Design and Implementation of Visualization Application Platform of Enterprise Operation Data Based on R Language

He Zhang

85649133@qq.com

Guangdong Industry Polytechnic, Guangzhou, Guangdong, China

Abstract: This paper studies the design and implementation process of enterprise operation data visualization application platform based on language. In this system, ggplot2 in R language tool is used as the graphic system for making and displaying statistical charts, and javaweb technology is used to develop application programs. R language and Java middleware choose Rserve 1.8.5. Enterprise operation data storage uses big data technology to build hadoop cluster to collect, clean, calculate and save data. In the data classification function module of the system, Bayesian algorithm is used to realize the scientific classification of data. The design of the application system can help the enterprise operators to automatically generate visual reports by selecting the data, which is convenient for users to consult the reference operation data and optimize the operation decisions in time, thus effectively improving the operation quality of the enterprise.

Keywords: R Language, Javaweb, Data Visualization, Big Data, Enterprise Operation.

1 INTRODUCTION

At present, under the background of data explosion caused by the all-round popularization of network equipment, the work of enterprise operation has been inseparable from the evidence support of effective data. This makes enterprises need to continuously increase investment in the collection, analysis and storage of big data. Especially in the operation of enterprises, it is necessary to accurately budget and adjust the operation decisions of enterprises according to the data. Secondly, the person in charge of the operation needs to query and browse the operation data, so as to grasp the overall situation of the enterprise's production and operation, and it is easy to directly attack the weak links of enterprise management to locate the problems, analyze the reasons, realize the improvement of the operation decision, promote the continuous improvement of the enterprise, and form a closed loop of enterprise operation monitoring application. Finally, more market data can be mined through the enterprise operation data, which is convenient for enterprises to find more business opportunities. Data visualization application can help enterprises to provide more efficient guidance. Therefore, this paper studies and designs the visualization application platform of enterprise operation data based on R language. [8]

BDEDM 2023, January 06-08, Changsha, People's Republic of China

DOI 10.4108/eai.6-1-2023.2330365

2 KEY TECHNOLOGIES

2.1 Javaweb

Javaweb refers to the category of application systems developed and designed by applying Java language. Java application uses B/S structure, and its architecture usually consists of four categories, namely servlet or JSP scripting language, HTML page, Java class controlled by business logic and other additional resources. Servlet component is the most representative of servlet interface implementation class. However, JSP components, usually in some specific conditions, are converted into servlet by HTML documents with JAVA programs, such as when clients request JSP files. The definition of Java class refers to the class defined by developers related to web application development programs. Additional resources refer to some static documents and client class files. Static files mostly refer to HTML files stored in the server file system. Client files generally refer to the qpplet applet and some basic configuration files such as web.xml that run by the client. [1]

2.2 R Language

R is not only a language environment composed of toolkits for data operation, calculation and display, but also has a powerful statistical analysis function. The R language environment contains various statistical analysis and related function models of drawing, which can not only meet the related needs of users' statistical calculation, but also realize the related functions of graphic analysis of data visualization. Besides, R is free and open source software, and it supports multiple platforms, such as Windows, Linux and UNIX. The function of R language can be enhanced according to the expansion package, so as to realize the related functions of quickly manipulating data call, input and output. Users can also use R language to change the source code of the application by expanding packages and plug-ins, so as to realize functions such as extended function loop and self-definition. Through the use of extension packages and plug-ins, R language can even conduct interactive real-time dynamic data analysis through other software, which greatly improves the compatibility and expansibility of R language. [3]

The commonly used charts of data visualization in R language include scatter chart, histogram, histogram, bar chart, box chart, area chart, hotspot chart and relationship chart. The functions used in R language to make charts are ggplot (), geom_histogram, geom_point (), geom_boxplot (), etc. Therefore, R language is used in the visualization application platform of enterprise operation data studied in this paper to process data images and related algorithms. Through R language, intuitive and effective graph list drawing can be automatically generated according to the effective data in the data cluster, and the generated results can be presented to users. The realization process of data visualization is shown in Figure 1. After the data is cleaned and saved by hadoop, the R language and mapreduce are used to construct the drawing function to construct the visual image, which is then presented to the user page of the system through javaweb. [7]

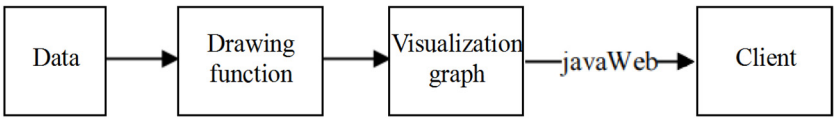

Figure 1: Data processing module diagram

2.3 Rserve

Rserve is an open source data processing middleware. The visualization application platform of enterprise operation data studied in this paper not only uses R language to process and analyze data, but also uses javaweb to deliver data results to users in the form of application page display. Therefore, Reserve is chosen to connect the two technologies. Because Reserve supports TCP/IP protocol of data and network transmission between R language and other languages, it also supports the use of multiple development languages. In this way, the statistical modeling, data analysis and visualization functions of R language can be displayed in javaweb applications through the remote connection of Reserve. [9]

2.4 Development Environment

The development environment of enterprise data visualization application platform is divided into two parts, one is the setting environment of data cluster, the other is the development environment of javaweb application design. The establishment scale of big data cluster, taking into account the data scale of medium-sized enterprises, uses five physical machines to build a big data cluster. In addition to the local management system server of the enterprise, the data comes from a large number of external data of relevant marketing markets obtained by using scrapy and crawler technology. Big data cluster is divided into one primary node and four secondary nodes, and hadoop ecological components are installed on five physical machines to clean, analyze and save big data. The main components of hadoop installed include HDFS distributed storage, mapreduce distributed computing, yarn resource scheduling and hive data warehouse. [6]

The operating system of the physical machine in big data cluster and application design is 64-bit Windows 10, and the CPU is configured as Intel core i7. Rserve, the middleware of R language and Java, chooses version 1.8.5. In the R language tool, ggplot2 is used as the graphic system for making and displaying statistical graphs. The installation package instruction to load ggplot is install.packages("ggplot2"); library (ggplot2). The development language of the application is Java, the development environment is jdl1.7, Apache tomcat 7.54 is selected for the server construction, and MySQL 7.6 is installed on the database server. The framework is SSM framework of spring+springmvc +mybaties. In order to ensure the subsequent development and secondary development of the application system, the relevant configuration of the enterprise operation data visualization application platform is stored in the form of configuration files. The system can obtain the MySQL data connection of JDBC by connecting the database configuration files and reading the key value. The configuration file code of the system is shown in Figure 2.

dbconfig.properties

```
# Database configuration file
hibernate.dialect=org.hibernate.dialect.MySQLDialect
validationQuery.sqlserver=SELECT 1
jdbc.url.jeecg=jdbc:mysql://10.50.0.22:3306/dscaodemo?useUnicode=true&characterEncoding=UTF-8
jdbc.username.jeecg=root
jdbc.password.jeecg=root
jdbc.dbType=mysql

hibernate.hbm2ddl.auto=update
```

Figure 2: Database configuration file code

3 REQUIREMENTS ANALYSIS

3.1 Functional Requirements

The user groups of R-language-based visualization application platform for enterprise operation data are mainly related personnel in charge of enterprise operation. The main functions of the system include data query, data classification, data analysis and data visualization report. Users can make the system automatically generate visual reports by selecting data, which is convenient for users to consult the reference operation data and adjust the operation decision in time. Among them, operation data includes debt repayment ability data, operation ability data, profitability data and financial data analysis. Among them, the operational capacity data mainly includes accounts receivable data and inventory turnover data. According to the relevant industry data captured by big data, the system can also generate a line chart comparing the enterprise with the industry average data. [5]

3.2 Overall Design

As shown in Figure 3, the overall design framework of the enterprise data visualization application platform can be divided into R language data visualization system and javaweb application subsystem. The two subsystems are connected by Rserve middleware, and the communication protocol is TCP/IP. R language subsystem needs to provide data conversion interface and graphics rendering interface. The function of the data conversion interface is to convert the processed data stored and collected by hadoop into a data category recognizable by R language and capable of data processing. And the image rendering interface is used to convert the images generated by R language tools into information forms recognizable by javaweb. The javaweb application part is mainly responsible for the realization of the main functions of the system. This subsystem can be divided into business department and management system. Among them, the business system is a functional module that generates lists and forms for data presentation. The main function of the management system is to realize the permission setting of the system and the setting of user information management module. As a complete application system, its R&D is designed with traditional B/S architecture, so it can't be separated from the technical support of database server. The database server of this system adopts relational database MySQL. [2]

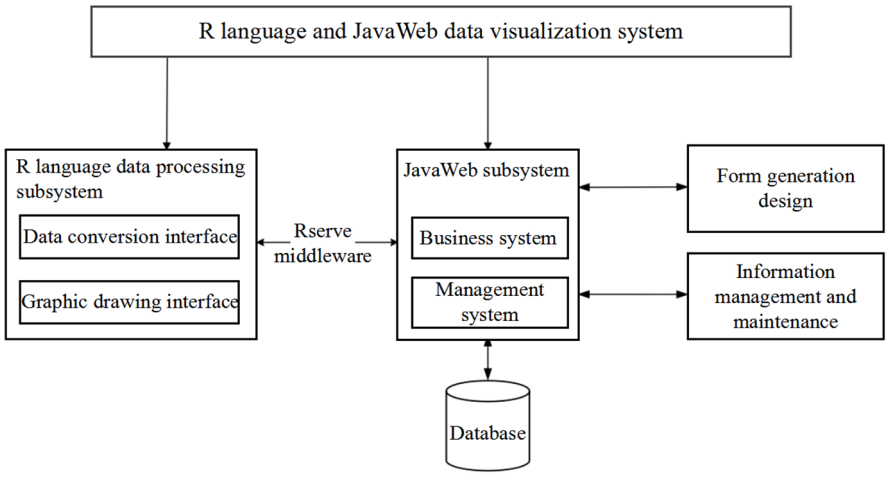

Figure 3: Overall system architecture diagram

4 FUNCTIONAL IMPLEMENTATION

The main functions of the system include data query, data classification, data analysis and data visualization report. Data classification is a very important link in data processing. This paper uses Bayesian classification algorithm to classify data. The principle of Bayesian algorithm is to first determine that an individual variable is assumed to belong to a random classification, and then calculate the probability that this variable is correct in this classification. Then calculate the probability values of all the classifications in this variable, and compare these probability values to get the best classification of this variable. In the calculation process, we determine the probability value by calculating the ECM misjudgment loss value. The calculation formula of ECM is shown in Formula 1, where P_1 and P_2 respectively represent the correct probability that a variable belongs to X1 or X2. After that, the minimized ECM value needs to be calculated by Bayesian algorithm. This process is shown in Formula 2, and then the comparison results of the probabilities of all categories are obtained. [4]

$$ECM = (R1, R2) = L(1/2) * P(1|2) * P_1 + L(1/2) * P(1|2) * P_2 \quad (1)$$

$$R = \{x| \frac{f_1(x)}{f_2(x)} > \frac{L(1|2)}{L(2|1)} * \frac{P_2}{P_1}\} \quad (2)$$

Take Company A as an example. After logging in to the system, the operation leader of Company A can select the data required for data reporting, such as the total market value data of the industry as of August 30, 2022. The system will summarize the collected public data of enterprises in the same industry, and draw the market value data chart of Company A in the same industry. The circular chart version drawn is shown in Figure 4. It can be seen from the figure that the industry valuation of Company A has reached 49.5 billion yuan, reaching the fourth position in the industry, which is higher than the industry average of 10.2 billion yuan

and the industry median of 387 million yuan. The important index of IPO is the market value. According to the change of market value, operators can judge the overall market positioning of the company on the basis of industry, and then analyze the advantages and disadvantages of the company from multiple dimensions of profit and development, so as to achieve the purpose of deepening the company's operation data. [10]

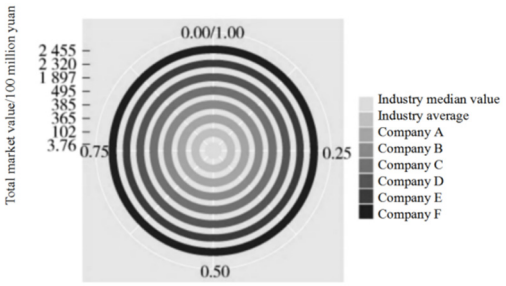

Figure 4: Ring chart of market value data of company A in the same industry generated by the system

This system mainly evaluates the operation of a company through debt repayment ability X and profitability Y indicators. Taking Company A as an example, the initial characteristic value of debt repayment ability is calculated to be 5.075, with a variance of 49.63%, while the initial characteristic value of profitability reaches 3.859, with a variance of 36.58%, with a total of 86.21%. The variance contribution rate of these two indicators exceeds 85%, which proves that these two ability indicators X and Y can represent the company's operating level. The final evaluation formula is shown in Formula 3.

$$V = \frac{\sigma 1}{\sigma} * X + \frac{\sigma 2}{\sigma} * Y(\sigma = \sigma 1 + \sigma 2) \quad (3)$$

The asset-liability ratio of Company A, which represents the debt repayment ability, is 0.871, the EPS of profitability is 0.824, the net profit score is 0.654, and the return on assets is 0.618. Therefore, the final debt repayment ability is 1.56, and the profitability score is 0.76, among which the net asset income chart of Company A shows the results as shown in Figure 5. What is reflected behind the return on net assets data is the ratio efficiency of shareholders' equity to its own capital, which can often reflect the relationship between a company's profit and the owner's interests. From the chart of return on net assets of Company A generated by R language technology in the system, we can easily see that the return on net assets of Company A has fallen below the average level of the industry, which means that the management has encountered bottlenecks in recent years. Therefore, although Company A's profitability is high, the profit source is mostly the sale of equity, so the overall profitability of Company A is not high, and operators need to pay attention to adjustment. The above data shows that Company A has low profitability but high debt repayment ability, and it still has good operation ability and good development prospects.

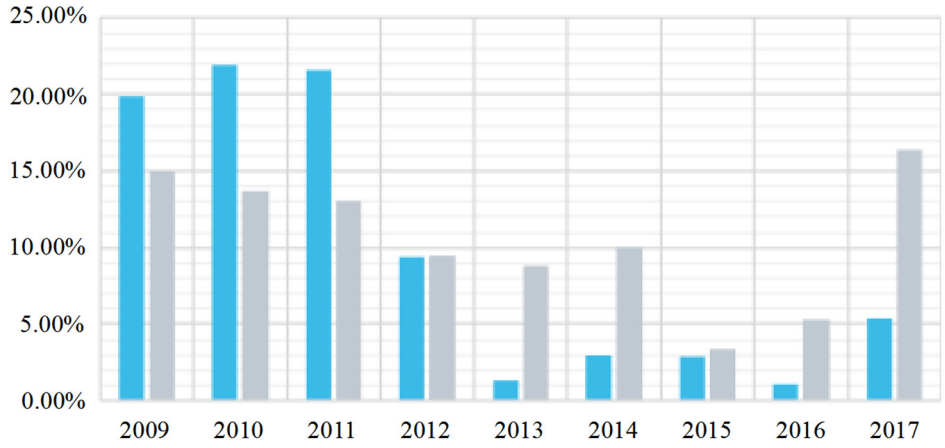

Figure 5: Comparison of average return on equity between a company and industry

5 CONCLUSIONS

This paper provides a good reference for the research of enterprise operation data visualization platform based on R language. However, due to the author's limited ability and time, this paper still lacks more in-depth research and deepening of functions. It is hoped that the follow-up can be improved by relevant experts. The client of the enterprise operation data visualization platform studied in this paper only supports PC browser login, which is slightly insufficient in the current era of popularization of mobile devices. In the future, we will try to study the related content of mobile data processing and further improve the R language operation data visualization algorithm. Nowadays, web crawler technology has developed to the stage where it can generate visual graphics. However, the use of web crawler technology in this paper is limited to text documents, and it needs to be improved.

REFERENCES

[1] Cao Dongsheng. (2017) Research and Implementation of Data Visualization Method Based on web. Xidian University.06.

[2] Gu Chenggong. (2020) Analysis on the Architecture and Implementation Technology of Telecom Big Data Operation Platform. Digital Communication World.02.

[3] Li Miao. (2021) Problems and Innovative Countermeasures of Enterprise Operation and Management Mode in the Era of Big Data. Enterprise Reform and Management.11.

[4] Li Tianci. (2019) Research on the Application of Visualization Method of Financial Data Based on R Language. Heilongjiang University.03.

[5] Li Xiangkun. (2022) Design and Implementation of Internet Data Mining and Analysis Platform. Technology Wind.04.

[6] Wang Wenjie, Kang Yumei. (2022) Analysis and Development of Visual Application of R Language Big Data Audit. Finacial Management Reaserch.07.

[7] Wang Yitai. (2020) Research on the Enabling Application of Big Data to Enterprise Operation. Communication Design and Application.07.
[8] Wu Dongmei. (2017) Application Research on How to Realize Data Visualization in Computer R Language Teaching. Digital Space.05.
[9] Wu Mingxin. (2017) Research on the Application of Financial Big Data Audit Based on R Language. Nanjing Audit University.05.
[10] Xian Jinqian. (2020) Design and Implementation of Enterprise Big Data Operation Management and Control System. Guangdong University of Technology.06.