Enterprise profiling is a user profiling of an enterprise. It is a model or concept that comprehensively describes the characteristics, behaviors and attributes of an enterprise through data and analysis. The application of this concept is intended to help companies better understand themselves and their competitors to optimize strategic decisions, marketing and operations management. Alan Cooper, the father of

interaction design, first proposed the concept of Persona [1], which is a virtual representative of real users. By labeling users with attributes and characteristics, Persona can be painted to achieve more accurate marketing and personalized recommendation for users and improve user experience. Davies et al. put forward the concept of A Corporate Character Scale [2], which can realize the evaluation of corporate brands and help enterprises enhance brand value by quantifying and labeling the indicators of corporate brand image. In recent years, with the steady development of the domestic economy, the number of enterprises and the data information accompanying the production and operation of enterprises are increasing rapidly, and the application of enterprise information display is also increasing. At the same time, with the emergence of big data technology, the application of enterprise profiling through various indexes is rising with the big data technology to extract enterprise characteristics and build an indicator system. Jinpin Index[3] Based on the standard index calculation model of enterprise brand communication influence and consumer word-of-mouth evaluation as the core, published the enterprise communication influence brand index for the first time, focusing on three dimensions of brand communication index, brand investment index and brand management index to describe the enterprise's communication influence. In the study of enterprise profiling based on big data platform, Tian Juan et al. [4] described the enterprise profiling in five dimensions and established the enterprise label model system. In particular, they used convolutional neural network (CNN) and other technologies to extract features for the dimensions, which greatly improved the efficiency and effect of feature processing. Wang Qingfeng et al. [5] used Markov logic reasoning in the research and implementation of enterprise profiling technology based on knowledge graph, and carried out enterprise profiling in the aspects of basic information, operation status, intellectual property rights, human structure and related prediction, providing reference information for the public and the government to understand the enterprise. Liu Yang [6] Based on the enterprise credit information service platform, used TextRank algorithm and other data mining algorithms to extract labels from enterprise profiling, designed a enterprise profiling system, and visualized the investment relationship between enterprises, the employment relationship between people and enterprises, and the shareholder relationship. Wu Zihang [7] used the expert evaluation method to screen indicators and the analytic hierarchy process to calculate the weights of indicators in the study of enterprise profiling on the enterprise management simulation platform and obtained the designated class target tag system of enterprise profiling, which improved the accuracy and validity of enterprise profiling. Wang Li et al. [8] mentioned that the enterprise profiling system is based on the enterprise as the main body. Through collecting and analyzing the information of the enterprise scale, business scope, company number, registered capital, financial status, recruitment status, latest business dynamics and other information, the enterprise profiling with different labels is constructed.

Government data refers to all kinds of data related to government generated or obtained by government agencies in the process of exercising their functions and powers. The main sources of government data are government departments, public institutions and enterprises and

institutions cooperating with the government. These data are mainly important information for the government to provide reference and decision-making support. In 2016, Dalian National Tax Bureau used "Internet +" and big data technology to collect the characteristics of enterprise management, integrity, growth, risk, contribution and habits, and supplemented by the application to display the basic information of enterprises, behavior patterns and analysis results visualization [9], making the monitoring of tax risks more accurate. And be more targeted in providing tax services. In 2018, Jiangsu Province's integrated financial Service Platform officially launched the "Enterprise Public Credit Information Report (Basic Version)", involving more than 100 data items, covering the public credit information of 59 provincial departments and 13 municipal governments with districts. From the aspects of basic information, association, business information, qualification certification and negative information, the public credit information of enterprises is depicted (reported) to help financial institutions connect with the financing needs of small and medium-sized enterprises, further tap high-quality and honest small and medium-sized enterprise customers, and effectively reduce the reliance of bank risk control on measures such as collateral (pledge) and guarantee. Wen Xiwei [10] used enterprise profiling technology in his study on the application of customs taxpayer management to establish enterprise profiling specimens for direct use in the territory according to industry, commodity and other identifiable standards. Through the analysis of enterprise characteristics and the grasp of dynamics, the purpose of scientific management of territorial taxpayers was achieved, accurate supervision and efficient service were realized, and administrative efficiency was improved.

Based on the enterprise information of big government data in Dongguan City, this paper uses big data technology to make enterprise portraits of the development health of enterprises in Dongguan City, so as to provide data reference for the government to supervise and provide efficient services.

4. Conclusion

Based on the partial enterprise data shared by government affairs across multiple functional departments, this paper studies an enterprise profiling technology based on government affairs big data. Through the implementation of key technology paths of big data, an enterprise profiling model and its application are constructed to finally reflect the health of enterprise economic operation, thus helping the public and the government to understand the comprehensive development quality of enterprises. Provide more powerful reference data for government supervision and service enterprises. At present, the research of enterprise profiling technology is still in the development stage of exploration and gradual application, then adjustment and re-application. With the development of "Internet + government", the government has accelerated the safe and orderly sharing and opening of government data, and the integrity of enterprise data has also been continuously improved. In addition, with the support of cloud computing, big data and artificial intelligence and other

technologies, the integration of dynamic data such as corporate news, Weibo and evaluation outside of government data, the model and application of enterprise profiling based on government big data will be further developed and deepened, and enterprise profiling will become more and more three-dimensional and accurate.

Acknowledgments

The school-level research project of Guangdong University of Science and Technology (GKY-2021KYYBK-18) in 2021 is "Research on an Enterprise Profiling Model Design Method Based on Government Big Data".

2022 Dongguan Social Development Science and Technology Top-level Project "Research and Design of an Emergency Supply Assurance Monitoring System in the Post-epidemic Context" (20221800903432).

This work was supported by Dongguan Science and Technology Special Envoy Project - 3C industry electronic circuit board solder spot CCD automatic detection and intelligent repair solder joints (Project No.: 20221800500692).

Thank you to my beloved and supportive family, whose love always reminds me to work harder to realize my dreams.

Thank you to my colleagues, who have given me invaluable help and many comments and sound advice during the process of writing my dissertation.

References

- [1] Alan Cooper, *The Road to Interactive Design: Bringing High-tech Products Back to Humanity* [M], Publishing House of Electronics Industry, 2006, pp. 116-127.
- [2] Davies G, Chun R, Da Silva R V, et al. *A Corporate Character Scale to Assess Employee and Customer Views of Organization Reputation*[J]. *Corporate Reputation Review*, 2004, 7(2):125-146. DOI:10.1057/palgrave.crr.1540216.
- [3] Chen Yiping. *Today's Product Index: Using Big Data to Profile Enterprises*[J]. *China Times*, 2015-07-06(021).
- [4] Tian Juan, Zhu Dingju, Yang Wenhan. *A review of research on enterprise portrait based on big data platform*[J]. *Computer Science*, 2018, 45(S2): 58-62.
- [5] Wang Qingfeng. *Research and implementation of enterprise portrait technology based on knowledge graph*[D]. Harbin Institute of Technology, 2019. DOI:10.27061/d.cnki.ghgdu.2019.000872.
- [6] Liu Yang. *Design and implementation of an enterprise portrait system*[D]. Hebei Normal University, 2019.
- [7] Wu Zihang. *Research on Enterprise Portrait of Enterprise Management Simulation Platform*[D]. Beijing Jiaotong University, 2022. DOI:10.26944/d.cnki.gbtfju.2022.001370.
- [8] Wang Li. *Optimization design of enterprise portrait system based on tag library*[J]. *Information and Computer*, 2020, 32(18):2.
- [9] Wang Lei. *Dalian State Taxation Bureau: Big data provides accurate portraits of enterprises*[N]. *China Taxation News*, 2016-10-17(A04).
- [10] Wen Xiwei. *Application of big data driven enterprise portrait in customs taxpayer management*[J]. *China Port Science and Technology*, 2020, (08): 37-40.