Disruptive Human Resource Management
S. Patnaik (Ed.)

doi:10.3233/ATDE240424

Optimization Method for Human Resource Decision Based on Hadoop Big Data Platform

Xueqi Zhang [a1]
[a] *Shandong Management University, Jinan, Shandong, 250357, China*
ORCiD ID: Xueqi Zhang https://orcid.org/0009-0003-6502-4193

Abstract. When enterprises face a large amount of enterprise data, how to effectively utilize a large amount of information to better manage talent has become a key issue that urgently needs to be solved. Therefore, this article proposes a human resource decision-making optimization method based on the Hadoop big data platform. By taking steps such as data collection, feature extraction, model construction, and decision optimization, the scientificity and accuracy of enterprise human resource management can be improved. This article compares the performance of the algorithm used with machine learning (ML) and deep learning (DL) algorithms, and the results indicate that the F1 value of the method designed in this article is 95.8%, which is higher than 84.7% of DL and 77.9% of ML. Experiments have shown that this method can effectively enhance the accuracy and efficiency of human resource decision-making, providing strong support for human resource management in enterprises.

Keywords. Hadoop Technology; Big Data Platform; Human Resource Management; Decision Optimization; Distributed Framework

1. Introduction

As the scale of enterprises continues to expand and market competition intensifies, how to apply big data to the human resource management of enterprises is an urgent problem to be solved. Therefore, how to effectively carry out human resource management in enterprises is the key to selecting and retaining talents, but traditional data processing methods are difficult to cope with the massive amount of human resource data and cannot effectively meet the optimization needs of human resource management in enterprises. Therefore, this article proposes a human resource decision optimization method on the basis of the Hadoop big data platform, and studies human resource decision optimization methods for large-scale data environments.

Resource management is not only the key to the development of enterprises, but also the key to enhancing their competitiveness. Wahyoedi S optimized group collaboration using multi-agent enhanced learning methods and extracted features from team members using deep neural networks. Experiments showed that his proposed algorithm could effectively improve the efficiency of group cooperation [1]. Putra R

[1] Corresponding Author: Xueqi Zhang, qishukejing777@163.com.

used simulation and optimization methods to establish and solve human resource decision-making problems, thereby achieving the optimal human resource allocation and performance evaluation strategy. His experiment showed that the method proposed in this article could effectively improve the effectiveness of human resource management decision-making [2]. Pourkhodabakhsh N used a multi-criteria decision-making technique to identify factors affecting employee turnover and used a best-worst approach to assess these variables [3]. Based on the Java programming language, Chen S combined the spring model-view-controller web application system with a B/S (browser/server) framework to construct and implement an intelligent enterprise human resource management information system [4]. Wang X used a recruitment algorithm that combines transfer learning and multi-objective optimization to capture the characteristics of candidates using pre-trained neural networks, and then used multi-objective algorithms to select the best candidate. Compared with traditional recruitment methods, this algorithm had significantly improved both recruitment efficiency and candidate quality [5]. These scholars mentioned above have not used big data processing platforms, and when they need to process data, there could be limitations. Often, in order for enterprises to make the right decisions, the human resource department needs to process a massive amount of data.

Hadoop is a distributed computing platform typically used to process large-scale data, and many scholars have used this method for human resource decision-making optimization. Mahmood Q U A collected cross-sectional data from 189 respondents working in different organizations in Pakistan, and his research results revealed a significant and positive impact of all green human resource practices on performance [6]. Bonner R L constructed a new recruitment model. He used Hadoop to analyze a large amount of employment data and used neural networks to predict the performance of job applicants [7]. Varma D constructed a talent recruitment model driven by Hadoop big data. This model extracted features from massive resumes and employee's information, and then used ML methods to predict the performance of applicants [8]. Berk L applied Hadoop in the analysis of a large amount of employment data and utilized integrated learning algorithms to predict job candidate's performance [9]. While using the Hadoop big data platform for human resource decision-making optimization, it also faces challenges such as data security risks, real-time limitations, and customization requirements.

Based on the research of these scholars, this article proposes an efficient human resource management method for massive data based on Hadoop. Then this article conducts a detailed study on the Hadoop based big data platform, and effectively improves the human resource management decision-making of enterprises through the collection, storage, processing and analysis of massive data. This article also constructs a human resource decision-making optimization model based on Hadoop by collecting data and feature extraction, and finally, through experiments, an example analysis is conducted on the proposed model. The results suggest that the model proposed in the article is effective and feasible.

2. Optimization Method for Human Resource Decision-making Based on Hadoop

2.1. Data Collection

This article uses the MapReduce component in Hadoop to collect receipts distributed from the public human resources database LinkedIn, and obtains relevant data (as shown in Table 1). At the same time, this article has conducted deprivacy processing on these data to ensure privacy protection and legality.

Table 1. Human resources data

Employee ID	Name	Gender	Age	Position	Section	Date of employment	Performance rating	Capability assessment
001	Zhang San	Man	30	Manager	Sales	2020-01-01	85	High
002	Li Si	Woman	28	Assistant	Personnel	2020-03-15	70	Middle
003	Wang Wu	Man	35	Advanced	Technology	2019-05-10	90	High
004	Zhao Liu	Man	32	Supervisor	Finance	2018-09-20	80	Middle
005	Xiao Ming	Woman	25	Internship	Sales	2021-02-28	60	Low

2.2. Feature Extraction

In order to extract features related to human resource management decisions from massive data, it is necessary to analyze them. These characteristics can include personal basic information, work experience, performance, etc. Among them, basic personal information can be obtained through simple statistics and calculations on the original data. The feature extraction of work experience can be obtained through statistical and classification of work information [10-11]. The calculation of working years can be obtained by subtracting the starting working time from the current working time. The calculation method for salary increase is based on the salary ratio of the previous year. The corresponding calculation formula is as follows.

$$N = G_n - K_n \tag{1}$$

In Formula (1), N is the length of service; G_n is the current date; K_n is the start date of work.

The calculation formula for salary increase is [12]:

$$W = {}^{X-S}/_{S} \times 100\% \quad (2)$$

In Formula (2), W is the rate of wage increase; X is the current annual salary; S represents the previous year's salary.

By collecting and categorizing employee evaluation results, the characteristics of performance evaluation can be obtained, and by calculating performance indicators, the ratio of performance indicators to target indicators can be obtained. The calculation of team collaboration score is based on evaluations from colleagues. The corresponding formula for performance evaluation calculation is as follows [13]:

$$P = {}^{Q}/_{R} \times 100\% \quad (3)$$

In Formula (3), P is the proportion of achievement of performance; Q is the actual number of completed performance; R is the performance target number.

2.3. Model Construction

In the modeling process, this article uses Hadoop as a distributed computing architecture to establish corresponding modeling methods and achieve deep analysis and mining of data. Hadoop is a trusted framework that can handle massive amounts of data and has scalability and fault tolerance capabilities [14-15]. With the help of Hadoop technology, parallel processing and analysis of data can be achieved on multiple nodes to accelerate computation and improve prediction accuracy.

During the modeling process, this article utilizes the programming mode of MapReduce in Hadoop to optimize the algorithm and solve it. MapReduce model divides the calculation work into two steps: mapping and reduction. In the mapping process, the input data is decomposed into several smaller blocks, and then a set of key value pairs is generated by manipulating each block [16]. During the reduction process, it aggregates those keywords and keyword pairs and produces the final result. With the help of the MapReduce model, the Hadoop big data platform model constructed in this article can fully leverage Hadoop's powerful distributed computing capabilities and achieve efficient parallel processing of data. Figure 1 shows the architecture of the Hadoop big data platform model constructed in this article:

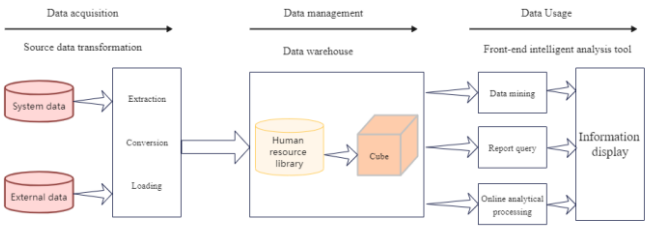

Figure 1. Architecture of Hadoop big data platform model

2.4. Decision Optimization

Decision optimization is a comprehensive analysis and simulation of the current and future situation of the enterprise. On the basis of the analysis results of the Hadoop big

data platform model constructed in this paper, targeted optimal decision-making solutions can be proposed. This article uses the constructed Hadoop big data platform model to transform the human resource decision-making problem into a linear programming problem. Specifically, this article takes the recruitment, selection, training, deployment, and motivation of human resources as decision variables, and constrains the company's strategic goals, market competition environment, and employee's capabilities. The objective function designed in this article is to minimize costs and maximize benefits. The minimum objective function is as follows:

$$min \sum(a_i * d_j) \tag{4}$$

The constraint condition for the objective function is:

$$s.t. A * d_j \geq b \tag{5}$$

In Formulas (4) and (5), a_i is the function coefficient vector of the i-th objective; d_j is the decision variable vector for the j-th objective; s.t. is the constraint condition; A is the coefficient matrix of equality constraints; b is the value vector of the equality constraint.

3. Experiment on Optimization Method for Human Resource Decision Based on Hadoop Big Data Platform

3.1. Comparison of Algorithms' Performance

The research on human resource management using big data has attracted increasing attention due to the coming era of big data. At present, in addition to Hadoop, there are more and more research hotspots in fields such as ML and DL. On this basis, this article compares the human resource decision-making optimization method based on the Hadoop big data platform with ML and DL algorithms to verify which algorithm has the best application effect in the optimal human resource decision-making. This article compares the performance of these algorithms mainly by comparing their F1 values. Figure 2 displays the experimental results.

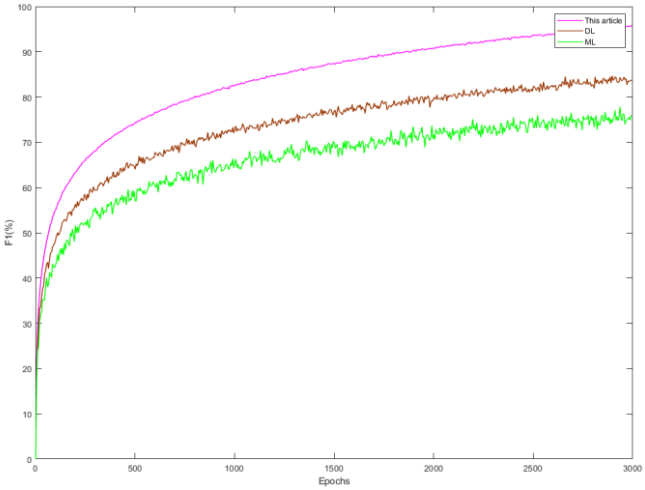

Figure 2. Comparison of algorithms' performance

Figure 2 shows that as the number of iterations increases, the F1 value of each algorithm gradually increases and reaches a steady state after the iteration is completed. The curve fluctuation of the method in this article is the smallest, which is the most stable, and the fluctuations in DL and ML are relatively significant. The F1 value of the method in this article remained stable at 95.8%, while the values of DL and ML are ultimately 84.7% and 77.9%. The experimental results suggest that compared with these two algorithms, the human resource decision optimization method based on the Hadoop big data platform in this article has superior performance.

3.2. Testing in Data Processing

Data processing has gradually penetrated into various aspects of enterprises and institutions. Therefore, how to effectively process massive data is currently an issue that needs to be addressed urgently. Faced with massive amounts of data, traditional processing methods are no longer suitable, resulting in the emergence of big data. Among these technologies, Hadoop-based big data processing platforms have become industry standards. In view of this, this article conducted a comparative experiment between Hadoop based data analysis platform and traditional computing platform to select a platform with better data processing accuracy. Figure 3 shows the experimental results of data processing accuracy.

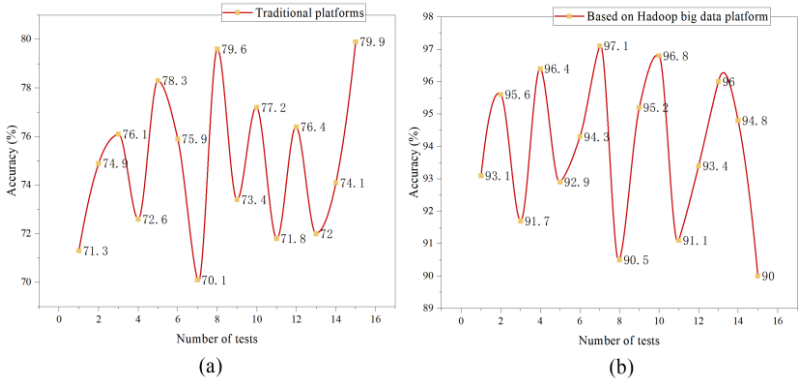

Figure 3. Comparison of data accuracy

Figure 3(a). Accuracy of traditional platforms

Figure 3(b). Accuracy of platform based on Hadoop big data

From Figure 3(a), it can be seen that the data fluctuates greatly, with accuracy ranging from 70.1% to 79.9%, with an average of 74.9%. As shown in Figure 3(b), the accuracy ranges from 90% to 97.1%, with an overall average of 93.9%. The experimental results suggest that the accuracy of the big data analysis platform established on Hadoop has significantly improved compared to traditional platforms. From the data fluctuation range of these two platforms, it can be seen that in most cases, Hadoop-based big data platform can process data more accurately and more stably. Overall, the Hadoop big data platform has higher accuracy in data processing, higher average accuracy, and more stable performance.

In addition to comparing accuracy, this article also compares the effectiveness of two platforms in processing data. The experiment involves two platforms

simultaneously processing data of the same scale, and then tests the processing time, processing efficiency, data integrity, and effectiveness in applicable scenarios. Table 2 shows the results.

Table 2. Comparison of data processing effects

Processing large scale data	Traditional platforms	Based on Hadoop big data platform
Amount of data processed (GB)	500	500
Processing Time (hours)	60	15
Processing efficiency (GB per hour)	10	40
Data integrity	Low	high
Applicable scenarios	Small scale data processing	Complex, large-scale data processing

From Table 2, it can be roughly understood how the data processing capabilities of the two platforms are. Further analysis shows that on the Hadoop big data platform, traditional platforms require 60 hours to process 500GB of data, while the Hadoop big data platform only requires 15 hours, indicating a significant improvement in the processing efficiency of Hadoop for massive data. Another more intuitive data is that the Hadoop big data platform can process 40GB of data per hour, while traditional platforms can only process 10GB of data per hour. Big data on Hadoop has higher integrity, but in traditional environments, data integrity is not high. In terms of application scenarios, traditional platforms are only suitable for small-scale data processing, while Hadoop based big data platforms are suitable for processing complex and large-scale data.

Overall, the Hadoop big data platform has stronger processing power, faster processing speed, higher data integrity, and is more suitable for processing large-scale data.

3.3. Comparison of Method Applicability

In order to compare the applicability of traditional methods and the Hadoop big data platform-based methods proposed in this article in different scenarios, corresponding experiments are designed in this paper. This experiment is conducted in three different human resource decision-making scenarios: enterprise recruitment, staff training, and performance evaluation. Through this experiment, the applicability of traditional methods and methods based on the Hadoop big data platform in different scenarios can be obtained, providing more scientific and effective method choices for human resource decision-making. Figure 4 shows the specific experimental results:

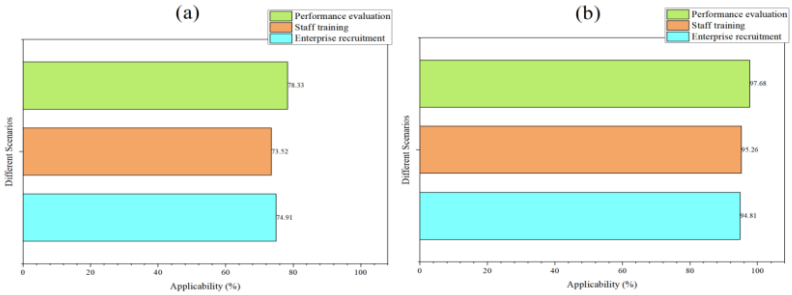

Figure 4. Comparison of method applicability

Figure 4(a). Applicability of traditional methods

Figure 4(b). Applicability of the method in this article

From Figure 4, it can be seen that the two methods are applicable in different scenarios, and the corresponding analysis is as follows.

In Figure 4(a), the applicability of traditional methods in the three scenarios of enterprise recruitment, staff training, and performance evaluation is 74.91%, 73.52%, and 78.33%, respectively, with an average applicability of 75.59%. In the applicability test results of the method in Figure 4(b), 94.81%, 95.26%, and 97.68% are achieved for these three scenarios, respectively, with significantly higher data compared to traditional methods. This can fully demonstrate the universal applicability and strong generalization ability of the method proposed in this paper.

This article also conducts in-depth research on the two algorithms from several perspectives, including stability, interactivity, data processing ability, decision optimization effectiveness, and model scalability. Through comparison, their respective characteristics and application scenarios can be clearly recognized, and suitable methods can be selected for processing based on this. Table 3 displays the experimental results.

Table 3. Comparison of other indicators

Parameters	Traditional method	Method based on Hadoop big data platform
Stability	Poor	High
Data interactivity	Limited	Highly interactive
Data volume processing capability	Low	High
Decision optimization effect	In general	Better
Model extensibility	In general	Better

It can be found in Table 3 that the method proposed in this article has obvious advantages in stability, interactivity, processing capability, decision optimization effect, extensibility and other aspects. This method can not only effectively process massive data, but also provide support for real-time queries and analysis. It can also improve the optimal efficiency of decision-making to a certain extent, and can extend the model, so in the big data environment, Hadoop-based big data analysis is a more suitable solution. However, in small-scale data environments, traditional algorithms are more suitable for small-scale data analysis as they do not require additional big data processing platforms and technologies.

4. Conclusions

In the big data environment, this article is based on Hadoop to study the optimization problem of talent resource allocation. Example analysis proves that the method proposed in this paper can increase the accuracy and efficiency of personnel decision-making in enterprises effectively, and provides a strong support for the personnel management work of enterprises. However, this method also has certain limitations, such as dependence on data quality and feature engineering, which require further research and improvement. Future research directions can include further improving the method, promoting its application to other fields, and strengthening its integration with other advanced technologies.

References

[1] Wahyoedi S, Suherlan S, Rijal S. Implementation of Information Technology in Human Resource Management[J]. Al-Buhuts, 2023, 19(1): 300-318.
[2] Putra R, Ali H. Organizational behavior determination and decision making: analysis of skills, motivation and communication (literature review of human resource management)[J]. Dinasti International Journal of Digital Business Management, 2022, 3(3): 420-431.
[3] Pourkhodabakhsh N, Mamoudan M M, Bozorgi-Amiri A. Effective machine learning, Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover[J]. Applied Intelligence, 2023, 53(12): 16309-16331.
[4] Chen S, Xu H. Meta-heuristic algorithm-based human resource information management system design and development for industrial revolution 5.0[J]. Soft Computing, 2023, 27(7): 4093-4105.
[5] Wang X, Wang Y, Zheng R. Impact of human capital on the green economy: Empirical evidence from 30 Chinese provinces[J]. Environmental Science and Pollution Research, 2023, 30(5): 12785-12797.
[6] Mahmood Q U A, Ahmed R, Philbin S P. The moderating effect of big data analytics on green human resource management and organizational performance[J]. International Journal of Management Science and Engineering Management, 2023, 18(3): 177-189.
[7] Bonner R L, Neely A R, Stone C B. Triaging your human capital: an integrative perspective on strategic human capital asset allocation[J]. Management Research Review, 2023, 46(3): 467-482.
[8] Varma D, Dutta P. Empowering human resource functions with data-driven decision-making in start-ups: a narrative inquiry approach[J]. International Journal of Organizational Analysis, 2023, 31(4): 945-958.
[9] Berk L, Bertsimas D, Weinstein A M. Prescriptive analytics for human resource planning in the professional services industry[J]. European Journal of Operational Research, 2019, 272(2): 636-641.
[10] Zhou J, Liu L, Wei W. Network representation learning: from preprocessing, feature extraction to node embedding[J]. ACM Computing Surveys (CSUR), 2022, 55(2): 1-35.
[11] Li X, Zhao H, Yu L. Feature extraction using parameterized multisynchrosqueezing transform[J]. IEEE Sensors Journal, 2022, 22(14): 14263-14272.
[12] Sijabat S. Decision Support System for Determining the Level of Employee Salary Increase in CV. Hera Citra Abadi by Using the Simple Additive Weghting (SAW) Method[J]. International Journal of Basic and Applied Science, 2019, 8(3): 97-102.

[13] Zhu L, Karim M M, Sharif K. SDN controllers: A comprehensive analysis and performance evaluation study[J]. ACM Computing Surveys (CSUR), 2020, 53(6): 1-40.
[14] Kalia K, Gupta N. Analysis of hadoop MapReduce scheduling in heterogeneous environment[J]. Ain Shams Engineering Journal, 2021, 12(1): 1101-1110.
[15] Aggarwal R, Verma J, Siwach M. Small files' problem in Hadoop: A systematic literature review[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(10): 8658-8674.
[16] Dong S, Kryczka A, Jin Y. Rocksdb: Evolution of development priorities in a key-value store serving large-scale applications[J]. ACM Transactions on Storage (TOS), 2021, 17(4): 1-32.