INTELLIGENT CLASSIFICATION OF FINANCIAL DATA AND INTELLIGENT INDUCTION OF EFFECTIVE DATA IN MACHINE LEARNING

XiongZhongBo

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Beliatskaya T.N. – Doctor in Economics, Associate Professor

Annotation. Discribed the developed system of intelligent classification of financial data and intelligent induction of effective data in machine learning based on big data technologies, text processing and math statistics methods

Keywords. Text classification; logical regression; financial data, natural language

The significance of economic data in the era of big data is very high. In the past 20 years, large-scale data growth has occurred in various fields.

Including healthcare and scientific sensors, user-generated data, Internet and financial companies, supply chain systems, etc. The International Data Corporation (IDC) reported that the total amount of data created and copied globally in 2011 was 1.8 1 B (1 Z B ≈ 1021B), which has increased nearly 9 times in just 5 years, and this number is expected Will double at least every two years.

In order to make reasonable use of the value brought by these data, we need to effectively classify the data and give value to various fields.

Common methods for screening big data, except manual screening, computer language screening, and artificial intelligence screening until fully automated.

The screened data can be used for the value of internal human cost of the enterprise, the standardization of management, the internal risks, the transaction risks between the enterprise and the enterprise, the assessment of the housing market price trend, and a new method for the police to investigate criminals, for medical analysis to find effective treatments, and prevention.

Business process of intelligent classification of financial data include the following steps.

First, survey users' demand for financial data and divide them into several categories

Second, the financial data classification system is designed and screened based on logistic regression algorithm.

Thirdly, the feature extraction system, feature weighting and logistic regression algorithm are applied to the system.

Fourthly, the system is applied to financial data classification, and the classifier is improved.

The traditional website classifies the data into culture, economy, military, entertainment, etc., so we subdivide the financial field into several

Table 1 classification of financial data

Quotation	Characteristic	Research
Shanghai and Shenzhen Stock Markets	economic data	Industrial economy
Hongkong stock market	money flow	Macro research
US stock market	IPO data	market research
Global index	Investment banking industry	Comments on people's livelihood
Fund market	Annual report and quarterly report	In depth report
Futures Options	Information hot spot	research report

For the automatic classification of news, announcement and securities consultation, the system trains the sample data through the logistic regression algorithm to get the classifier, and processes the data through the classifier.

After the original data, text preprocessing, feature selection, feature weighting, and special vector, the model is obtained based on the algorithm, and the final model file is obtained by testing. The new input data is obtained by the classifier to get the vector, and then the probability value is compared.

Objectives of the classification system:

(1) The historical financial data is obtained from the database, and the historical financial data is divided into training samples and test samples. The training samples are trained to generate classification models.

(2) After preprocessing the test samples, the classification effect of the classification model is tested, and the training parameters of the model are adjusted according to the effect. When the performance of the classification model tends to be stable, the final classification model is obtained

(3) It is applied to real-time classification to realize the preprocessing of new data, calculation of features, and prediction of the returned results by classification model.

Developed system extracts the sample file from the sample database. After data preprocessing, the sample file is divided into training sample and test sample. The training sample is used to train the classifier, and the test sample is used to evaluate the classification performance of the classifier. According to the evaluation results, the training process of the classifier is adjusted accordingly to get the final classifier. The classifier can be used for data classification in financial data real-time database, in which the data in financial data real-time database is the result of real-time crawling of web pages. The real-time data is classified by the classifier and finally stored in the classification result database for the display of classification results.

References

1. Yang Y M, Pederson JO. A Comparative Study on Feature Selection in Text Categorization [A] //Proceedings of the 14th international Conference on Machine learning[C], Nashville Morgan Kaufmann 1997:412－420

2. Yan Yan. "Research on Text Representation and Classification Method Based on Deep Learning" University of Science and Technology Beijing, 2016.

3. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

5. Beliatskaya, T. Intellectual capital in digital economy: education dimension / T. Beliatskaya, V. Knyazkova // Mechanisms of interaction between competitiveness and innovation in modern international economic relations : in 4 vol. / ISMA Univ. ; red. M. Bezpartochnyi. – Riga, 2017. – Vol. 1. – P. 84–93.

6. Beliatskaya, T. N. The methodology of e-economy infrastructure research / T. N. Beliatskaya, V. S. Knyazkova // Managing service, education and knowledge management in the knowledge economic era : proc. of the Annu. Intern. conf. on management a. technology in knowledge, service, tourism & hospitality 2016 (SERVE 2016), 8–9 Oct. 2016 a. 20–21 Oct. 2016, Jakarta & Vladimir State Univ., Vladimir / ed.: F. L. Gaol, N. Filimonova, F. Hutagalung. – Boca Raton, 2017. – P. 95–100.с.