

Application of Data Mining in the Field of Human Resource Management: A Review

Xin Ma^{1, a}, Ying Zhang^{2, b}, Yanjie Song^{1, c}, Erbao Wang^{3, d}, Feng Yao^{1, e} and Zhongshan Zhang^{1, f, *}

¹ College of Systems Engineering, National University of Defense Technology, Changsha 410073, China.

² College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410073, China.

³ Security Center of Human Resources, Beijing 100034, China

^a Maxin1995Oct@163.com, ^b zy2826938789@163.com, ^c songyanjie17@nudt.edu.cn, ^d 632148046@qq.com, ^e yaofeng@nudt.edu.cn, ^{f, *} zszhang@nudt.edu.cn

Abstract. Data mining methods play an important role in customer management, manufacturing management, and financial management. This paper focuses on the latest research progress from 2012 to the present. The analysis of the progress of existing research will help researchers to better engage in research work in the future. The literature review is classified according to the research field and the data mining methods involved. In each of the research areas, we conduct research classification based on the data mining tools used. Through literature analysis, it can be seen that the degree of attention of neural networks has increased continuously in recent years.

Keywords: Data Mining; Human Resource Management; Review.

1. Introduction

Data mining is an effective means of obtaining useful information by identifying data features. Human resources in the enterprise generate various data, which can be used in the formulation of corporate strategies and the selection of employees. At present, the application and application of data mining methods for human resource management (HRM) can be divided into customer management, manufacturing management, and financial management. Customer management includes the management of specific customers and relationships with them, including Customer Identification, Customer Attraction, Customer Retention and Customer Development [1, 2]. Manufacturing management manages raw materials, design processes, design processes, and product quality and failure conditions involved in the manufacturing process, including Quality control, Job shop scheduling, Fault diagnosis, Condition based monitoring, Manufacturing process, Defect Analysis, Manufacturing system [3]. Financial management focuses on financial fraud detection (FFD) and it can be divided into Bank fraud, Insurance fraud, Securities, commodities fraud, and other financial violations [4].

Because of the variety of research areas and data, researchers can use many kinds of data mining methods. Data mining methods for human resource management are mainly classification methods, cluster analysis, association analysis, regression analysis, and neural networks. Through literature review, we analyze the application of data mining methods in the field of HR, provide reliable and effective decision support for enterprises, and promote the development of enterprises.

This article consists of three parts. The first part introduces the basic situation of data mining methods in HRM. After that, a literature review will be given in the second part. Finally, the third part of this paper will summarize the research in this paper and give the future development trend.

2. Literature Review of the Articles

The literature review is based on the research field of data mining research on customer management, manufacturing management, and financial management. This classification method can

select effective methods based on problem characteristics and provide researchers with guidance for conducting relevant research.

2.1 Literature Review in the Field of Customer Management

Classification method [5] used a method of supervised classification to try to apply in the CRM field. [6] proposed a classification framework for Partial Focus Feature Reduction. In order to solve the problem of noise and imbalance of classification methods, a classification method combining feature selection is proposed [7]. [8] used a variety of classification algorithms to classify customers in health sector operating companies, assist in business decisions and promote sales activities, based on product groups and risk factors.

Clustering analysis [9] transformed the linguistic data sequence into a fuzzy data sequence, and finally derives the fuzzy equivalence relation. A clustering algorithm is developed by using this relationship. [10] used a two-stage cluster analysis method to improve efficiency through secondary clustering. [11] used a two-layer clustering analysis algorithm to extract the characteristics and preferences of mobile phone users. [12] analyzed the selection of the starting position and the determination of the optimal number of clusters in the CRM analysis using K-means clustering.

Regression analysis [13] considered the variable correlation of sequences and proposes a variable length regression model. [14] used a cluster analysis method based on logistic regression to improve the accuracy of customer churn prediction. [15] combined logistic regression methods and decision trees to predict the loss of users in the telecommunications industry. [16] used regression methods to analyze the impact of organizational culture factors on hotel customers.

Neural Network [17] used the BP neural network model in the e-commerce customer management system, which is better than RFCA. In order to improve the convergence speed of neural networks, the Legendre wavelet neural network model was constructed to improve the convergence speed while ensuring simplicity and accuracy [18]. [19] used artificial neural networks to manage customer impressions after data segmentation and classification. [20] proposed a recurrent neural networks (RNNs) based on the client loyalty number (CLN), recency, frequency, and monetary (RFM) variables.

2.2 Literature Review in the Field of Manufacturing Management

Classification method [21] used the decision tree or neural network to determine the inventory strategy of the manufacturing enterprise after classifying according to the inventory style. [22] used the Analytic Hierarchy Process (AHP) and the Supply Chain Operations Reference (SCOR) to develop an inventory strategy for fresh vegetables. [23] developed a risk framework for suppliers and proposed four criteria for the classification framework. [24] combined the characteristics of maintenance and logistics to propose a multi-standard classification method.

Clustering analysis [25] used fuzzy clustering methods to evaluate oilfield companies' reservoir management units (RMUs). [26] used information entropy with improved clustering algorithm to process high-dimensional data and consistent data in discrete manufacturing quality management systems. [27] provided a management recommendation based on a hybrid latent class model for important risk factors in the milk manufacturing industry. [28] proposed a fuzzy clustering method for manufacturing lean management. Then, taking the specific enterprise as an example, the usage requirements of the method are introduced, and the effectiveness of the method is verified.

Regression analysis [29] used a combination of regression and neural network methods to analyze the improved possibilities of thin film transistor-liquid crystal displays (TFT-LCD). [30] compared the effects of ANN, SVM and multiple logistic regression (MLR) in the fault diagnosis of rolling element bearings. The multi-logistic regression effect is superior to the other two methods. In order to monitor the reliability of wind turbines, many methods such as Fast Fourier Transform (FFT), Hilbert Yellow Transform (HHT), feature extraction, and logistic regression (LR) were used [31].

Neural Network Job Shop Scheduling is a difficult issue in the manufacturing sector and an important part of research. [32] proposed a Hopfield Neural Network (HNN) algorithm to solve the shop scheduling problem. [33] studied the rescheduling problem of semiconductor manufacturing

systems and proposed a fuzzy neural network model. [34] designed a neural network model that predicts the impact of knowledge management on manufacturing performance. [35] combines a deep confidence network with a top-level regression layer to form a deep neural network.

2.3 Literature Review in the Field of Financial Management

Classification method [36] considered the generalized beta distributions for non-negative random variables and the generalized skew-Student distribution for random variables distributed on the real line, and establishes two parameterized classification tree methods to achieve financial risk management. The companies in the China Economic Research Database are classified into two categories by classification [37]. [38] established a Bayesian network model to classify bad debts in hospitals. [39] compared the performance of Weighted Support Vector Machines (SVM), Naive Bayes (NB) and K-Nearest Neighbors classifiers for financial statement fraud, the weighted SVM method is more effective.

Clustering analysis [40] proposed the average linkage hierarchical clustering algorithm for the analysis of financial indicator networks. [41] used data envelopment analysis (DEA) and axiomatic fuzzy set (AFS) clustering to assess online banking performance. [42] used gray clustering methods to assess financial risk. [43] used clustering analysis method to achieve the scale reduction of financial variables and non-financial variables as early warning indicators.

Regression analysis [44] used regression methods to analyze the relationship between working capital management (WCM), working capital policy (WCP), and business profitability (PFT). [45] considered the nonlinear expectation, a sub-linear expectation nonlinear regression model is constructed to identify the financial risks faced. Several parameter estimations and model predictions were suggested, and the asymptotic normality of the estimation

and the mini-max property of the prediction are obtained. [46] used fuzzy linear regression to improve the original earnings management model. The probability of a high-level manager's flow through logistic regression is positively correlated with the company's size and negatively correlated with ownership concentration [47].

Neural Network [48, 49] developed a self-organizing artificial neural network model combined with statistical tools to monitor fraud in financial statements. [50] established an artificial neural network model to predict the possibility of exchange rate changes. Compared with the time series model, the artificial neural network model has a better prediction effect. [51] uses a probabilistic neural network model to predict the risk of credit defaults. Six real-world credit data sets have been applied for effectiveness and feasibility purposes.

3. Conclusion

Data mining methods are well applied in all areas of human resource management. Overall, data mining has a good prospect in the field of HRM. The use of data mining tools has a positive effect on supporting management and policy development. A trend in research is the combination of data mining methods and some optimization methods to improve the accuracy of classification and prediction. At the same time, the combination of different data mining tools is also a content that can be studied in more depth. In addition, in the data mining method, artificial neural network, deep neural network, belief network and other tools show a rapid development trend, which is consistent with other research in the field of artificial intelligence, and may become a hot research field in future research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 71690233, 61473301.

References

- [1]. Ling, R., & Yen, D. C. 2001. Customer relationship management: An analysis framework and implementation strategies. *Journal of Computer Information Systems*, 41, 82–97.
- [2]. Kracklauer, A. H., Mills, D. Q., & Seifert, D. 2004. Customer management as the origin of collaborative customer relationship management. *Collaborative Customer Relationship Management- taking CRM to the next level*, 3–6.
- [3]. Choudhary, Alok & Harding, Jenny & Tiwari, Manoj. 2009. Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*. 20. 501-521. DOI: 10.1007/s10845-008-0145-x.
- [4]. Ngai E W T, Hu Y, Wong Y H, et al. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559-569.
- [5]. Lessmann S., Voß S. 2008 Supervised Classification for Decision Support in Customer Relationship Management. In: Bortfeldt A., Homberger J., Kopfer H., Pankratz G., Strangmeier R. (eds) *Intelligent Decision Support*. Gabler.
- [6]. Tu Y, Yang Z. 2013. An enhanced Customer Relationship Management classification framework with Partial Focus Feature Reduction. *Expert Systems with Applications*, 40(6):2137-2146.
- [7]. Natchiar S U, Baulkani S. 2015. Customer relationship management classification using data mining techniques. *International Conference on Science Engineering and Management Research*. IEEE, 2015:1-5.
- [8]. Alsaç A, Çolak M, Keskin G A. 2017. An integrated customer relationship management and Data Mining framework for customer classification and risk analysis in health sector. *International Conference on Industrial Technology and Management*. IEEE, 2017:41-46.
- [9]. Singh A B, Singh T B. 2012. Fuzzy max- Δ equivalence relation-based clustering method for customer relationship management. *Adv Fuzzy Sets Syst*, (2):97-113.
- [10]. Jayaraj V, Rajakumar M, J. Lavanya J L, et al. 2015. Enhancing Customer Relationship Management by Integrating Customer and Product Values based on Clustering and Indexing Techniques. *Thrombosis & Haemostasis*, 115(2).
- [11]. Chang C I, Ho J C. 2017. Two-layer Clustering Model for Mobile Customer Analysis. *It Professional*, 19(3):38-44.
- [12]. Syakur M A, Khotimah B K, Rochman E M S, et al. 2018. Integration K-Means Clustering Method and Elbow Method for Identification of The Best Customer Profile Cluster. *336(1):012017*.
- [13]. Miguéis V L, Poel D V D, Camanho A S, et al. 2012. Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, 39(12):11250-11256.
- [14]. Lu N, Lin H, Lu J, et al. 2014. A Customer Churn Prediction Model in Telecom Industry Using Boosting. *IEEE Transactions on Industrial Informatics*, 10(2):1659-1665.
- [15]. Dalvi P K, Khandge S K, Deomore A, et al. 2016. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. *Colossal Data Analysis and NETWORKING*. IEEE.
- [16]. Rahimi R. 2017. Organizational Culture and Customer Relationship Management: A Simple Linear Regression Analysis. *Journal of Hospitality Marketing & Management*, 2017:1-7.

- [17]. Zhou M L, Hao J J. 2012. Development of Customer Relationship Management System in E-Commerce Based on BP Neural Network. 6-7:995-999.
- [18]. Liu X. A new algorithm for electronic customer relationship management. 2012. International Conference on Computing and Convergence Technology. IEEE, 2012:359-362.
- [19]. Ogwueleka F N, Misra S, Colomo-Palacios R, et al. 2015. Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank. *Human Factors in Ergonomics & Manufacturing*, 25(1):28-42.
- [20]. Salehinejad H, Rahnamayan S. 2017. Customer shopping pattern prediction: A recurrent neural network approach. *Computational Intelligence*. IEEE, 2017:1-6.
- [21]. Xue D J, Gao T Y, Mu G Y, et al. 2013. Research of Parts Inventory Classification and Management Strategy in Manufacturing Enterprises. *Applied Mechanics & Materials*, 433-435:2359-2362.
- [22]. Guritno A D, Fujianti R, Kusumasari D. 2015. Assessment of the Supply Chain Factors and Classification of Inventory Management in Suppliers' Level of Fresh Vegetables. *Agriculture & Agricultural Science Procedia*, 3:51-55.
- [23]. Phusavat K, Anussornnitisarn P, Pongrakhananon T, et al. 2015. Applications of benchmarking and classification framework for supplier risk management. *Benchmarking*, 22 (2): 275 -289.
- [24]. Teixeira C, Lopes I, Figueiredo M. 2018. Classification methodology for spare parts management combining maintenance and logistics perspectives. (4):1-20.
- [25]. Li A Q, Chang Y R, Kong X H. 2012. Production performance appraisal rating for reservoir management units based on fuzzy clustering. Hindawi Publishing Corp.
- [26]. Wang P F, Shu H P, Zheng J L, et al. Application of Evolution of Clustering in Discrete Manufacturing Quality Management. *Journal of Sichuan University of Science & Engineering*, 2013.
- [27]. Tremblay M, Hess J P, Christenson B M, et al. 2016. Customized recommendations for production management clusters of North American automatic milking systems. *Journal of Dairy Science*, 99(7):5671-5680.
- [28]. Niu Z, Yang F, Jing S. 2017. A Study of the Classification of Lean Management Tools in Manufacturing Enterprises Based on Fuzzy Clustering Analysis. *Industrial Engineering Journal*.
- [29]. Hsieh K L. 2010. Yield management in TFT-LCD manufacturing by using regression and neural network techniques. *International Journal of Manufacturing Technology & Management*, 20(1):300-315.
- [30]. Pandya D H, Upadhyay S H, Harsha S P. 2014. Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. *Soft Computing*, 18(2):255-266.
- [31]. Bodla M K, Malik S M, Rasheed M T, et al. 2016. Logistic regression and feature extraction-based fault diagnosis of main bearing of wind turbines. *Industrial Electronics and Applications*. IEEE, 2016:1628-1633.
- [32]. Fnaiech N, Hammami H, Yahyaoui A, et al. 2012. New Hopfield Neural Network for joint Job Shop Scheduling of production and maintenance. *IECON 2012 -, Conference on IEEE Industrial Electronics Society*. IEEE, 2012:5535-5541.

- [33]. Zhang J, Qin W, Wu L H, et al. 2014. Fuzzy neural network-based rescheduling decision mechanism for semiconductor manufacturing. *Computers in Industry*, 65(8):1115-1125.
- [34]. Li P T, Wong K Y. 2017. A Neural Network Approach for Predicting Manufacturing Performance using Knowledge Management Metrics. *Cybernetics & Systems*, 48(1):1-17.
- [35]. Bai Y, Li C, Sun Z, et al. 2017. Deep neural network for manufacturing quality prediction. *Prognostics and System Health Management Conference*. IEEE, 2017:1-5.
- [36]. Adcock C J, Meade N. 2016. Using parametric classification trees for model selection with applications to financial risk management. *European Journal of Operational Research*, 259(2).
- [37]. Pan W T. 2012. The use of genetic programming for the construction of a financial management model in an enterprise. *Applied Intelligence*, 36(2):271-279.
- [38]. Shi D, Zurada J, Guan J. 2016. A Bayesian Network Approach to Classifying Bad Debt in Hospitals. 2016:3298-3307.
- [39]. Moepya S O, Akhoury S S, Nelwamondo F V. 2015. Applying Cost-Sensitive Classification for Financial Fraud Detection under High Class-Imbalance. *IEEE International Conference on Data Mining Workshop*. IEEE, 2015:183-192.
- [40]. Kumar S, Deo N. 2012. Correlation and network analysis of global financial indices. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 86(2):026101.
- [41]. Tao L, Liu X, Chen Y. 2013. Online banking performance evaluation using data envelopment analysis and axiomatic fuzzy set clustering. *Quality & Quantity*, 47(2):1259-1273.
- [42]. Zhu R N, Yin H H, Management S O. 2014. Assessment of the financial risk listed companies based on the grey clustering. *Science-Technology and Management*.
- [43]. Bao X Z, Tao Q Y, Hong-Yu F U, et al. 2015. A Study of Financial Distress Warning based on R-variance Clustering and COX Proportional Hazards Model. *Journal of Systems & Management*.
- [44]. Temtime Z T. 2016. Relationship between Working Capital Management, Policies, and Profitability of Small Manufacturing Firms. *Dissertations & Theses - Gradworks*.
- [45]. Song Y, Lin L. 2013. Sublinear Expectation Nonlinear Regression for the Financial Risk Measurement and Management. *Discrete Dynamics in Nature and Society*, (1):937-959.
- [46]. Höglund H. 2013. Fuzzy linear regression-based detection of earnings management. *Expert Systems with Applications*, 40(15):6166-6172.
- [47]. Rađen D, Stanisic N. 2017. Financial Distress and Managerial Turnover: The Case of the Republic of Serbia. *Social Science Electronic Publishing*, preprint (6).
- [48]. Fanning K M, Cogger K O. 2015. Neural network detection of management fraud using published financial data. *Intelligent Systems in Accounting Finance & Management*, 7(1):21-41.
- [49]. Fanning K, Cogger K, Srivastava R. 2015. Detection of management fraud: a neural network approach. *Intelligent Systems in Accounting Finance & Management*, 4(2):113-126.
- [50]. Chaudhuri T D, Ghosh I. 2016. Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework. *Papers*.
- [51]. Abedin M Z, Chi G, Colombage S, et al. 2018. Credit Default Prediction Using a Support Vector Machine and a Probabilistic Neural Network. *Social Science Electronic Publishing*.