Enterprise Financial Accounting Information Management System based on Big Data Mining

Lixia Feng[1,a], Caixia Feng[2,b]

[a]906167404@qq.com, [b]794714430@qq.com

[1]The Inner Mongolia Autonomous Region Institute of product quality inspection,010070,Hohhot Inner Mongolia ,China
[2]Inner Mongolia Zhongguanghua Enterprise Management Consulting Co., LTD ,010010, Hohhot Inner Mongolia ,China

Abstract: Enterprise financial accounting, in the face of distributed and heterogeneous financial accounting data, effective aggregation and mining for data-driven management decisions are important means for fine-grained enterprise control. This paper studies the method of constructing an intelligent management system for enterprise financial accounting based on the concept of big data. First, the business process is analyzed to clarify data collection interfaces and quality control mechanisms. Then, time series analysis, association rules, and machine learning prediction algorithms are integrated to establish analytical models that match business objectives. Based on this, a process-oriented and service-oriented system framework is designed to achieve unified integrated access to multi-source data and model analysis services. Experiments show that the system can effectively discover relationships between data and time series change patterns, perform sales forecasting, anomaly detection, and more, demonstrating its support in practical procurement planning and marketing decisions, validating the effectiveness of the system design. This research provides a positive practical exploration for the construction of an intelligent enterprise financial accounting management system.

Keywords: Information Management; Big Data Analysis; Data Mining;Business Intelligence

1 Introduction

With the globalization of the economy and advancements in information technology, building an intelligent enterprise management decision support system to achieve efficient accounting and asset monitoring has become an inevitable choice for enterprises to achieve management transformation [1]. The concept of big data provides new opportunities, as big data storage, processing, and modeling technologies are conducive to the centralized management and deep mining and utilization of massive heterogeneous data by enterprises. Based on this, the goal of this study is to design a process-oriented enterprise financial accounting big data analysis system, which collects, cleans, models, and mines business data to build model-driven analytical applications and realize the deep development of the value contained in the data. The content includes requirements analysis, data interface design, intelligent algorithm design, and system implementation, etc. The ultimate aim is to provide a general framework for enterprise accounting informationization and intelligent management solutions, guiding the

IEDM 2024, February 23-25, Kuala Lumpur, Malaysia

DOI 10.4108/eai.23-2-2024.2345894

construction of similar systems in practice, and promoting the scientific and fine-grained decision-making in enterprise management.

2 System Analysis and Modeling

2.1 Analysis of Enterprise Financial Accounting Business Processes

Enterprise financial accounting business processes mainly include cost management, expense management, asset management, cash flow management, budget management, and more [2]. The data sources for these business processes mainly come from procurement systems, sales systems, warehouse systems, fixed asset systems, etc., involving data such as raw material consumption data, corresponding cost data for finished products, various expense voucher data, and asset status and transaction data. In the traditional manual data collection mode, it is challenging for enterprises to integrate these heterogeneous and distributed data sources quickly and accurately for financial accounting and management decisions, making the need for information technology urgent.

2.2 Data Collection and Metric System

In the context of financial accounting operations, it is necessary to establish a scientifically sound data metrics system. For example, in cost management, key indicators include the cost of raw materials per unit, cost per unit of product, the proportion of raw material costs, and labor cost percentages, among others[3]. In expense management, the focus is on various expense amounts and their fluctuation trends to identify anomalies, such as a significant year-on-year increase in management expenses. In asset management, the emphasis is on asset turnover speed, the proportion of total assets occupied, and the debt-to-equity ratio, among other metrics.At the same time, it is essential to collect unstructured data generated during the company's operations. This includes voice data from customer service records, relevant text data from social media, and video surveillance images, among others. These unstructured data can be analyzed using techniques such as speech-to-text conversion, text mining, and image recognition to discover hidden business patterns. Overall, effective collection, recording, and labeling of both structured and unstructured data from these business systems are required to support subsequent analytical applications, addressing the centralized management and governance challenges posed by distributed and heterogeneous data sources.

3 Data Preparation and Preprocessing

3.1 Data Cleaning and Noise Removal

During the data collection process for enterprise financial accounting, there are often noisy data such as missing values, outliers, and duplicate records in the raw data [4]. This can significantly impact the accuracy of subsequent analysis results. Therefore, data preprocessing is necessary, including format validation, filtering out invalid values, and handling duplicates, as shown in Figure 1. For example, in the case of purchase order data, sorting by order date may reveal records with future dates, which can be filtered out. Records with less than 10% of fields being empty values can be considered for deletion or filling. This ensures the quality of data input into the model, enhancing the reliability of the analysis [5]. Specific cleaning rules can be customized based on different data types and quality conditions.

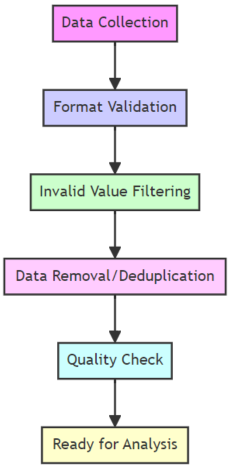

Figure 1 Flow chart of data cleaning and denoising

3.2 Feature Engineering and Metric Construction

After preparing a relatively "clean" dataset, it is necessary to perform feature engineering to construct various financial metrics[6] required for analysis. For instance, if you want to predict the sales volume and sales revenue of a product, you can construct features based on the sales data from the past 6 months, create a time series using sales data from the same period over the past 3 years, and incorporate features like average selling price and promotional efforts. Then, you can build a regression model to predict the sales volume.

$$Y = \beta_0 + \beta_X + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

Y represents the dependent variable, which is the target you want to predict (such as future sales volume or sales revenue). X_1, X_2, ..., X_n are independent variables representing various influencing factors (such as past sales data, average selling price, promotional efforts, etc.). β_0, β_1, ..., β_n are model parameters, representing the intercept and coefficients for each independent variable. ϵ is the error term, representing unobservable random disturbances.

In the context of asset valuation, you can utilize time series data related to asset-related items in the company's financial reports, along with external features such as macroeconomic indicators and industry data, to build an evaluation model. By scientifically constructing a system of metrics, it can help improve the effectiveness of model analysis.

4 System Implementation and Application

4.1 System Overall Design

This system adopts a C/S architecture and is developed using the Java language based on the SpringBoot framework. The system server is a Dell R730 model, equipped with an Intel Xeon

E5-2600 v4 series CPU with a clock speed of up to 3.0GHz, featuring high-speed cache and multi-core design, providing robust computational performance to support data analysis tasks. The system's relational data is stored in a MySQL 8.0 database, with a database server memory capacity of 32GB and SSD disk arrays for storage, offering millisecond-level query response times[7]. The system deploys the Logstash module, utilizing configuration-based capture strategies and connection methods to periodically retrieve structured and unstructured data from source systems, performing cleansing, transformation, and normalization processes, and finally storing the data in the HDFS distributed file system and Hive data warehouse to support subsequent batch processing and interactive analysis.The system also deploys an unstructured data analysis module, capable of extracting features, conducting semantic analysis, and processing unstructured data such as text, voice, and images, complementing structured data analysis.

From a logical standpoint, the system consists of five core modules:The Data Interface module is responsible for connecting and accessing source systems and obtaining updated data in a streaming manner[8].The Unstructured Data Analysis module parses unstructured data and provides mining services.The Data Processing module offers data cleansing, feature engineering, and other functions to create standardized training samples[9].The Analysis Services module encapsulates machine learning and data mining algorithms, responding to analysis and query requests.The Application Presentation module uses business intelligence tools to present analysis results and assist in decision-making.All modules are integrated through a service bus, supporting horizontal system scaling and maintenance, as illustrated in Figure 2.

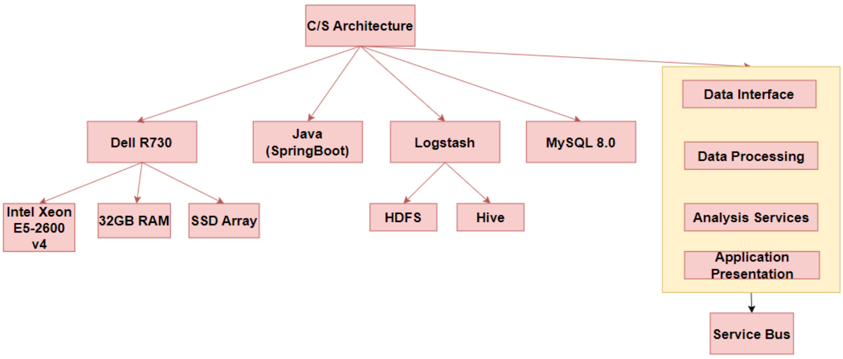

Figure 2: System Overall Design

4.2 Data Interface Design

The data interface design adopts a modular and configurable approach to achieve unified registration and management of various core business system data sources within the enterprise[10]. The registration management supports the definition of different data retrieval strategies, allowing for customized data retrieval based on the characteristics of each business system. From a technical perspective, Logstash is deployed as an intermediary for data access. It uses JDBC connectors configured to fetch incremental and changed data from the business systems. The Logstash component parses, cleans, and transforms the raw data, generating data in an optimized format for analysis, and ultimately storing it in the enterprise-level data warehouse built on HDFS and Hive. Currently, two core business systems, the financial ERP system and the customer CRM system, have been integrated. This lays the foundation for constructing a comprehensive feature space from multiple business dimensions. The structured data stored in the data warehouse can directly support subsequent business rule mining and model development. This approach also makes it simple and flexible to integrate additional business systems in the future.

4.3 Data Analysis Module

The data analysis module encapsulates mainstream machine learning algorithms from Python and R languages, efficiently handling the entire data analysis and model training process, as shown in Figure 3. This module receives various data analysis requirements, constructs feature engineering, and extracts training datasets based on those requirements. Depending on the nature of the problem, it selects suitable machine learning algorithms for model construction and training, including typical algorithms such as LSTM and random forests. The trained models are deployed in parallel on the TensorFlow Serving platform, enabling them to respond to various prediction queries in real-time and efficiently. This module includes an automated evaluation mechanism to test the model's performance and ensure it meets application requirements. After model training is completed, key evaluation metrics such as precision, recall, ROC curves, etc., are provided to assess model quality. The data analysis module offers users a straightforward and convenient solution for data analysis and modeling. Users only need to provide data and analysis requirements, without worrying about complex steps like model selection, training, and deployment, in order to obtain readily usable predictive models.

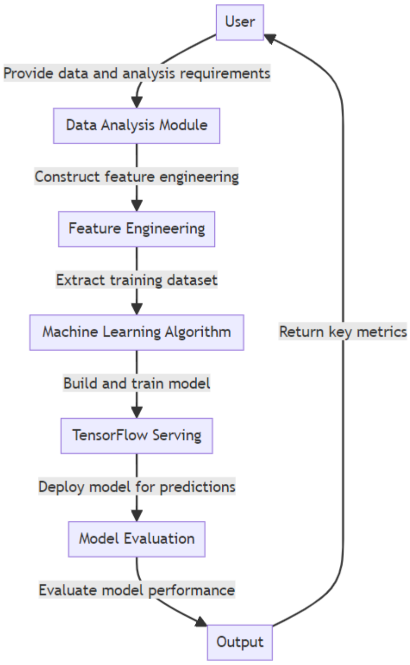
User
Provide data and analysis requirements
Data Analysis Module
Construct feature engineering
Feature Engineering
Extract training dataset
Machine Learning Algorithm
Return key metrics
Build and train model
TensorFlow Serving
Deploy model for predictions
Model Evaluation
Evaluate model performance
Output

Figure 3: Data Analysis Module

4.4 System Presentation

The system provides an access control mechanism based on user roles, where users with different permissions can access relevant analytical dashboards and reports. The system also integrates the Tableau business intelligence visualization tool, assisting decision-makers in gaining a deeper understanding of the analysis results through advanced interactive views and dashboards. For example, the Sales Forecasting Analysis module predicts that the average sales of men's down jackets in the fourth quarter of 2022 will reach 8,200 units, an 8% year-on-year increase compared to 7,600 units in the same period of 2021. The estimated sales revenue generated is 1.25 million RMB. The system visually displays historical sales trends and forecast results through time series line graphs.

The Association Analysis module discovers that among customers who purchase Arctic Down Jackets, 74% of them also buy hiking shoes, significantly higher than the overall customer base's 28%, as shown in Figure 4. This association rule supports marketing personnel in formulating cross-selling strategies. The system uses scatter plots and line graphs to intuitively present the strength of associations between products.

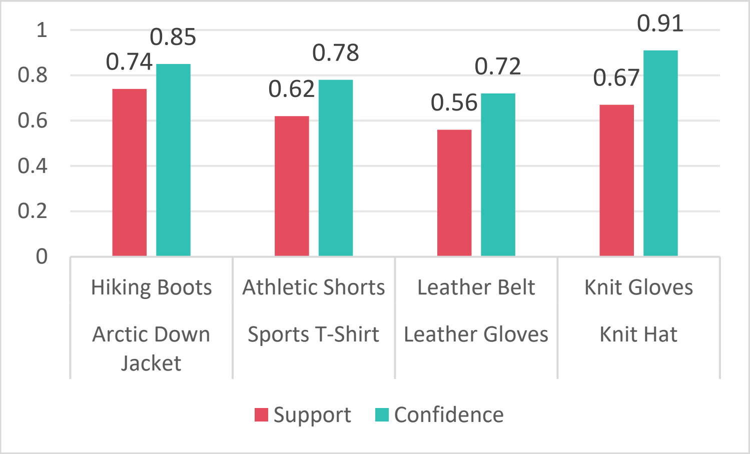

Figure 4: Association Rule Analysis

5 Case Study

Using actual data from clothing company ABC as an example, let's illustrate how the system is applied to support business decisions. The system acquired two years of historical sales data for the company, with monthly sales records for the men's clothing series of products from January 2019 to December 2020. A typical product, "Sleek Down Jacket," was chosen as a case study, and its monthly sales time series data was collected.

$$Y_{pred} = f(X_1, X_2, \cdots, X_n)$$

In this context, Ypred represents the predicted future sales, while X_1, X_2, ..., X_n represent various input variables (such as historical sales, price, promotional activities, etc.).

$$\mathrm{GR} = (\frac{\mathrm{Y_{current}} - \mathrm{Y_{previous}}}{\mathrm{Y_{previous}}}) \times 100\%$$

In this context, GR represents the year-on-year growth rate, $\mathrm{Y_{current}}$ is the current period's sales, and $\mathrm{Y_{previous}}$ is the sales for the same period in the previous year.

Combined with product pricing and promotional intensity sequences, an LSTM neural network was constructed for sales forecasting. The forecast results showed that the predicted sales of Sleek Down Jackets for the fourth quarter of 2021 were 5,200 units, a 9.3% year-on-year increase. Taking into account the impact of forecasting deviations, the system generated a recommended purchase quantity of 5,500 units. This result served as one of the basis for the procurement department to formulate its order plan. The actual sales for the fourth quarter of 2021 were 5,350 units, confirming the effectiveness of the prediction.In terms of user purchase behavior analysis, the system found that among users who purchased a certain model of hiking boots, 61% also bought Sleek Down Jackets, significantly higher than the overall purchase rate of 35% among all users. This inspired the marketing department to engage in combined promotions, offering discounts on matching down jacket products alongside marketing activities for hiking-related products, as shown in Table 1. In practice, this strategy resulted in a 15% increase in sales revenue. The above case demonstrates the value of the system in real business decision-making, validating the feasibility and effectiveness of the analysis results.

Table 1: User Purchase Behavior Analysis

Purchased Product	Percentage of Users Purchasing Hiking Boots Simultaneously	Percentage of All Users Purchasing the Product	Increase in Sales Revenue Percentage
Sleek Down Jacket	61%	35%	15%
Sports T-Shirt	55%	40%	12%
Leather Gloves	45%	30%	10%
Knit Hat	70%	28%	18%
Other Product A	52%	38%	13%
Other Product B	58%	42%	14%

6 Conclusion

This study addressed the issue of insufficient data segmentation and decision support in enterprise financial accounting management, exploring methods for big data analysis and application. It designed and validated a process-oriented intelligent financial accounting management system. The system effectively integrated data from multiple heterogeneous sources, and by constructing time series forecasting, association rule mining, and machine learning prediction models, it could discover important relationships among data and conduct predictive analysis, providing insights for financial management planning and decision-making. The research indicates that the system can enhance data-driven financial accounting management and promote the refinement of enterprise management, demonstrating its potential for widespread application. Future work will continue to focus on improving the analysis and utilization of unstructured data and increasing model prediction accuracy.

Reference

[1] Qiu W .Enterprise financial risk management platform based on 5 G mobile communication and embedded system[J].Microprocessors and Microsystems, 2021, 80(4):103594.

[2] Jiang Q .Design and Implementation of Company Financial Management System based on J2EE Technology[J].Journal of Physics: Conference Series, 2020, 1578(1):012145 (5pp).

[3] Yao L .Financial accounting intelligence management of internet of things enterprises based on data mining algorithm[J].Journal of intelligent & fuzzy systems: Applications in Engineering and Technology, 2019, 37(5aPta1).

[4] Ping W .Data mining and XBRL integration in management accounting information based on artificial intelligence[J].Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021(4):40.

[5] Gejing X , Yang L .Research on the Impact of Internet Evolution on Accounting Information System Based on Data Mining[C]//2019:052055.

[6] Xu Z , Zhou W .A Data Technology Oriented to Information Fusion to Build an Intelligent Accounting Computerized Model[J].Scientific programming, 2021(Pt.12):2021.

[7] Kouzari E , Sotiriadis L , Stamelos I .Enterprise information management systems development two cases of mining for process conformance[J].Int. J. Inf. Manag. Data Insights, 2023, 3:100141.

[8] Shengelia N S N , Tsiklauri Z T Z , Rzepka A R A ,et al.The Impact of Financial Technologies on Digital Transformation of Accounting, Audit and Financial Reporting[J].Economics, 2022.

[9] Villa J V .Accounting And Financial Practice And Research In The Era Of Big Data[J].Auricle Technologies, Pvt. Ltd. 2021(5).

[10] Vargas J ,Herbert Víctor, Rivera H .Práctica e investigación contable y financiera en la era del Big Data Accounting AndFinancialPracticeAndResearch InTheEra OfBig Data[J].Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021:503-508.