2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy. October 6-9, 2019

A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization*

Jiamin Liu, Tao Wang, Jiting Li, Jingbo Huang, Feng Yao, and Renjie He

Abstract— **In the era of big data and industry 4.0, the mode of human resource management (HRM) will be changed. As the role of talents rises, enterprises need to pay more attention to human capital. Meanwhile, intelligent production also urges intelligent HRM. Based on the accumulation of data, enterprises should use big data and artificial intelligence technologies to analyze employees, predict the future, and support businesses. Thus, the paper attempts to put forward ideas on data-driven solution to the promotion issue in HRM, and focus on the influence from the position of organization. The data come from a state-owned enterprise in China. Here the features of organizational position are chosen to analyze employee promotion and forecast employee prospect. From the analysis based on statistics and networks, as well as the prediction based on machine learning, we find that structural position plays a more critical role than geographic position. Besides, employees can benefit from working in the place where special experience is available, where mobility is more stable, or where resource is more abundant. But organizations should be concerned with the fair development of workers. The experimental results also validate that the prediction model is practical and effective.**

I. INTRODUCTION

In the era of big data and industry 4.0, the production and operation mode of enterprises become increasingly intelligent. Accordingly, the management mode, including human resource management (HRM), must be improved [1,2]. As the new production mode requires employees to have control over the more complicated work and become more responsible, the role of humans in the development of industry 4.0 has been rising [3]. Moreover, the design and manufacture of products cannot be completed independently by the Cyber Physical Systems (CPS) [4], so the participation of human intelligence and labor are demanded. Enterprises need employees with stronger design and planning capabilities to strengthen real-world production and logistics. Only higher-quality staff can meet the needs of enterprise development nowadays [5]. Therefore, human resource is especially valuable.

*Research supported by the National Natural Science Foundation of China (grant number: 71701206, 61473001, 71690233).

Jiamin Liu is with College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: liujm_1020@163.com).

Tao Wang is with College of Systems Engineering, National University of Defense Technology, Changsha, China (corresponding author to provide e-mail: wangtao@nudt.edu.cn).

Jiting Li is with College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: lijiting@nudt.edu.cn).

Jingbo Huang is with College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: nudt_huang317@163.com).

Feng Yao is with College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: yaofeng@nudt.edu.cn).

Renjie He is with College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: renjiehe@nudt.edu.cn).

With the knowledge and skills of employees becoming the core competence of a company, HRM has been upgraded from strategic HRM to human capital management. In order to adapt to the changes of management itself and the current industrial upgrading, intelligent HRM is necessary. Based on the accumulation of data, enterprises should use big data and artificial intelligence technologies to analyze employees and organizations, predict the future, and support businesses. Currently, some companies like Baidu, Google and IBM have adopted big data and artificial intelligence to arrange manpower, recruit newcomers and forecast resignation [6-10].

Therefore, based on the data-driven method, the paper tries to study the intelligent transformation of HRM in the era of industry 4.0 and big data. Specifically, we focus on employee promotion. Promotion study not only can help workers obtain development opportunities, but also assist enterprises to select and retain talents [11,12]. Some experts found that employee promotion was affected by gender, age, education and communication patterns [13-16]. Also, nowadays, college students prefer to find jobs in big cities, and many employees want to be transferred to high-level institutions. They all believe that will be helpful for their promoting. Is this sense correct? Whether one's promotion can be influenced by the level of his organization? Can the location of the job reveal the worker's future? To look for the answer, the research studies the impacts of organization position on promotion, and attempts to forecast employee prospect with these findings.

In this paper, the position of an organization contains two aspects: position in geography and position in organizational structure. The research discovers the correlation between organization position and employee promotion, especially for better educated workers. Fig. 1 shows the architecture of the data-driven solution to learn promotion. At first, we get staff data from the HRM information database of the enterprise. Then features about the organization position are extracted from the dataset, while networks are constructed to learn staff transfers between different positions. Next, statistics is applied for correlation analysis, in order to preliminarily examine the relationship between positions and promotion. Based on the features, supervised learning is used to forecast employee prospect, and identify important factors. The good results of experiments verify the significance of position features and the effectiveness of the prediction model.

A few studies have found that the organization position affected individual development. M. Dickmann proposed that people could develop better in London where could have more opportunities [17]. S. Spilerman and T. Petersen discovered that organizational structure would affect regimes and the ceiling barrier of promotion [18]. However, these studies mainly use methods like questionnaires and interviews which have limitations in sample size and quantification explanation,

978-1-7281-4569-3/19/$31.00 ©2019 IEEE

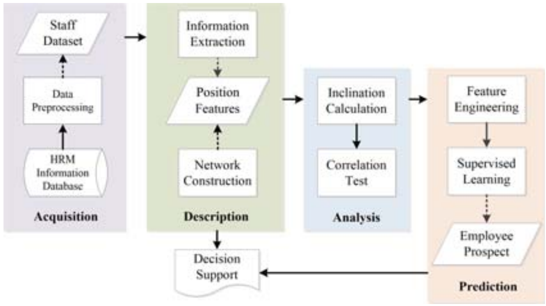

Figure 1. Architecture of the data-driven method

causing problems in applicability and accuracy. Traditional ways are not suitable for the fast-changing market [19].

Compared to questionnaires and interviews, data-driven methods are more objective and efficient. Recently, S. Fraiberger has found insights about artists' career trajectories with the help of big data [20]. T. Safavi has learned career transitions by network science and predict individual retention by machine learning [21]. H. Li forecasted career development by survival analysis and multi-task learning [22]. J. Yuan analyzed employee mobility by social networks and logistic regression [23]. However, none of these focus on the correlation between the position of organization and promotion. In this case, the paper discovers the significance of organization position and proposes a practical way to forecast employee promotion. Moreover, the study not only can help managers understand employees and distinguish talents timely, but also help individuals to make better career planning.

The rest of this paper is organized as follows. The second section describes the dataset. The third section introduces position features and their impacts on promotion. Details and results of experiments are illustrated in section IV. The last section summarizes the work of this paper and points out the future research direction.

II. Data Description

The data come from a state-owned enterprise in China. It contains personnel information of employees from 2006 to 2016, including gender, university, job title, department and so on. Moreover, we select employees who graduate from higher education colleges and major in science and engineering. Thus, employees in the staff dataset are knowledge staff. The dataset is consisted of 800,572 records which involve 34,391 people. Each record corresponds to an employee's details at that time, covering all changes and developments since he or she joined the company. After dropping records which contain missing data and abnormal values, information of 17,704 people is remained in the final staff dataset.

We calculate the increase of job rank for every employee, and the distribution is shown in Fig. 2. Negative value means the person leaves before being promoted. People who promote by only 1 rank are the most, which is followed by -1 and 2. However, less than 400 people have been promoted by 4 or more ranks. Therefore, we classified employees whose job rank increases by no less than 2 as 'promoted' employee (marked as '1'), while the others are classified as 'non-promoted' employee (marked as '0'). The statistics of the promotion status are shown in TABLE I.

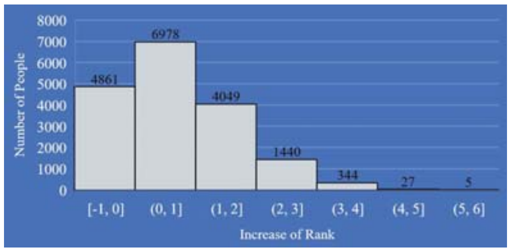

Figure 2. Distribution of rank growth

TABLE I. Statistics of Promotion Status

Promotion Status	Frequency	Percentage
1	5865	33.1%
0	11839	66.9%
Sum	17704	100%

III. The Position of Organization

A. Dimensions

The paper regards 'organization' as a social entity in which people work and live. Its scope is not fixed. Namely, an organization can be as large as a company, or as small as an office. The position of organization contains the geographical position and the structural position. The former one refers to the place where the organization locates, so we can describe the geographical position of an employee by the data about his working city. In this case, the scope of an organization is like a subsidiary or a branch company. Meanwhile, the latter one indicates the organization's role in the organizational structure, which can be depicted by the information about one's office. Then, we divide these two aspects of positions into six different dimensions. Each dimension has one or more related features. The six dimensions and their detailed features are listed in TABLE II.

B. Features

Set the staff dataset as $\mathbf{X}=[\mathbf{X}_1,\ldots,\mathbf{X}_i,\ldots,\mathbf{X}_{N1}]$, the number of employees is $|\mathbf{X}|=N_1$, and x_i represents the individual associated with $\mathbf{X}_i$. The data matrix of employee x_i is $\mathbf{X}_i=[\mathbf{v}_{i1},\ldots,\mathbf{v}_{ij},\ldots,\mathbf{v}_{iN2}]$, the number of fields in the dataset is $|\mathbf{X}_i|=N_2$. The change vector of x_i on the j-th field is $\mathbf{v}_{ij}=(u_{ij1},\ldots,u_{ijt},\ldots,u_{ijN3})$, where u_{ijt} is the value of the j-th field in the t-th time period. The number of elements in the change vector is $|\mathbf{v}_{ij}|=N_3$, and μ_{ij} is the set of elements in $\mathbf{v}_{ij}$.

1) Area

The 'area' has only one feature: main working area (F_{21}). This feature represents 'the area where the main organization of the employee is located'. It is calculated by:

$$\begin{cases} t_m = \underset{t}{argmax}\ max\left\{\psi_d\left(u_{ij_at}\right) \mid u_{ij_at} \in \mu_{ij_a}\right\} \\ F^{(i)} = u_{ij_a t_m} \end{cases}. \quad (1)$$

In (1), the j_a-th attribute is the associated field in staff dataset, the function ψ_d calculates the duration of value u.

Since personnel transfers rarely occur between different areas, changes of areas are not considered.

2) Administrative Division

The 'administrative division' has five features. There are

TABLE II. POSITION FEATURES

	Dimension	Feature	Symbol
Geographical Position	Area	Main working area	F_{11}
	Administra-tive Division	Initial administrative division	F_{21}
		Main administrative division	F_{22}
		Final administrative division	F_{23}
		The number of administrative divisions	F_{24}
		The change of administrative division	F_{25}
	Special Region	Initial special region	F_{31}
		Main special region	F_{32}
		Final special region	F_{33}
		The number of special regions	F_{34}
		The change of special region	F_{35}
Structural Position	Department	Initial department	F_{41}
		Main department	F_{42}
		Final department	F_{43}
		The number of departments	F_{44}
		The change of department	F_{45}
	Level	Initial organizational level	F_{51}
		Main organizational level	F_{52}
		Final organizational level	F_{53}
	Size	The size of main organization (larger)	F_{61}
		The size of main organization (smaller)	F_{62}

TABLE III. SUB-FEATURES ABOUT THE CHANGE OF POSITIONS

Sub-features	Symbol#
BC of initial position in the network	$F_{\tau 5a}$
BC of main position in the network	$F_{\tau 5b}$
BC of final position in the network	$F_{\tau 5c}$
ID of initial position in the network	$F_{\tau 5d}$
ID of main position in the network	$F_{\tau 5e}$
ID of final position in the network	$F_{\tau 5f}$
OD of initial position in the network	$F_{\tau 5g}$
OD of main position in the network	$F_{\tau 5h}$
OD of final position in the network	F_{k5i}
Ratio of OD and ID of initial position in the network	$F_{\tau 5j}$
Ratio of OD and ID of main position in the network	$F_{\tau 5k}$
Ratio of OD and ID of final position in the network	$F_{\tau 5l}$
The number of occurrences of transferring paths in the network	$F_{\tau 5m}$

In the symbols, variable τ={2,3,4}

five levels of administrative divisions in the dataset, from high to low levels are: province capital, prefecture-level city, county-level city, village and town, remote district. Main administrative division is the administrative division where an employee works the longest, while the initial one and the final one are the administrative division of the first and the last place where an employee work. F_{22} can be calculated by (1), and F_{21}, F_{23} can be obtained by (2), (3), respectively.

$$F^{(i)} = u_{i,j_a,1} \,, \tag{2}$$

$$F^{(i)} = u_{i,j_a,N_3} \,, \tag{3}$$

As people may work at organizations in different administrative divisions during career, the paper also considers the number of administrative divisions and the change of administrative division.

The number of administrative divisions (F_{24}) is calculated by (4).

$$F^{(i)} = \left|\mu_{ij_a}\right| . \tag{4}$$

Then, network is used to describe the changes in the organization's positions for employees within different promotion status. A node represents one position, and an edge indicates a transfer between the two positions. The weight of the edge means the frequency of such mobility. Specifically, several network attributes are used to construct sub-features about the change of position [24]:

- In-degree (ID). It is the number of edges from a node to its neighbors (considering weight).
- Out-degree (OD). It is the number of edges from a node's neighbors to itself (considering weight).
- Betweenness Centrality (BC). It is the number of shortest paths that pass through the node.

These sub-features are listed in TABLE III. The 'network' in the table refers to the 'whole' network.

Fig. 3 shows the network of administrative division for different employees. The size of the node is proportional to its in-degree, while the node's color is associated with the betweenness centrality. Nodes with deeper color are more important. In addition, the thicker the edge, the more frequent the mobility between the nodes. Although prefecture-level city occupies an important position in the networks, promoted employees are more likely to be transferred to province capital. Obviously, the most important mobility patterns in the whole sample are 'county-level city to prefecture-level city', and 'prefecture-level city to province capital'. The transition from remote district to province capital is common for promoted employees, but non-promoted employees tend to move between organizations located in province capital, county-level city as well as villages and towns.

3) *Special Region*

The 'special region' has five features. There are three types of regions in the dataset: undeveloped regions (i.e. border region, poverty region, and race region), special zones (i.e. special economic zone, Macao, and Hong Kong), and non-special regions. Main special region (F_{32}) is the region where an employee stays the longest, which is defined by (1). Moreover, initial region (F_{31}) and final region (F_{33}) are the first and last region where an employee works, which are calculated by (2), (3), respectively. The number of special regions (F_{34}) is calculated by (4).

Besides, we also use network to describe the changes in special region. Associated sub-features are created according to TABLE III. Fig. 4 shows the networks of special region for different workers. Among all the regions, non-special region is the most important. Many people go to undeveloped regions then return to non-special regions. In particular, the mobility of promoted employees is more frequent in organizations located in non-special region and border region 1.

4) *Department*

The 'department' also has five features. It illustrates which department of the company the employee's organization (e.g. office, workshop) belongs to. These departments, located in different positions in the organizational structure, are of various levels, sizes, and duties. Main department (F_{42}) is the department where an individual works for the most time, which is defined by (1). Besides, initial department (F_{41}) and final department (F_{43}) are constructed by (2), (3). The number of departments (F_{44}) is designed by (4).

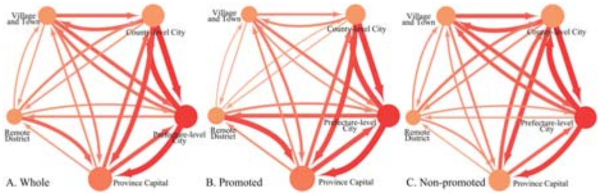

Figure 3. Networks of administrative division

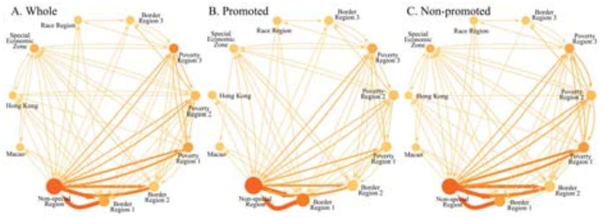

Figure 4. Networks of special region

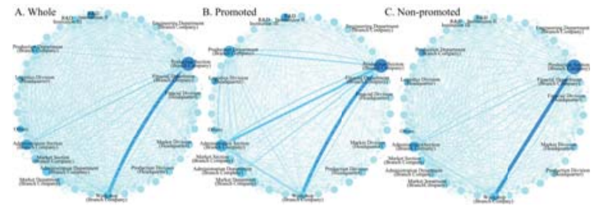

Figure 5. Networks of department

Associated sub-features of the change of department are constructed based on TABLE III. Fig. 5 shows the networks of department. The production section of branch company is the department with the largest inflow of personnel, and it is the most critical node in departmental transfer. The administration section of branch company also has a vital position. Compared to non-promoted people, promoted employees are more likely to be exchanged between the production section of branch company, the production department of branch company, and the administration department of branch company. In addition, R&D institution II plays a more important role in promoted group.

5) Level

The 'level' has three features. In the enterprise, every organization has a predefined level, ranging from level 1 (the lowest) to level 10 (the highest), revealing its authority and scope. Level 1, 3, and 7 are with the largest proportion of employees. The initial organizational level (F_{51}), the main organizational level (F_{52}), and the final organizational level (F_{53}) are constructed according to (2), (1), (3), respectively.

6) Size

The 'size' has two features. In the paper, the organizational structure of the Chinese state-owned enterprise is pyramid, so the size of an organization can reflect its structural position. The size of the main larger organization (F_{61}) is the number of people in the branch company that the employee spends the most time. Similarly, the size of the main smaller organization (F_{62}) is the number of people in the longest-serving office or workshop of an employee. They are calculated by:

$$\begin{cases} t_m = \underset{t}{argmax}\ max\left\{\psi_d\left(u_{ij_d t}\right) \mid u_{ij_a t} \in \mu_{ij_a}\right\} \\ F^{(i)} = \psi_p(u_{ij_a t_m}) \end{cases}, \quad (5)$$

where the function ψ_p calculates the number of people of an organization.

C. Correlation

To further discover the correlation between categorical features and promotion, an inclination index ρ is designed to represent the deviation between actual and expected situations for each promotion status. The inclination of category value f_c in feature f under promotion status s is defined as:

$$\rho^{s}_{f,f_c} = P\left(f_c \mid status = s\right) - P\left(f_c\right),\ s \in \{0,1\}\ . \quad (6)$$

P denotes the probability. If $\rho>0$, staff in the category f_c tends to be status s; otherwise, people with f_c are less likely to be s.

The inclination of main working area is shown in Fig. 6. In general, staff of organizations in economically and politically developed areas, such as Beijing and Guangdong, have more opportunities to be promoted.

The inclination of different administrative divisions is shown in Fig. 7. It seems that the staff in organizations located in provincial capital develop very well. But the promotion of employees working in middle-level areas is poor.

The promotion inclination for staff in different special regions is shown in Fig. 8. People in non-special regions develop much better than those who work in the organizations located in undeveloped regions. Additionally, the beginning of a career seems to have a more significant impact.

Fig. 9 shows the promotion inclination of department. The five departments with the largest number of people are chosen as representatives. Staff in the workshop of branch company is poorly developed, while workers in the administration section of branch company are promoted much higher. Moreover, the

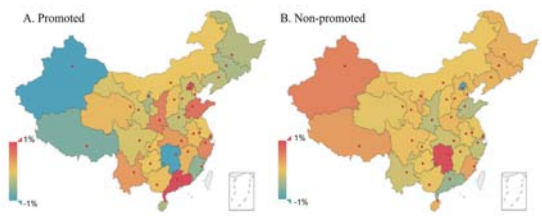

Figure 6. Inclination of areas

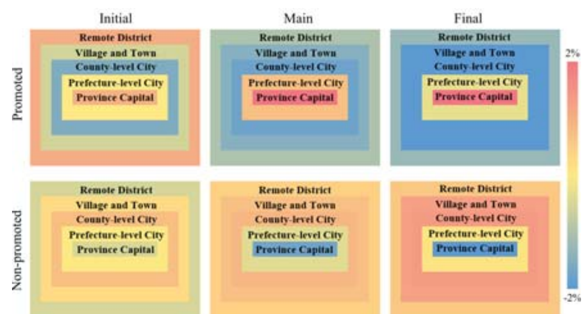

Figure 7. Inclination of administrative divisions

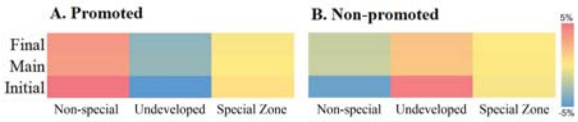

Figure 8. Inclination of special regions

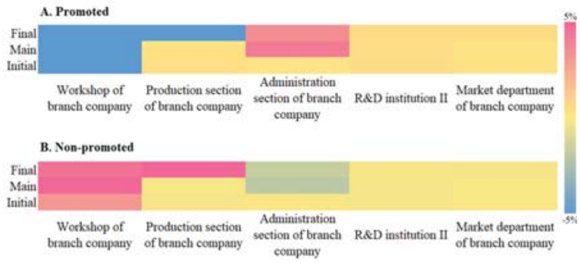

Figure 9. Inclination of major departments

Feature	Pearson	Feature	Pearson	Feature	Pearson	Feature	Pearson	Feature	Pearson	Feature	Pearson
F_{24}	0.249	F_{25h}	-0.006	F_{35b}	0.057	F_{35j}	-0.147	F_{45d}	-0.009	F_{45l}	-0.228
F_{25a}	0.005	F_{25i}	-0.031	F_{35c}	0.057	F_{35k}	-0.049	F_{45e}	0.001	F_{45m}	-0.310
F_{25b}	0.029	F_{25j}	-0.001	F_{35d}	0.121	F_{35l}	0.019	F_{45f}	-0.273	F_{51}	0.106
F_{25c}	0.030	F_{25k}	-0.056	F_{35e}	0.053	F_{35m}	0.083	F_{45g}	-0.089	F_{52}	0.318
F_{25d}	-0.010	F_{25l}	-0.103	F_{35f}	0.060	F_{44}	0.364	F_{45h}	-0.304	F_{53}	0.406
F_{25e}	0.055	F_{25m}	-0.116	F_{35g}	0.117	F_{45a}	-0.018	F_{45i}	-0.376	F_{61}	-0.053
F_{25f}	0.092	F_{34}	0.126	F_{35h}	0.053	F_{45b}	-0.035	F_{45j}	-0.086	F_{62}	-0.116
F_{25g}	-0.023	F_{35a}	0.140	F_{35i}	0.061	F_{45c}	-0.287	F_{45k}	-0.318		

-0.3 0.3

Figure 10. Pearson correlation coefficient of numerical features

promotion of R&D staff is relatively stable.

The inclination of categorical features to every promotion status has been analyzed above. For numerical features, we use Pearson correlation coefficient to find their inclination. The result is shown in Fig. 10. The number of administrative divisions, the number of departments, main organizational level, and final organizational level have strongly positive correlations with promotion. However, the OD of employee's main department, the OD of employee's final department, the ratio of OD and ID of employee's main department, and the number of occurrences of employee's mobility path on department have great negative correlations with promotion.

In addition, we use Chi-Square test (for categorical features) and Analysis of Variance (for numerical features) to learn whether these features are significantly correlated with promotion. P-values of most features are less than significant level α=0.05, except F_{25a}, F_{25d}, F_{25h}, F_{25j}, F_{45d} and F_{45e}. According to the result, the position of organization has significant influence on employee promotion.

IV. PREDICTION

A. Problem Statemen

The aim of the prediction is to estimate employee prospect and identify potential staff. Also, it attempts to validate the effects of organization position on promotion, distinguishing important factors. That can be achieved by solving the following problem:

- Given an employee x_i, forecast the prospect of x_i based on the position of his or her organization.

As the general believes that workers who go higher and farther in the enterprise have brighter future, the prospect of an employee can be determined by their ability to promote. Thus, we take promotion status as the target label of prediction. When an employee's promotion status is predicted to be '1', his or her prospect is good. On the contrary, if the promotion status is predicted to be '0', the staff's prospect is poor. In that case, forecasting employee prospect becomes a binary classification problem on imbalanced dataset. Moreover, factors that affect promotion strongly often play a greater role in prediction. Hence, we measure the importance of features by their predictive ability.

B. Methods

Supervised learning is adopted to forecast employee promotion. We choose logistic regression (LR) [25], random forests (RF) [26] and AdaBoost (AB) [27] to construct models. They are efficient in many situations and have solved numerous prediction problems. The experiment of promotion prediction is carried out within Python. At first, we extract features and construct feature vectors. The feature set consists of all the features about the position of organization, i.e. F_1 to F_6. Five-fold cross-validation is used to split training set and test set, avoiding random factors. Synthetic minority over-sampling technique (SMOTE) [28] is adopted to deal with class-imbalance problem. Grid search is chosen to adjust hyper parameters, determining the best classifier. In order to explore the predictive capability of different models and features, we conduct experiments for several times on features of each dimension and by every classifier. Moreover, we identify the most influential factors through a feature selection strategy. The performances of models are evaluated by accuracy, area under curve (AUC), recall and precision. Specifically, the AUC is calculated by trapezoidal rule.

C. Experiments

At first, we test the effectiveness for each dimension under every classification model, and results are shown in Fig. 11. For models, RF outperforms the others. LR is superior to AB and has high recall. For features, the performances of F_4 and F_5 are the best, while that of F_1 is the worst. The recall of special region is the highest, indicating its outstanding ability to find out all the promoted staff. However, the prediction results are not very ideal as the accuracy of most of them is less than the baseline (0.669). Thus, we combine different dimensions to improve the prediction ability of the model.

TABLE V is the result of prediction on combined features. The accuracy of most combinations is much greater than the baseline. It is obvious that combination can effectively improve the prediction ability of models. Particularly, the accuracy, AUC, recall and precision of all these feature sets have positive growth, except 'F_1+F_3', 'F_1+F_4', 'F_1+F_6', 'F_2+F_3'. It indicates that structural position plays a more critical role than geographic position. Moreover, the '*ALL*' group which contains all the features performs the best. Therefore, we carry out the next experiment based on it.

After one-hot encoding, the number of features increased dramatically, so we design a multi-method combination voting strategy for feature selection. Four feature selection methods are chosen: variance filtering, F-value filtering, L1 regularization, and recursive feature elimination (RFE). Each

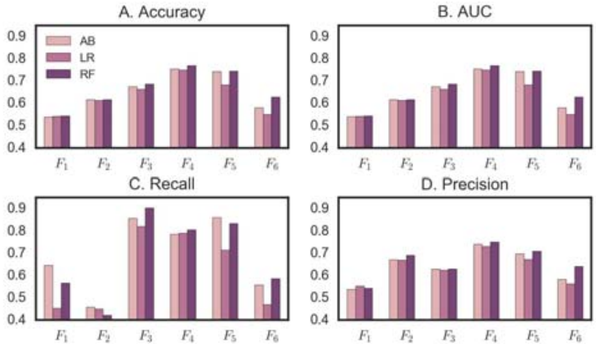

Figure 11. Prediction result of each dimension

TABLE IV. PREDICTION RESULT OF DIMENSION COMBINATION

Feature	Accuracy	AUC	Recall	Precision
F_1+F_2	0.643*	0.643*	0.617**	0.651
F_1+F_3	0.685	0.685	0.821(-)	0.645
F_1+F_4	0.770*	0.770*	0.821(-)	0.745*
F_1+F_5	0.744**	0.744**	0.817**	0.713*
F_1+F_6	0.589	0.589	0.540(-)	0.599
F_2+F_3	0.689	0.689	0.815**	0.651(-)
F_2+F_4	0.776*	0.776*	0.817**	0.756
F_2+F_5	0.744*	0.744*	0.805	0.718
F_2+F_6	0.663*	0.663*	0.624**	0.676
F_3+F_4	0.800*	0.800*	0.880*	0.759*
F_3+F_5	0.778*	0.778*	0.848	0.744*
F_3+F_6	0.699*	0.699*	0.781*	0.673*
F_4+F_5	0.782	0.782	0.825	0.760
F_4+F_6	0.775**	0.775**	0.828**	0.749*
F_5+F_6	0.748*	0.748*	0.805**	0.723*
$F_1+F_2+F_3$	0.708*	0.708*	0.812**	0.680*
$F_4+F_5+F_6$	0.797**	0.797**	0.849**	0.769*
ALL	0.832**	0.832**	0.877**	0.805**

* The amelioration/deterioration is greater than 0.05.

** The amelioration/deterioration is greater than 0.1.

method can vote for the features in its selected feature subset. Specifically, RFE involves three different models and has three votes. Features with more than half votes are retained, while the rest are deleted.

The voting result is shown in Fig. 12. Sixteen features with six votes are of high importance. They are F_{24}, F_{25g}, F_{25m}, F_{31} (province capital), F_{34}, F_{35a}, F_{35j}, F_{35m}, F_{44}, F_{45i}, F_{45m}, F_{51}, F_{52}, F_{53}, F_{61}, F_{62}. It can be concluded that the richness of organizational experience, the work experience in depressed areas, the change of position, organizational level and size have strong impacts on employee promotion. Referring to the correlation analysis in section III, we guess that people who experience more positions tend to be promoted. Moreover, both high organizational level and small organizational scale are helpful for promotion. Besides, promoted employees initially prefer to work in regions with less outflow or high importance, and they eventually tend to be transferred to departments with fewer leavers. The uniqueness of mobility path also contributes to better prospects.

In Fig. 13, the performances of models with and without feature selection are compared. The AUC and accuracy almost coincide. Feature selection does not significantly ameliorate or deteriorate the prediction ability, but it greatly improves the speed of the algorithm, making the prediction more efficient.

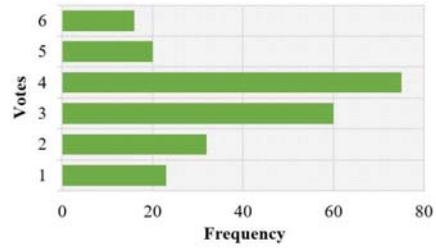

Figure 12. Vote result

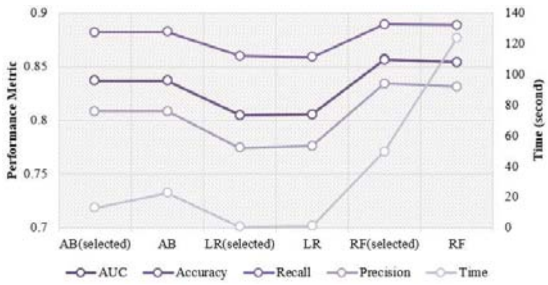

Figure 13. Comparison of models with/without feature selection

Besides, RF outperforms the others and its time consumption is also acceptable. Therefore, we use the RF classifier based on feature set '*ALL*' and multi-method combination voting strategy as the final prediction model, with accuracy and AUC of 0.856, recall of 0.889, precision of 0.834. In the similar research from J. Yuan, the recall and precision of promotion prediction are with a synthesized value ($F1$ score) of about 0.85 [23], while that of our experiment is 0.861. Thus, the proposed data-driven method has good performance and is not inferior to the other published method.

V. DISCUSSION AND CONCLUSION

The paper puts forward ideas on data-driven solution to the issue of promotion, and made preliminary practice, trying to provide insights for promotion study and even intelligent HRM study in the era of big data and industry 4.0. Statistics, networks are adopted to analyze the correlation between the position of organization and promotion, and then machine learning is used to forecast employee prospect. From the results, we can conclude that:

- There are obvious personal advantages on experience in staff promotion. The number of experienced positions and the uniqueness of experience have strong impacts on promotion.
- There are obvious regional differences in staff promotion. For example, employees that initially work in province capital or developed areas tend to have better prospect.
- There are obvious departmental tendencies in staff promotion. For instance, employees in administration sections have more promotion chances, while workers in production section are difficult to be promoted.

To sum up, working in positions where mobility is more stable, where resource is more abundant, or where special experience is available, will contribute to promoting. Thus, we believe that people should work in different regions and departments to enrich their experiences. They also should seize the opportunity to work in high-level organizations and big cities. For managers, they must give employees with different geographical and structural positions the same chances for development, ensuring organization fairness. In this way, organizational commitment and job satisfaction of employees can be guaranteed [29], achieving performance improvement and talent retention.

Besides, the data-driven approach is more objective and efficient than traditional methods. From the experiments, the position of organization influences employee prospect significantly, and the proposed model is effective, which can adapt to the demands of HRM under industry 4.0. Enterprises can identify and nurture potential employees based on the findings, enhancing the human capital of the organization. Meanwhile, employees can realize their own value more clearly and better serve the enterprise. Admittedly, this approach still has some limitations. For example, the definitions of 'promoted' and 'non-promoted' employees are subjective, and the prediction costs a bit long time.

In the future, we plan to find multiple ways to measure promotion. Also, we will enrich the experiments by making more comparisons between different data-driven methods. On top of this, more factors and the state-of-art computing techniques will be added to ameliorate our data-driven solution, providing more powerful decision support for HRM.

REFERENCES

[1] S. Shamim, S. Cang, H. Yu, and Y. Li, "Management approaches for Industry 4.0: A human resource management perspective," in *Proc. 2016 IEEE Congress on Evolutionary Computation*, Vancouver, 2016, pp. 5309-5316.

[2] D. Trotta and P. Garengo, "Industry 4.0 key research topics: A bibliometric review," in *Proc. 7th International Conf. Industrial Technology and Management*, Oxford, 2018, pp. 113-117..

[3] F. Hecklau, M. Galeitzke, S. Flachs, and H. Kohl, "Holistic approach for human resource management in Industry 4.0," *Procedia CIRP*, vol. 54, pp. 1-6, 2016.

[4] E. A. Lee, "Cyber Physical Systems: Design Challenges," in *Proc. 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing*, Orlando, 2008, pp. 363-369.

[5] L. Prifti, M. Knigge, H. Kienegger, and H. Krcmar, H. "A Competency Model for" Industrie 4.0" Employees," in Proc.13th International Conf. on Wirtschaftsinformatik, St. Gallen, 2017, pp. 46-60.

[6] H. Lin, H. Zhu, Y. Zuo, C. Zhu, J. Wu, and H. Xiong, "Collaborative company profiling: Insights from an employee's perspective," in *Proc. 31st AAAI Conf. Artificial Intelligence*, San Francisco, 2017, pp. 1417-1423.

[7] C. Zhu, H. Zhu, H. Xiong, C. Ma, F. Xie, P. Ding, and P. Li, "Person-Job Fit: Adapting the Right Talent for the Right Job with Joint Representation Learning," *ACM Transactions on Management Information Systems*, vol. 9, no. 3, article 12, 2018

[8] T. Xu, H. Zhu, C. Zhu, P. Li, and H. Xiong, "Measuring the popularity of job skills in recruitment market: A multi-criteria approach," in *Proc 32nd AAAI Conf. Artificial Intelligence*, New Orleans, 2018, pp. 2572-2579.

[9] D. Wei, K. Varshney, and M. Wagman, "Optigrow: People analytics for job transfers," in *Proc. 2015 IEEE International Congress on Big Data*, New York, 2015, pp. 535-542.

[10] R. Horesh, K. Varshney, and J. Yi, "Information retrieval, fusion, completion, and clustering for employee expertise estimation," In *Proc. 2016 IEEE International Conf. Big Data*, Washington D.C., 2016, pp. 1385-1393..

[11] H. Lai, "Study on influence of employee promotion system on organizational performance," *International Journal of Organizational Innovation*, vol. 5, no. 1, pp. 231, 2012.

[12] P. Nguyen, C. Dang, and L. Nguyen, "Would better earning, work environment, and promotion opportunities increase employee performance? An investigation in state and other sectors in Vietnam," *Public Organization Review*, vol 15, no. 4, pp. 565-579, 2015.

[13] F. Blau, and J. DeVaro, "New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires," *Industrial Relations: A Journal of Economy and Society*, vol. 46, no. 3, pp. 511-550, 2007.

[14] S. Adams, "Passed over for promotion because of age: an empirical analysis of the consequences," *Journal of Labor Research*, vol. 23, no. 3, pp. 447-461, 2002.

[15] S. Spilerman, and T. Lunde, "Features of educational attainment and job promotion prospects," *American Journal of Sociology*, vol. 97, no. 3, pp. 689-720, 1991.

[16] A. Woolley, C. Chabris, A. Pentland, N. Hashmi, and T. Malone, "Evidence for a collective intelligence factor in the performance of human groups," *Science*, vol. 330, no. 6004, pp. 686-688, 2010.

[17] M. Dickmann, and T. Mills, "The importance of intelligent career and location considerations: exploring the decision to go to London," *Personnel Review*, vol. 39, no. 1, pp. 116-134, 2009.

[18] S. Spilerman, and T. Petersen, "Organizational structure, determinants of promotion, and gender differences in attainment," *Social Science Research*, vol. 28, no. 2, pp. 203-227, 1999.

[19] S. Mu, *New Logic of Human Resource Management*, Beijing: Xinhua Publishing House, 2015.

[20] S. Fraiberger, R. Sinatra, M. Resch, C. Riedl, and A. Barabási, "Quantifying reputation and success in art," *Science*, vol. 362, no. 6416, pp. 825-829, 2018.

[21] T. Safavi, M. Davoodi, and D. Koutra, "Career Transitions and Trajectories: A Case Study in Computing," In *Proc. 24th ACM SIGKDD International Conf. Knowledge Discovery and Data Mining*, London, 2018, pp. 675-684.

[22] H. Li, Y. Ge, H. Zhu, H. Xiong, and H. Zhao, "Prospecting the career development of talents: A survival analysis perspective," In *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, 2017, pp. 917-925.

[23] J. Yuan, Q. Zhang, J. Gao, L. Zhang, X. Wan, X. Yu, and T. Zhou, "Promotion and resignation in employee networks," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 442-447, 2016

[24] G. Chen, X. Wang, and X. Li, *Introduction to Complex Networks: Models, Structures and Dynamics*. Beijing: Higher Education Press, Beijing, 2012.

[25] D. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society*, pp. 215-242, 1958.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[27] Y. Freund, R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proc. European Conf. Computational Learning Theory*, Barcelona, 1995, pp. 23-37.

[28] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.

[29] A. McDowall, and C. Fletcher, "Employee development: an organizational justice perspective," *Personnel Review*, vol. 33, no. 1, pp. 8-29, 2004.