

Research on Business Data Analysis and Mining based on Deep Learning

Xiaoyan Gao

School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang), Suzhou 215100, China

1994049542@qq.com

Abstract. At present, the database system can efficiently realize the functions of data entry, query and statistics, but it can't find the rules of relationships in the database, and can't predict the future development trend according to the existing data. Using data mining technology, we can effectively find the regularity behind a large number of data, provide the hidden knowledge and means in the data, and eliminate the phenomenon of data "explosion but lack of knowledge". Based on DL (Deep Learning), this paper studies the analysis and mining of commercial data. In order to better assist investors to evaluate and make decisions on financial data, it puts forward the need to build a reliable and effective financial data forecasting model. In the original data obtained, it is necessary to interpolate by certain methods in data cleaning, and some repeated data will be duplicated. This paper analyzes and makes full use of these massive commercial information through DL, finds out the potential laws of business and market operation, and finds some unknown and hidden useful information from it, which helps enterprises to grasp the development direction of the industry and sales market from a macro perspective and better serve the vast number of consumers.

Keywords: Deep Learning; Business Data Analysis; Excavate.

1. Introduction

A lot of transaction data accumulated in the field of commercial circulation contains a lot of knowledge that is beneficial to business decision-making. It is difficult to extract this knowledge by traditional analysis methods. Using DL, accurate and timely information can be obtained. Based on the data warehouse of enterprises, decision-makers make decisions on the professional knowledge of planners through online analysis and processing [1]. At present, the database system can efficiently realize the functions of data entry, query, statistics, etc., but it can't find the rules of the relationship existing in the database, and can't predict the future development trend according to the existing data. Using DL can effectively find the regularity behind a large number of data, provide the hidden knowledge and means in the data, and eliminate the phenomenon of data "explosion but lack of knowledge" [2]. In order to better assist investors to evaluate and make decisions on financial data, this paper puts forward the need to build a reliable and effective financial data forecasting model, and on the basis of financial data analysis, integrates DL algorithm to analyze financial data [3]. The real decision-making based on data analysis is mainly concentrated in the banking, insurance, telecommunications and e-commerce industries, and has not yet expanded to all areas of operation management. Enterprise decision-making units cannot effectively use the existing information, and even make decision-making behavior chaotic. If useful information and knowledge can be mined from a huge database through DL for decision support, enterprises will have a certain degree of competitive advantage [4]. At present, most mainstream financial data analysis systems need to pay for intelligent diagnosis. As an important part of stock data, investors set different data, such as stock price, index, length and financial ratio, which can judge industries, intelligently compare hot spots, monitor news, etc. Users have different needs for diagnosis and analysis, and determine diagnostic data and conditions. This paper uses the massive data accumulated by enterprises for a long time to realize four applications: customer classification and characteristic analysis, marketing strategy analysis, operating cost and income analysis, fraud analysis and prevention. These hidden patterns and relationships can be found through DL [5]. By analyzing and making full use of these massive

commercial information through DL, we can find the potential laws of business and market operation, and find some unknown and hidden useful information, which can help enterprises grasp the development direction of the industry and sales market from a macro perspective and better serve the vast number of consumers [6].

2. Application Value and Advantages

2.1. Application Value of Data Analysis and Mining Models

Data mining, as a popular data processing method in recent years, is a method that can extract hidden, unknown, and potentially valuable information from databases. Usually, data mining includes two aspects: statistical analysis and knowledge discovery. Among them, statistical analysis mainly includes correlation analysis, regression analysis, and discriminant analysis.

At present, some large enterprises have invested tens or even billions of dollars in information construction, such as using SAP, ERP, and CRM for data analysis. However, data between departments has not been connected, and platform data cannot be shared, unable to support business needs in a timely manner. With the intensification of market competition, consumer demand is becoming increasingly high, posing higher requirements for business managers to aggregate dispersed data and efficiently analyze and mine these data. The relevant information submitted to the remote server, such as the user's personal information, the product information that the user wants to order, and some questions and requirements raised by the user. However, many users are unwilling to disclose their real personal information, so they will not fill in the registration form truthfully, which will result in the low quality of the original data of data mining [7]. In this case, it is necessary to infer the user's background information from the browser's browsing information and then utilize it. The financial data analysis system also analyzes and mines representative stocks in corresponding sectors based on different data lists, such as historical trends and K-line trends of stocks. In order to better understand the information of individual stocks, financial analysis is conducted by comparing the obtained historical data according to different types or examining their corresponding market information in detail [8-9]. The application of data analysis and mining models can help enterprises solve the above problems, achieve systematic analysis of multi-dimensional data, timely detect anomalies and propose improvement plans, and help enterprises achieve intelligent decision support. Data mining is an ability to extract hidden, valuable, and decision-making guidance from a data warehouse. It can uncover unknown patterns and information, which is of great significance for the study of practical problems. The original data types include structured data, semi-structured data and heterogeneous data. The method of discovering knowledge can be either mathematical or non mathematical; It can be deductive or inductive.

2.2. Advantages of Business Data Platforms

You only need to access the source data of the data source, and you can do multidimensional analysis of the data in any dimension without any preprocessing. You can gain business insight through the data simply by dragging and dropping, and people without any technical foundation can analyze the data by themselves. The original records obtained from data sources are not only huge in data, but also may contain a lot of noise data, redundant data, sparse data or incomplete data, so it is very difficult to mine them directly [10]. In fact, data preparation plays a vital role in the final success of data mining and whether it has economic benefits. Data preprocessing mainly includes data cleaning, integration, selection and transformation. Each candidate set will be visited once, and the data generated before will not be used, resulting in most of the information being discarded. Therefore, opening up new storage space, accessing the database once and recording relevant information and pruning by using the properties of frequent itemsets can reduce the number of candidate itemsets and speed up the generation of frequent itemsets. Hundreds of data sources can be accessed with one click, processed, analyzed and mined, and presented visually by multiple terminals, which can manage the

data in the whole value chain. Build a complete data server cluster, provide strong and stable data computing ability, and respond to hundreds of millions of rows of data in one second.

3. Business Data Analysis and Mining based on Deep Learning

3.1. Business Data Analysis and Mining

Keras is a Python project initiated by the open source community, based on TensorFlow and the highly modular neural network library running on Theano. Keras supports the back-end calculation of TensorFlow and Theano framework. The most important feature of Keras framework is that a variety of efficient encapsulation functions simplify the programming of deep neural network models and improve the efficiency for the expansion of neural networks. We compare and analyze many kinds of machine learning and DL frameworks at present, and choose TensorFlow and Keras as the basic framework of the system data analysis platform, and SCiKit-LEAM as the algorithm framework of machine learning. Now, due to the realization of business automation in various industries, a large number of business data are produced in the commercial field, which are no longer collected for the purpose of analysis, and the analysis of these data is no longer simply for the needs of research, but mainly to provide truly valuable information for business decision-making and then gain profits. By importing business data such as enterprise production data, sales data, customer data and financial data into BDP business data platform, all kinds of data are stored, managed, modeled, calculated and analyzed, and then presented to enterprises that need data in the form of visual charts for analysis and management decisions. In the aspect of data processing and other algorithms that have certain requirements for performance, the front-end and back-end are developed separately, data analysis is carried out, and the analysis results are fed back to the front-end page, and the current system is accessed and stored by using an open source database management system. Conduct commodity analysis, business monitoring and analysis, member data analysis, KPI assessment of various departments, cost control, sales forecast, and decomposition of solution data analysis index system, as shown in Figure 1.

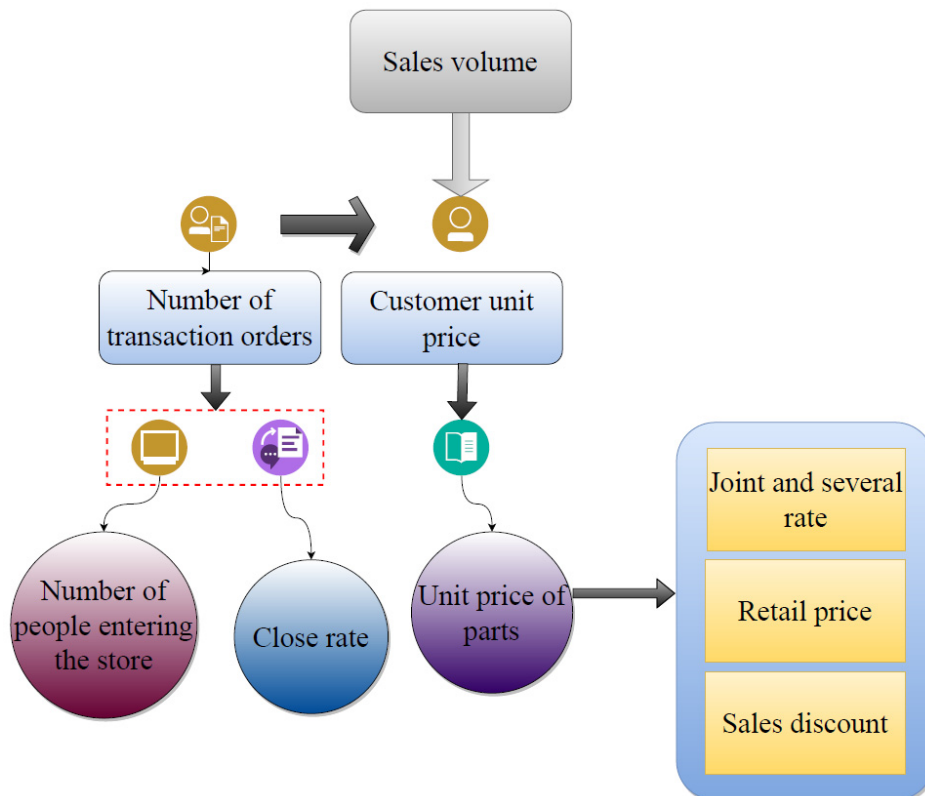


Figure 1. Decomposition of Data Analysis Indicator System

Collect customer needs, then extract customer data sources, import them into the business data platform, establish data analysis and mining models on the business data platform, design calculation logic, and automatically run calculations on the platform. Finally, the platform outputs visual charts and analysis results to guide commercial enterprises in formulating strategic plans and conducting corresponding management.

3.2. Data Analysis Module

This paper constructs a deep neural network and trains to learn the rules between historical transaction data, and then selects data from a certain period of time to predict it based on the generated model. TuShare is one of the main sources of data acquisition, selecting stocks within a specific time period to obtain data on their trading days, such as historical highs, lows, opening prices, closing prices, trading volume, and other data. In the data analysis stage, analyze stock trading data and textual information based on the corresponding analysis theme. Mainly from two aspects of fundamental analysis and technical analysis, the data is analyzed and explained from different dimensions, including prices, indicators, trends, and industries, and the statistical results are presented in intuitive visual charts. The collection of analytical data is not comprehensive enough and lacks indirect data; Timely identify, lock in, and adjust current issues; A comprehensive real-time analysis is required for analyzing the inventory structure of goods, in order to guide delivery and distribution; Use BDP machine learning Analysis of algorithms to analyze the commodity joint rate, transaction rate, customer unit price, etc. In order to ensure the characteristics of financial data time series, it is necessary to discretize or normalize specific data. Table 1 shows the functional use case description of the collection module.

Table 1. List of Collection Module Functions

Number	
Performer	Data acquisition module
Basic Path	The obtain stock list is compared with that list in the database.
Business Rules	According to the filtering conditions, the corresponding data types are collected, updated synchronously and saved in the database.
Non Functional requirement	According to the title, content, time and other key features, the full-text search is carried out according to the items.

In the original data obtained, we often encounter incomplete data that are not updated after being suspended from the market. In data cleaning, we need to interpolate by certain methods, and some duplicate data will be duplicated. In the process of data analysis, different dimensions of data need to be analyzed, and each analysis type needs different types of data, so it is necessary to extract data according to the actual analysis type. Multi-dimensional analysis of customer volume, year-on-year, month-on-month, comparison between new and old stores, etc. Analysis of sales promotion; Commodity analysis: commodity classification analysis, profit contribution of various commodities, cross analysis of commodities and members; Member analysis: data analysis model of member system, analysis of new members, distribution of members and consumption habits of members.

4. Conclusion

With the rapid growth of data volume, most enterprises are facing challenges such as high user volume and concurrency, rapid changes in business requirements, poor response speed, shortage of data technology personnel, and opaque information in various departments. Problems such as data system paralysis and inconsistent data caliber are constantly emerging. This paper studies the analysis and mining of business data based on DL. The original records obtained in the data source not only have a huge amount of data, but also may have a large number of noise data, redundant data, sparse

data or incomplete data. It is very difficult to directly mine on them. In fact, whether data mining ultimately succeeds or not, and whether it has economic benefits, data preparation plays a crucial role. Data preprocessing mainly includes data cleaning, integration, selection, and transformation. Extract valuable business knowledge from the target data, analyze and validate the results to ensure their reliability and practicality. Through this design and development work, I learned the relevant knowledge of DL in the financial field, realized the importance of standardized software development and detailed operations, fully researched the analysis requirements of the system, and conducted overall and detailed functional design for it. On the basis of implementing and testing the functional modules, it is also necessary to continuously cultivate one's ability to learn new technologies and knowledge, in order to improve oneself and enhance technology.

References

- [1] Luo Y, Yao C, Mo Y, et al. A creative approach to understanding the hidden information within the business data using Deep Learning[J]. Information Processing & Management, 2021, 58(5):102615-102630.
- [2] Bai S, Li M, Lu Q, et al. A new measuring method of dredging concentration based on hybrid ensemble deep learning technique[J]. Measurement, 2022(188-):188.DOI:10.1016/j.measurement.2021, 33(10):26-34.
- [3] Schmitt M A. Deep Learning in Business Analytics: A Clash of Expectations and Reality[J]. Papers, 2022, 53 (16): 19-32.
- [4] Li L. Cross-Border E-Commerce Intelligent Information Recommendation System Based on Deep Learning[J]. Computational intelligence and neuroscience, 2022, 42(16):36-44.
- [5] Kraus M, Feuerriegel S, Oztekin A. Deep learning in business analytics and operations research: Models, applications and managerial implications[J]. Neural computing & applications, 2021, 19(8):19-24.
- [6] Fombellida, J. Martin-Rubio, I. Torres-Alegre, S. Andina, D. Tackling business intelligence with bioinspired deep learning[J]. Neural computing & applications, 2020, 32(17):26-34.
- [7] Li M. High performance deep learning techniques for big data analytics[J]. Concurrency and Computation Practice and Experience, 2022, 30(9):25-31.
- [8] Liang M. Optimization of Quantitative Financial Data Analysis System Based on Deep Learning[J]. Complexity, 2021, 2021(1):1-11.
- [9] Zhoujian C, He W, Jianyang Z, et al. Initial study on the application of deep learning to the Gravitational Wave data analysis[J]. Journal of Henan Normal University(Natural Science Edition), 2021, 16(5):11-19.
- [10] Zhijing T, Jinghao S, Zhizheng W, et al. Accurate marketing based on big data analysis[J]. Wireless Internet Technology, 2022, 23(12):20-36.