Research on Power Grid Position Allocation Decision Based on Multi-stage Modeling

Changjun Zhao, Xiaoyun Ding(✉), Shunyu Deng, Zhiwei Tan, Gaixia Kang, and Xiaogang Chen

State Grid Gansu Electric Power Company Marketing Service Center, Lanzhou 730300, China
254167329@qq.com

Abstract. The advanced post matching system is of great significance for enterprise development and personal career development. We conduct a statistical survey on employees and posts of a power grid unit, and use the random forest under the multi-stage modeling strategy to study the relationship between individuals and posts. The research shows that the random forest under the multi-stage modeling strategy can effectively realize the job matching. On this basis, based on the characteristics of random forest discovery, such as the identification of the person in charge of the post, years of entry, age, years of termination of the labor contract, time of graduation from current education, professional and technical qualification level, etc., it has certain explanatory power for the post matching of the power grid.

Keywords: position matching · multi-stage · random forest · management and decision-making

1 Introduction

In the process of digital transformation of power enterprises, the data of enterprises has grown rapidly and become the demand and standard structure of power enterprises [1]. Through data mining can create new value, in the current environment of fierce competition, electric power enterprises realize the importance of modern management technology application and rational utilization of talent, talent as the core competitiveness of electric power enterprises, in the management of employees, reasonable personnel call, position matching is conducive to the development of electric power enterprise, and improve the level of management [2].

As a high-dimensional data solution, machine learning has been widely concerned by academic circles through its research and its position matching in electric power enterprises, and has improved the corresponding measures in recent years [3]. Zhang Zhiyu et al. [4] applies the data mining technology to the person-position matching management of enterprises, and analyzes the application process of the data mining technology in the person-position matching problem in detail. Jiang Rong [5] establishes the index system and its standards for the ability required by the position and takes it as the input of BP neural network to build the person-position matching evaluation

C. F. Peng et al. (Eds.): MMET 2022, ASSEHR 703, pp. 298–305, 2023.
https://doi.org/10.2991/978-2-494069-51-0_41

model. The feasibility of the evaluation model is verified by the person-position matching in state-owned enterprises. Zhang Qin et al. [6] studied the relationship between the evaluation tools and the position through regression analysis, and explored that the current evaluation tools have a certain explanatory ability of the person and position matching. Wang Zelong [6] combined with the statistical analysis method, used the data mining technology to build the position portrait of industrial engineers under the background of new technology, analyzed and extracted the indicators from the portrait, and used it for the match of applicants and positions. Compared with the more subjective traditional interview and survey method, it has obvious advantages. Sang Haifeng et al. [7] conducted a follow-up visit and questionnaire survey on the promotion of college students after graduation. Based on the decision tree algorithm, they established the promotion factors of college graduates, analyzed the pre-processed data, extracted the potential rules influencing the promotion of college graduates, and conducted verification and analysis.

In this paper, we use the random forest to analyze the connections between the various attributes of the sample data, obtain the main potential influencing factors, and adopt a multi-stage modeling strategy for different positions to achieve position matching. The research result of this paper is the exploration and attempt of the digital transformation process of electric power enterprises, hoping to accumulate experience for the intelligent prediction of position matching.

2 Random Forest Position Matching Model Based on a Multi-stage Modeling Strategy

2.1 Data Preprocessing

According to the actual statistical data of the electric power enterprise, the data information is missing and discrete due to some factors in the statistical process. The absence of data leads to reduced sample information, which biases the analysis results of the data. Too many discrete values lead to errors and even errors in the analysis results. In order to increase the interpret ability and improve the prediction accuracy of the model. Therefore, the data need to be detected, cleaned, and feature-derived before modeling.

Since the data selected in this paper is under the same report cycle, that is, the variance of the report cycle is 0, and the variance attribute of 0 attribute has no impact on the prediction results, so consider removing such features. In addition, with the help of relevant business personnel, this paper has deleted a large number of redundancy and redundant attributes and attributes with excessive missing values, such as labor plan source, unit level, secondary unit name, etc. For parts with a small number of missing values, this paper achieves missing values based on the hierarchical median filling method based on practical business considerations. For more discrete properties, but with general rules, this paper uses feature recoding to improve the robustness of the model.

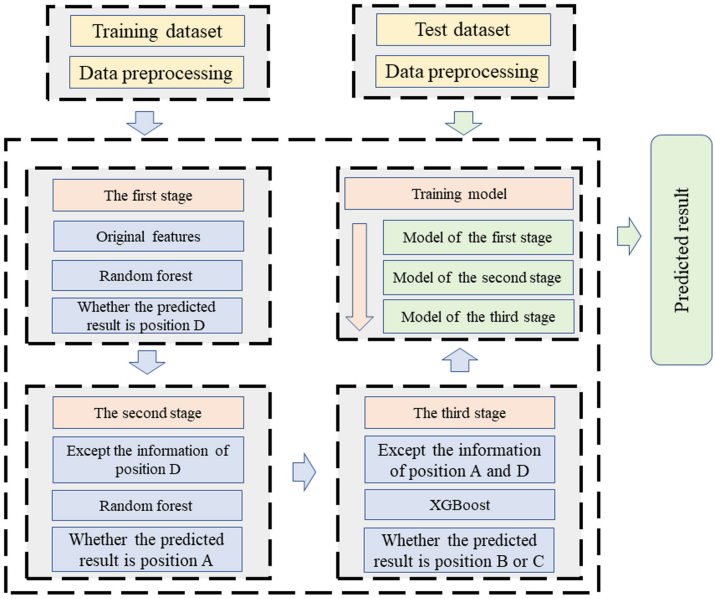

Fig. 1. The flowchart of the multi-stage model [owner-drawing]

2.2 Based on Random Forest and the XGBoost Algorithm to Implement Multi-stage Modeling

As the positions of this data are divided into four categories, they are transformed into four stages of positions according to the promotion rules of corporate positions. Random forests was used to model the four stages separately. Random Forest is an integrated model based on Bagging ideas. As the positions of this data are divided into four categories, they are transformed into three stages of positions according to the promotion rules of corporate positions. Random forests are used to model the first two stages separately. XGBoost [9] is applied to the third stages. Random Forest is an integrated model based on Bagging ideas [10], its base classifier is a decision tree model. By randomly sampling the original data, the original data set is divided into multiple decision trees. The random selection of samples and features causes the difference of the generated individual decision trees, thus improving the anti-noise performance and generalization ability of the algorithm, and making the performance more stable. Finally, the voting method is adopted to take the category with the largest number of votes as the judgment result. The flowchart of the multi-stage model is shown in Fig. 1.

Through the above method, 38 properties were finally obtained. Attributes of the dataset are shown in Table 1.

Table 1. Attributes and definition of the dataset [owner-drawing]

Attribute	Definition
G	Employee gender
PS	Whether it is a party member
IS	Cadres or workers
DS	Whether it is a demobilized soldier
CT	Term or no maturity
NAR	985/211/general universities
CMNAR	Classification of Majors under the now the academic record
GINAR	Graduation institution under the now the academic record
PTQS	Professional and technical qualification series
WETM	Whether it is engineering technology master
PTQL	Professional and technical qualification level
CP	Length of contract commencement to termination
IOLMC	Internal organization at the level of a major category
FTHDM	Full-time highest degree major
EE	Employment degree
CPSED	Classification of professions studied by employment degree
SC	Source Channel
IPP	Identification of the person in charge of the post
S	surname
PW	Place of work
POP	Place of origin(Province)
WOIP	Whether the origin information is in the province
A	Employee age
YGCD	The year of graduation of current degree
ENOWT	Engaged in other work time
YB	Year of birth
MB	Month of birth
QB	Quarter of birth
YPW	Year of participation in the workforce
MPW	Month of participation in the workforce
QPW	Quarter of participation in the workforce
YEC	Year of entry into the company

(*continued*)

Table 1. (*continued*)

Attribute	Definition
MEC	Month of entry into the company
QEC	Quarter entry into the company
YTEC	Year of termination of employment contract
MTEC	Month of termination of employment contract
QTEC	Quarter of termination of employment contract
CC	classification code

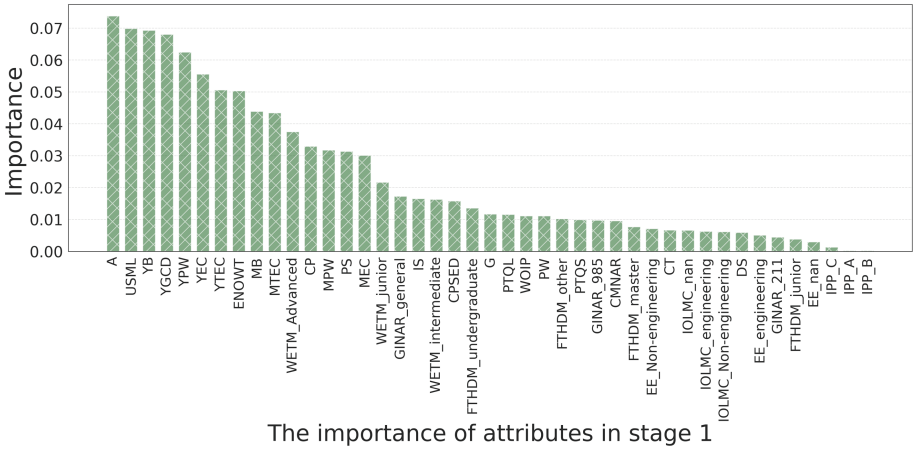

Fig. 2. Stage 1 attribute importance sort in the dataset [owner-drawing]

3 Experiment and Analysis

3.1 Attribute Analysis

In this paper, the preprocessed data is used as the input to the random forest to achieve the importance score of the attributes. The scoring results for the three stages can be plotted and shown in Figs. 2, 3 and 4.

As can be seen from Figs. 2, 3 and 4, the influence degree of the attributes at different stages varies. For the D position, among the data characteristics selected in this paper, the contract period signed when entering the electric power enterprise, the current education and graduation time, the age-related factors, and the technical level are the priority factors. For position A, whether it is the identification of the person in charge, the factors related to qualifications, and the technical level have an obvious influence. In position A, people with both qualifications and skills are preferred to be appointed as the person in charge of the position. In addition, the impact of 985 project in position A is greater than that of ordinary colleges and 211 project, while other positions have relatively small restrictions on education. For B and C positions are similar to D

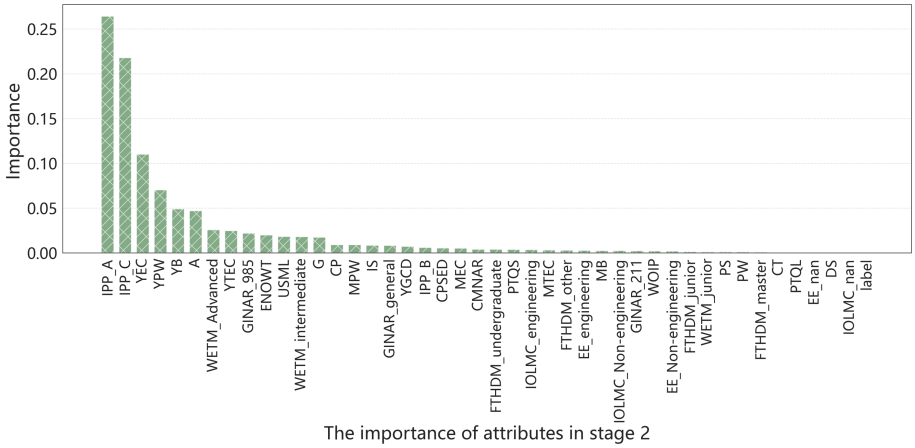

Fig. 3. Stage 2 attribute importance sort in the dataset [owner-drawing]

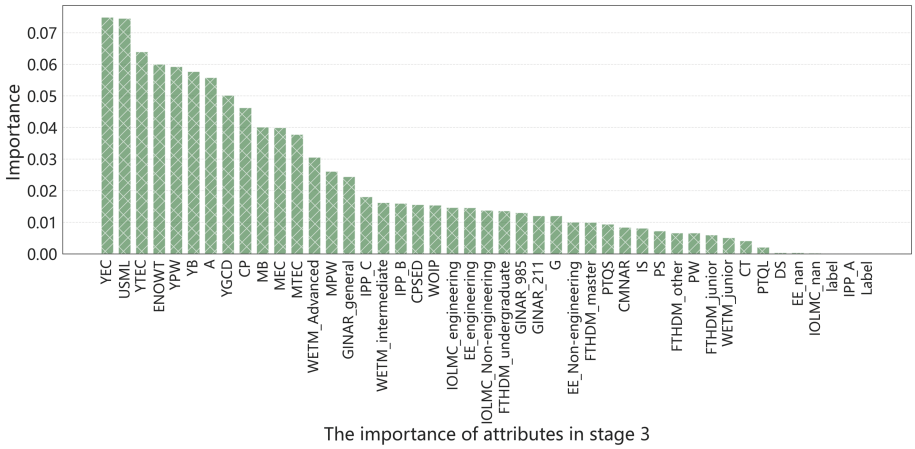

Fig. 4. Stage 3 attribute importance sort in the dataset [owner-drawing]

positions, but for B and C positions, the year of entering the unit is the primary influencing factor. Different institutions and factors related to qualifications are more important than technology. Therefore, there is a high probability of arranging D positions for new employees in electric power enterprises. In three different stages, not all attributes have effects, and these features that do not, are redundant in the corresponding stages. So it can be eliminated in the prediction stage.

3.2 Position Match

The attributes after the feature screening are used as the model input of the corresponding stage. By constantly adjusting the parameters of the model and comparing the prediction

Table 2. Parameter setting list [owner-drawing]

stage	parameter	
stage 1	n_estimators = 500	max_depth = 9
stage 2	n_estimators = 1000	max_depth = 3
stage 3	n_estimators = 500	max_depth = 8

Table 3. Evaluation results of the model [owner-drawing]

stage	accuracy	recall
stage 1	0.9	1.0
stage 2	1.0	1.0
stage 3	0.8125	0.95

Table 4. The accuracy of the different algorithms on the overall data prediction [owner-drawing]

stage	accuracy	recall
random forest	0.75	0.75
XGBoost	0.7	0.72

accuracy, the final parameters of each stage are shown in Table 2. The evaluation results are shown in Table 3.

As can be seen from Table 4, the multi-stage modeling strategy can accurately identify the position allocation at different stages. Table 4 shows the overall performance of different algorithms on this dataset. Compared with Table 3, the modeling method in this paper has significantly improved the accuracy and recall rate. To sum up, in this dataset, the multi-stage modeling strategy used in this paper has better identification accuracy, which can effectively help the power enterprise to complete the position allocation problem of talents, while a single modeling method cannot complete the effective position matching.

4 Conclusion

Through data pre-processing, feature selection and model training, the factors affecting the position of personnel are analyzed in an electric power enterprise in China, and a multi-stage talent position matching model is established. The results show that the analysis of personnel information by means of data science can help enterprises to allocate and assess the position personnel. In addition, using multi-stage modeling can effectively predict whether employees fit the company position, which is in line with the actual personnel allocation of the enterprise. The model outperforms the random forest

and XGBoost directly modeling the overall data. It is of more guiding significance to the actual personnel position allocation of electric power enterprises.

References

1. Liu Piao, Cheng Donghui, Gao Qiqi, Lu Chen. Based on the Big Data Job Analysis and Recommendation System [J]. The Smart City, 2021 ,7 (16):13–14.
2. Bai Min, Gu Huaixin, Cheng Wei. Research on person Resources Data Mining Technology and Its Application [J]. commercial culture, 2021(14):128–129.
3. Jiang Fei. Application of machine learning in person-position matching degree measurement model [D]. Hunan University of Technology, 2012.
4. Zhang Zhiyu, Lu Mingli, Li Cong. The Application of Data Mining in person-position Matching [J].China Person Resources Development, 2010(02):51–53.
5. Jiang Rong. Based on the evaluation model of managers of BP neural network [D]. Shenyang University of Technology, 2018.
6. Zhang Qin, Zhao Weidong, Xiong Jiabuo. Empirical study on the matching relationship between evaluation tools and positions [J]. Journal of TC (Social Science Section):1–8.
7. Sang Haifeng, Jiang Mingdi, Lu Zhong Qiao, conductor Jin, Wei Fuhong. Data mining algorithm for college promotion based on decision tree [J]. Journal of North China University (Natural Science Edition), 2019, 20(06):836–840.
8. Park Jaewon, Shin Minsoo, Heo Wookjae. Estimating the BIS Capital Adequacy Ratio for Korean Banks Using Machine Learning: Predicting by Variable Selection Using Random Forest Algorithms[J]. Risks,Volume 9, Issue 2. 2021. P 32-32.
9. Ma Meihong, Zhao Gang, He Bingshun et al. XGBoost-based method for flash flood risk assessment[J] Journal of Hydrology, 2021, 598
10. Li Xinhai. Application of the random forest model in classification and regression analysis [J]. Journal of Applied Entomology,2013, 50(04): 1190-1197.