



Research and Application on the Model of State-Owned Enterprise Cadre Portrait Under the Background of Big Data

Youping Liu¹(✉), Jiangang Shen², and Junshuai Xue³

¹ China North Talent Research Institute, Beijing, China
bqrclyp@163.com

² Beijing Mafumatou Technology Co. Ltd., Beijing, China

³ Xi'an Electronic Engineering Research Institute, Xi'an, China

Abstract. Cadre management is one of the key points and difficulties in the management of state-owned enterprises. On the basis of human resource big data, this paper introduces talent portrait technology into cadre management and puts forward a model of state-owned enterprise cadre portrait based on big data and natural language processing technology to solve the problems of appraisal, performance evaluation, and selection of cadres. The model, which collects the basic information, performance assessment, personality characteristics and other aspects of data of cadres, after data cleaning and conversion, uses expert prediction and information technology to generate labels, taking an enterprise cadre selection and appointment business scenario as an example to explain the application of the model.

Keywords: Cadre portrait · Big data · Natural language processing

1 Introduction

Cadre management is one of the important contents of human resource management (HRM) in state-owned enterprises. Meanwhile, in-depth understanding of cadres, objective evaluation cadres, comprehensive assessment cadres, and accurate selection of cadres are also management problems faced by many state-owned enterprises. In order to solve the problem of cadre management, more and more scholars have tried to introduce portrait technology into cadre management in recent years [1, 2].

Portrait technology originated from the e-commerce field of user portraits. In the context of big data, each specific information of users has been abstracted into labels, which are used to concretize the user profile, so as to provide targeted services for users. Portrait technology has been more and more applied to HRM in the process of existing talent and potential talent portraits in enterprises and public institutions, which finally contributes to talent portrait technology. This technology has been increasingly and successfully applied in talent recruitment, talent training, talent-post matching and performance evaluation and other fields in recent years.

Based on human resources and big data concerning cadre management, talent portrait method and natural language processing technology are innovatively introduced, and cadre portrait model is constructed to solve practical business problems such as cadre appraisal, performance evaluation, and selection. Firstly, the construction method of cadre portrait model is studied. Furthermore, the cadre portrait labeling system applicable to state-owned enterprises is proposed on the basis of this model, and the application of the model is illustrated by taking the cadre selection and appointment of a large group state-owned enterprise as an example.

2 Building of a State-Owned Enterprise Cadre Portrait Model

The portrait of state-owned enterprise cadres is a complex system project. Firstly, it is necessary to collect data from multiple information systems involved, such as basic information of cadres, performance assessment, personality characteristics and other aspects. Secondly, cleaning and converting these data to remove dirty data and convert them makes it indispensable to a unified caliber cadre management big data. Thirdly, the state-owned enterprise cadre portrait label system is extracted and generated based on the big data. Finally, according to the characteristics of business scenarios, we need to use the label system to portray the cadres. In general, it consists of four steps: data collection, data preprocessing, label generation and cadre portrayal.

2.1 Data Collection

The comprehensiveness and completeness of the original data will directly affect the result of the portrait. Therefore, data acquisition is an important work for the portrait of SOE cadres. The original data of portrait should not only include structured data such as personnel file and assessment, but also unstructured data such as organizational conversation, inspection materials, as well as psychological assessment data that can reflect personality characteristics and psychological factors.

2.2 Data Preprocessing

The original data collected from the associated system needs to be cleaned and converted before it can be used to generate labels.

(1) Data Cleaning.

Data cleaning is the process of supplementing, correcting, and cleaning the incomplete data, incorrect data, and repeated data in the original data to improve the data quality. In this paper, the data cleaning operation is carried out mainly by computer automatic cleaning and supplemented by manual cleaning.

(2) Data conversion.

Before generating labels, a lot of data conversion work needs to be done, which includes the following three parts.

Identity Unification. The data of cadres distributed in different systems need to generate the unique identity identification that can determine the identity of cadres through

the identity unification procedure, and associate the data of each system through this identification.

Code Conversion. Some fields in the original data are code values, such as gender, ranking, professional title, etc. and it is necessary to convert the code values of the same code item in each system into a unified value.

Calculation and Aggregation. The original data is often too fine-grained and needs to be converted into coarser-grained data after calculation and aggregation for the purpose of generating labels. For example, the date of birth shall be converted into age, work history data needs to be aggregated into working time (months) in basic organizations and working time (months) in authority organizations, and reward, honor and punishment data shall be aggregated into the times of reward, honor and punishment at different levels.

2.3 Label Generation

The methods of talent portrait label generation include expert prediction method, empirical pattern recognition method [3, 4], semantic labeling method based on multimodal subspace learning [5], and in recent years, some scholars have also used big data analysis and deep learning techniques to generate labels [6, 7]. In this paper, we use a combination of expert prediction and big data and natural language processing technologies to generate state-owned enterprise cadre portrait labels.

In the model of SOE's cadre portrait model studied in this paper, the label system includes five major categories of labels: basic information, work situation, personality characteristics, performance assessment, and reward, honor and punishment situation. Among them, the labels under the personality characteristics category need to be automatically extracted from the organizational conversation and quality assessment big data using natural language processing and big data technology, and the labels under the other four categories are generated by the discussion of experts engaged in human resource management and organizational management.

(1) Extracting labels from big data of organizational conversations.

Organizational conversation data is often stored in the form of word and PDF documents, which are the conversation records and investigation materials of the interviewed cadres by the superior leaders, colleagues at the same level and lower levels in the previous organizational work and inspection activities, including evaluations of political awareness, business ability, work experience, incorruptibility, contact with the masses and other aspects. Some personality characteristics of cadres can be extracted from them using natural language processing technology, and the extraction model is shown in Fig. 1.

Corpus. All the interview materials are used as the corpus, which is the base of the word segmentation and the statistics of word frequency.

Document Parser. First of all, to count the word frequency more accurately, all the document materials from the same interviewee need to be merged into one document.

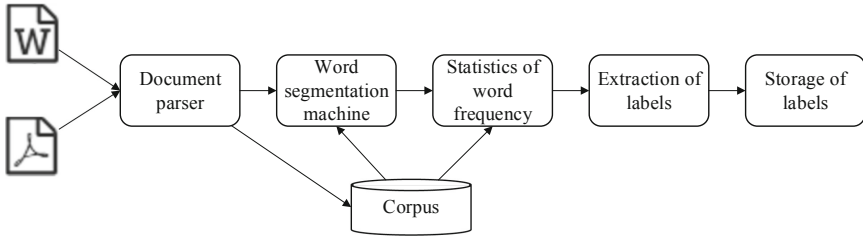


Fig. 1. The model of extracting labels from the big data of cadre interviews

Then the documents in word and PDF format are parsed and converted to plain text format for subsequent processing. Meanwhile, the parsed document is added to the corpus.

Word Segmentation Machine. Use Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS).

Chinese word segmentation tool to segment the content and mark the part of speech of the word segmentation results. According to the writing characteristics of the interviews and inspection materials of cadres in state-owned enterprises, the phrases of “noun+adjective (in Chinese)” (such as high cultivation of party spirit, strong management ability, strong affinity, etc.) are combined into one word for subsequent calculation of word frequency.

Statistics of word frequency. Using the TF-IDF algorithm, word frequency and inverse document frequency are counted for each word in the document and TF-IDF values are calculated.

$$TF - IDF = TF \times IDF \quad (1)$$

In formula (1), TF is the word frequency of a word t in the document.

$$TF = \frac{C_t}{C_{total}} \quad (2)$$

In formula (2), C_t is the number of occurrences of word t in the document, and C_{total} is the total number of occurrences of all words in the document.

In formula (1), IDF is the inverse document frequency of word t .

$$IDF = \ln\left(\frac{D_{total}}{D_t + 1}\right) \quad (3)$$

In formula (3), D_{total} is the number of all documents in the corpus, and D_t is the number of documents containing word t .

Extraction of Labels. TF-IDF values reflect the importance of a word in a document and is often used to extract keywords from a document. In this paper, we use the 5 words with the largest TF-IDF values in the document as the labels corresponding to the cadre.

Storage of Labels. The extracted labels are stored in the database for the subsequent cadre portrait procedure.

Update of Labels. On the one hand, when a new document is generated for an interviewee, the above process is automatically performed to update the labels of that interviewee; on the other hand, due to the increasing number of documents in the corpus, it is

Table 1. Indexes of quality evaluation scale for management talents

Level 1 index	Level 2 index
Ambition energy	success desire, power motivation, competitiveness, persuasion, determination, vitality
Thinking mode	innovation ability, curiosity, insight, foresight, rational ability, critical ability
Emotional adjustment	optimism, self-affirmation, seeking change, emotional stability, stress resistance, adaptability
Interpersonal interaction	empathy, gregariousness, openness, supportiveness, modesty
Task execution	responsibility, willpower, organization, standardization, rigor, independence

necessary to regularly update the labels of the previously generated interviewees using the latest corpus.

(2) Extracting labels from psychological evaluation big data.

Using the professional psychological evaluation scale to test SOE cadres, the personality and psychological characteristics of cadres can be obtained. The indexes included in the scale used in this paper are shown in Table 1.

In Table 1, each index has a full score of 10. In this paper, the indexes comments with scores higher than 7.5 points or lower than 2.5 points are extracted as the personality characteristics labels of the cadre, such as “strong career ambition”, “hesitation in decision-making” and “long-term vision”.

2.4 Cadre Portrait

After the portrait labels of are generated, they can be applied to cadre evaluation, cadre selection, cadre training and optimization of salary mechanisms. Different business scenarios can adopt different technical means to generate different forms of cadre portraits on the basis of the labeling system. Commonly used portrait methods include analytic hierarchy process [1], state diagram method [8], statistical analysis method [9], and clustering method [10], etc.

3 State-Owned Enterprise Cadre Portrait Labeling System

In this paper, the combination of expert prediction, big data and natural language processing technologies is used to generate the portrait label system of state-owned enterprise cadres, which is shown in Fig. 2.

The labeling system is divided into two levels, of which five first-level labels include basic information, work information, personality characteristics, assessment and rewards, honors and punishments in recent years. The personality characteristics are dynamically and automatically generated by big data and natural language processing technology, and the other 4 types of labels are determined by the experts.

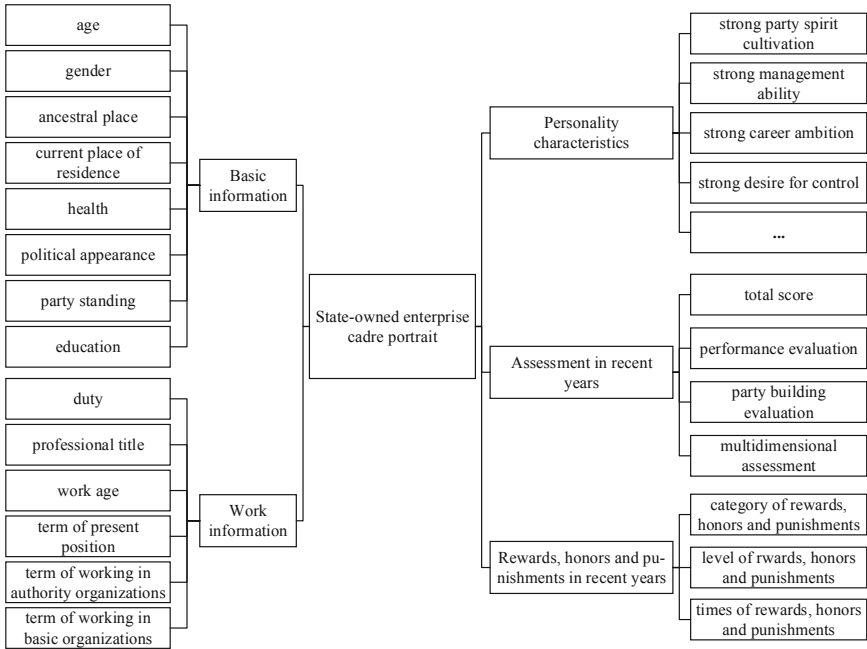


Fig. 2. The portrait labels of cadres in state-owned enterprises

4 Application Example

The selection and appointment of cadres is one of the typical application scenarios of cadre portraits. Taking the cadre selection and appointment scenario of a large group state-owned enterprise as an example, this paper illustrates the application method of the above mentioned cadre portrait model.

4.1 Data Sources

The original data of enterprise’s cadre portrait comes from the human resources management information system, the cadre assessment and evaluation system, the cadre interview whole-process documentation system, and the management talent quality evaluation system. The original data of cadre portrait is shown in Fig. 3.

(1) Human resource management system.

The system includes organizational structure management, employee information management, personnel change, salary management, contract management, statistical analysis and other modules. Up to now, more than 4 million pieces of personnel information data has been stored the system. In this paper, the personnel file information of the cadres is extracted from the human resources management system, including

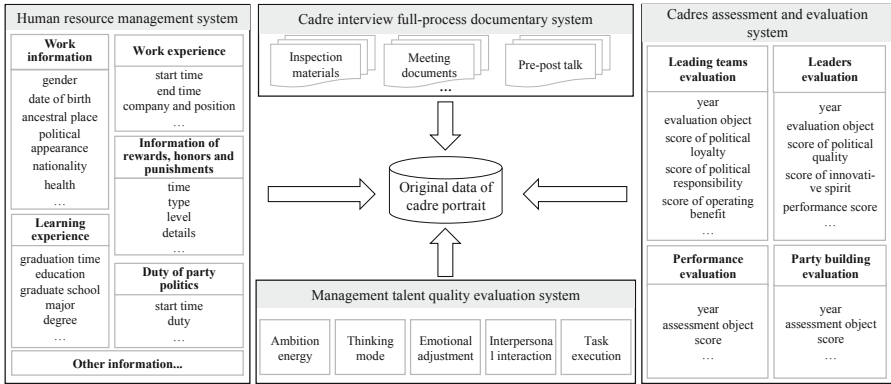


Fig. 3. Original data of cadre portrait in a state-owned enterprise

basic information data, the work information data, work history data, reward, honor and punishment and other data in the past five years.

(2) Cadres assessment and evaluation system.

The system is used to evaluate annual performance of the leading teams and cadres at all levels. Members of leading team, middle-level cadres, staff representatives and external directors score the leading members and cadres from multiple dimensions via mobile phones. In this paper, the data of party building evaluation, performance evaluation and multi-dimensional assessment data of cadres in the past three years should be selected from the system. At present, the system has contained more than 95000 assessment data of leaders over the years.

(3) Cadre interview full-process documentary system.

The system is used to record and query the materials of the conversation of the cadres by the relevant personnel of the superior, the peer and the subordinate level in the whole process management of cadres' selection, appointment, cultivation and retention. There are more than 62,000 documents and materials saved in the system. And, this paper collects the records and inspection materials of the organizational department on the talk about the term of office, assessment, selection and appointment of cadres from the documentary system.

(4) Management talent quality evaluation system.

The system comprehensively evaluates the psychological quality and personality characteristics of all the management talents in the enterprise, and assesses management talents by the Management Talent Quality Assessment Scale. At present, there are more than 125000 data of evaluation results. This paper collects the assessment result data of five aspects: ambition energy, thinking mode, emotional adjustment, interpersonal interaction and task execution.

After collecting the original data from the above systems, data preprocessing and label generation, portrait labels of 1236 candidate middle-level cadres from the enterprise were generated.

4.2 Cadre Portrait

According to the business characteristics of leading cadres’ selection and appointment, this study uses Analytic Hierarchy Process [8] to determine the weight of each label and uses the following formula (4) to calculate the matching degree between the employment positions P and the candidate cadres E.

$$Sim_{(E,P)} = \sum_{i=1}^n S_{(i,P)} \times W_i \tag{4}$$

W_i is the weight of the label i , $S(i, p)$ is the matching degree between label i of cadre E and the corresponding requirement of post P. $S(i, p)$ is calculated as follows.

- 1) Set the requirements corresponding to the labeling system for each post, such as the requirement of the Director selection of Science and Technology Division at the headquarters of the enterprise in 2022 {age: between 25 and 40 years old, gender: male, political appearance: party member, professional title: senior and above, education: Master’s degree or above, term of working in basic organizations: more than 3 years, strong party spirit cultivation, strong business ability, strong innovation ability, average of total scores in the past three years: > 90}. The weights of the above labels are normalized as shown in Table 2 (label weights that have nothing to do with job requirements are treated as 0).
- 2) Personality characteristic label: if the personality labels of cadre E contains a personality trait keyword required by position P, the matching degree $S(i, p)$ of the label is 1, otherwise it is 0.
- 3) Other category labels: if the value of the label of cadre E meets the corresponding requirements of post P, the matching degree of the label $S(i, p)$ is 1, otherwise it is 0. For example, post P requires age of the candidate cadre between 25 to 40, and the age of cadre E is 35, then $S(i, p) = 1$. The labels of some candidate cadres for this position are shown in Table 3.

After calculating the matching degree of all candidates for all positions through formula (4), the proposed results of cadre selection and appointment can be derived from the cadre portrait according to the matching degree sorted from highest to lowest, as shown in Table 4.

Table 2. Weight of each label of the Director of Science and Technology Division

Label	Weight	Label	Weight
Age	0.2	Gender	0.05
Political appearance	0.05	Professional title	0.05
Education	0.15	Term of working in basic Organizations	0.1
Strong party spirit cultivation	0.05	Strong business ability	0.1
Strong innovation ability	0.15	Average of total scores in the past three years	0.1

Table 3. Labels of some candidate cadres

Cadre	Age	Gender	Political appearance	Professional title	Education	Term of working in basic organizations (in years)
Cadre 1	29	man	party member	senior	master	4
Cadre 2	31	man	party member	senior	doctor	5
Cadre 3	44	woman	party member	professor	master	7
Cadre 4	37	man	party member	professor	bachelor	3
Cadre	Personality characteristics			Average of total scores in the past three years		
Cadre 1	strong party spirit cultivation, firm political stand, strong communication and strong innovation ability			97.28		
Cadre 2	strong career ambition, strong business and strong innovation ability			95.63		
Cadre 3	strong party spirit cultivation and firm political stand			90.51		
Cadre 4	strong business, strong innovation and strong management ability			99.63		

Table 4. Results of matching degree calculation for some candidate cadres

Cadre	Overall Matching	Age	Gender	Political appearance	Professional title	Education	Term of working in basic organizations
Cadre 2	0.95	1	1	1	1	1	1
Cadre 4	0.8	1	1	1	1	0	1
Cadre 1	0.75	1	1	1	1	1	1
Cadre 3	0.5	0	0	1	1	1	1

(continued)

Table 4. (continued)

Cadre	Overall Matching	Strong party spirit cultivation	Strong business ability	Strong innovation ability	Average of total scores in the past three years
Cadre 2	0.95	0	1	1	1
Cadre 4	0.8	0	1	1	1
Cadre 1	0.75	1	0	0	1
Cadre 3	0.5	1	0	0	1

5 Conclusion and Prospect

This paper researches the mechanism of cadre portrait of state-owned enterprises based on big data, and establishes the portrait model by using the idea of “data collection → data preprocessing → label generation → cadre portrait”, and in the label generation stage, the cadre personality characteristics labels are automatically generated and updated using big data and natural language processing technologies. Finally, a large conglomerate SOE cadre selection and appointment business scenario is taken as an example to illustrate the application of the portrait model proposed in this paper. In the future, in the process of continuous promotion and application of the cadre portrait model proposed in this paper, the method of automatic generation and updating of other categories of labels except personality traits can be further studied by technical means such as big data.

References

- Hou B., Bai X., Liu T. et al. (2021) Research on the construction of corporate economic responsibility audit portrait index from the perspective of big data. *Friends of Accounting*, 2021(23): 102–106. <https://kns.cnki.net/kcms/detail/14.1063.F.20211126.1522.026.html>.
- Dai L., Peng C. (2017) The mechanism of auditing and supervising of cadre personnel archives based on big data. *Archives & Construction*, 2017(6): 31–34+90. <https://kns.cnki.net/ms2/article/abstract?v=3uoqIhG8C44YLTIOAiTRKibYIV5Vjs7iAEhECQAQ9aTiC5kcBjCgn0Ru0QXXwcKR7urR5JAeHbF8pjSzCNSbT58x8WawSRsfg9&uniplatform=NZKPT>.
- Wang X., Liu J., Xu D. (2018) A practice of user profile in big data. *Telecommunications Science*, 2018, 34(5): 127–133. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTIOAiTRKibYIV5Vjs7i0kJR0HYBJ80QN9L51zrP0tqmYZmQdPpQqm3tOmRT-kfrnMIJmgnwubYeGEji-r9&uniplatform=NZKPT>.
- Xue L., Song W. (2020) Relation extraction method based on dynamic label. *Journal of Computer Applications*, 2020, 40(6): 1601–1606. <https://kns.cnki.net/kcms/detail/51.1307.TP.20200122.1128.006.html>.
- Tian F.g, Li X., Liu F., et al. (2020) A semantic tag generation method based on multi-model subspace learning. *Journal of Shandong University (Engineering Science)*, 2020, 50(3): 31–37+44. <https://kns.cnki.net/kcms/detail/37.1391.T.20200617.1404.010.html>.

6. ZHAO Y. (2021) Generation method of user portrait talent label system based on big data mining. *Information Technology*, 2021, (12):142-146. DOI: <https://doi.org/10.13274/j.cnki.hdzj.2021.12.025>.
7. Liu X., Liu B., Wang Y., et al. (2020) Research and Development of Multi-label Generation Based on Deep Learning. *Computer Science*, 2020,47(3):192–199. <https://kns.cnki.net/kcms/detail/50.1075.TP.20191125.1356.002.html>.
8. Jiao T., Yang Q., Lei Y., et al. (2020) Research on dynamic multi-dimensional development of new generation scientific research talents based on talent portrait technology. *Modern Business*, 2020, (32):60-62. DOI: <https://doi.org/10.14097/j.cnki.5392/2020.32.021>.
9. Cha L., Yu T., Shi H., etc. (2019) Construction and Application of High Potential Talent Portrait Based on Data Analysis. *Enterprise Reform and Management*, 2019, (4):83-84. DOI:10.13768/j.cnki.cn11-3793/f.2019.0457.
10. Jiang L. (2022) Research and Implementation of Topic Model-based Text Mining Technology for Building Talent Portraits. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFDTEMP&filename=1022662085.nh>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

