A Novel Data Mining Approach towards Human Resource Performance Appraisal

Pei Quan[1, 2] [0000-0001-6005-5714], Ying Liu[1, 2], Tianlin Zhang[1, 2], Yueran Wen[3], Kaichao Wu[4], Hongbo He[4], Yong Shi[2,5,6,7]

[1] School of Computer and Control, University of Chinese Academy of Sciences, Beijing, 100190 China
[2] Key Lab of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing, 100190 China
[3]School of Labor and Human Resources, Renmin University of China, Beijing, 100872 China
[4] Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100190 China
[5] School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190 China
[6] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, 100190 China
[7] College of Information Science and Technology, University of Nebraska at Omaha, NE, 68182 USA

quanpei17@mails.ucas.ac.cn

Abstract. Performance appraisal has always been an important research topic in human resource management. A reasonable performance appraisal plan lays a solid foundation for the development of an enterprise. Traditional performance appraisal programs are labor-based, lacking of fairness. Furthermore, as globalization and technology advance, in order to meet the fast changing strategic goals and increasing cross-functional tasks, enterprises face new challenges in performance appraisal. This paper proposes a data mining-based performance appraisal framework, to conduct an *automatic* and *comprehensive* assessment of the employees on their working ability and job competency. This framework has been successfully applied in a domestic company, providing a reliable basis for its human resources management.

Keywords: Performance Appraisal, Data Mining, Enterprise Strategy, Job Competency.

1 Introduction

The six modules of human resources: recruitment, configuration, training, development, performance management, compensation and benefit management, are interconnected. Among them, performance management is the core in practical businesses.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

With performance management, companies can reward and punish good or bad performance, and implement performance-based wages. Businesses can also identify weaknesses and deliver targeted training with proper performance management. Based on specific circumstances of internal and external recruitment, they can also achieve better matchings of positions and employees. Thus, a performance appraisal system which meets the requirement of enterprise strategic goals and current market conditions can fully release the potential of employees, and greatly mobilize their enthusiasm for the overall business development.

In practice, most employee performance appraisal approaches follow the traditional *manual* method for evaluation and supervision. It is very labor intensive, incomprehensive and unfair in domains where work is difficult to quantify, as well as large companies with thousands of employees and many departments. Therefore, the results of performance appraisal are not accurate, and cannot achieve the expectations. In addition, the market and policies of enterprises are changing rapidly, and their strategic objectives are also being constantly adjusted. Dynamically evaluating the relationship between actual work and strategic goals, and establishing real-time performance appraisal system are urgent problems in human resources management. In addition, with the development of society, the complexity of work is getting higher and higher, and job competition is becoming more intense. Thus, it is difficult to solve problems completely through employees' inherent knowledge. Therefore, it is necessary to *automatically* evaluate workability of staffs, based on the actual requirements of positions and the development of the employees. It is very useful to supervise the continuous growth of employees, as a basis for training and staffing.

In this paper, we use data mining algorithms to solve the above problems. The main contributions of our work include two aspects: **work performance** and **job competency**. We propose an automatic, comprehensive and fair performance appraisal framework which meets the strategic objectives of the enterprise and the needs of the market. Firstly, through text analysis of plans and summaries in the employee's work report, and the strategic objectives of the enterprise, the **work performance** of the employees can be evaluated from three aspects: job value, executive ability and content of the report. In the evaluation of **job competency**, the competency model of positions is extracted from the competency requirements of the job, and match with external knowledge sources such as books, images and other information in the internal knowledge base. Our model will automatically generate questions from the above core concepts. By investigating employee's answers, we can evaluate their job competency.

Currently, this performance appraisal framework has been highly recognized by human resources experts and has been widely used by thousands of employees at Company H and Company J. In addition, Company H is one of the largest high-tech companies in China. In practical application, this framework plays a role in encouraging staff to work actively and speeding up the realization of corporate strategic objectives, and contributes to the employee assessment and personnel adjustment.

The paper is organized as follows: Section 2 provides related work and backgrounds of human resource performance evaluation and data mining algorithms. Section 3 presents our methodology. Section 4 discusses implementation details and experiment results. Section 5 summaries this paper.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

2 Related Work

In the field of **performance appraisal**, it is generally difficult to have a comprehensive assessment of staff performance. Various performance appraisal methods have their own advantages and disadvantages. Therefore, the study of personnel performance appraisal theory still needs to be further improved, especially in fitting performance appraisal methods to be in line with actual needs. At present, the main research methods are as followed.

Key Performance Indicators (KPIs) are one of the most commonly used methods[1][2]. They are the key factors that determine the effectiveness of a business strategy. They turn a business strategy into internal processes and activities, and continuously strengthen the key competitiveness of enterprises and achieve high returns. The KPI method is based on annual target, combined with analysis of employee performance differences, and then periodically agreed on the key quantitative indicators of enterprises, ministries and individuals to build performance appraisal system.

360 degree assessment method is a more comprehensive performance evaluation method, also known as comprehensive evaluation method, with a wide range of sources of assessment results, and multi-level features[3]. 360 degree, as the name implies, refers to an all-round evaluation of employee performance. In terms of examiners, they include internal and external customers, as well as superior leaders, colleagues, subordinates, and employee themselves. The specific implementation process can be summarized as following: Firstly, the employees listen and fill out the questionnaire. Then, the managers evaluate the performance of different aspects of performance. When analyzing and discussing the assessment results, the two sides have conducted a full study and discussion to formulate the performance targets for the next year. The advantage of this method is to break the traditional way of superior evaluation of subordinates. It can avoid the phenomenon of "halo effect", "center trend", "personal prejudice and check blind spot" which is very common for the examiner in the traditional evaluations.

Date mining methodologies have been developed for exploration and analysis, by automatic or semi-automatic means, of large quantities of data to discover meaningful patterns and rules[4]. Indeed, such data including employees' seldom used data and work summary can provide a rich resource for knowledge discovery and decision support. Therefore, data mining is discovery-driven, not assumption-driven. Data mining involves various techniques including statistics, neural networks, decision tree, genetic algorithm.

Data mining has been applied in many fields such as marketing[5], finance[6], traffic[7], health care[8], customer relationship management[9], and educational data mining[10]. However, data mining has not been used well in human resource management. In particular, Chien C F [11] used data mining in the high-technology industry to analyze the ability of employees to improve personnel selection and enhance the quality of employees.

With the gradual development of data mining and text analysis, more and more fields apply data mining algorithms on domain specific data analysis, and gain positive results. For example, Tang et al. employ a multiview privileged SVM model to exploit complementary information among multiple feature sets, which can be an interesting

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

future direction for our work, as we process data from multiple sources [22]. However, there are few cases which combine performance evaluation and data mining at present. Therefore, this paper proposes a novel comprehensive performance appraisal framework based on data mining and text analysis, which combines a employees' work performance, corporate strategic objectives and position competence. It provides a promising way for human resource management.

3 Methodology

This paper constructs an automatic framework for human resource data mining to evaluate the employees’ work from their work summary and self-improvement. As the main contribution and novelty of our work, we extensively apply NLP and data mining technologies to areas of work performance, job competency and self-growth material recommendation. Under our methodology, working ability and job competencies could be quantified and the decision makers can have an easier and better understanding on employees’ comprehensive ability. The evaluation results can be used to effectively adjust enterprise position structure reasonably and improve matching of staff and posts. The performance appraisal framework is shown in Fig.1.

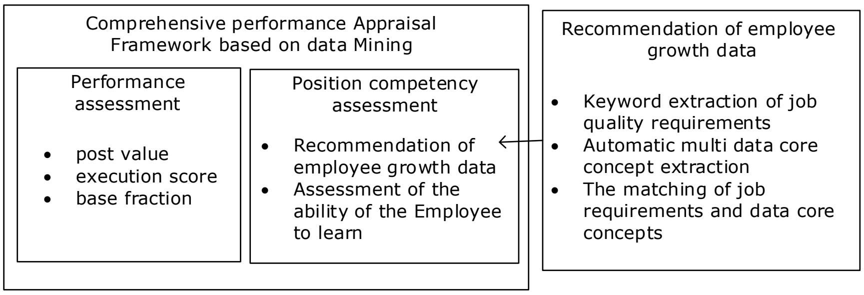

Fig. 1. The performance appraisal framework

3.1 Assessment of Work Performance of Employee Based on Text Analysis

Each employee submits a job report periodically, including the company's strategic objectives, the employee's expected plan, and a summary of the employee's actual work during that period. Since each report submitted is reviewed by the manager of the employee, the reliability of the report's content can be guaranteed. Therefore, our framework applies text analysis on the employee's work reports, and conducts analysis on the position value, the execution score and the basic score, and thus obtains the employee's work performance result. The specific assessment is as follows:

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

3.1.1 Position Score

The most intuitive manifestation of the value of an employee is the impact of his/her work on the strategic goals of the organization. Therefore, we correlate the work plan in the employee's work report with the strategic objectives of the enterprise. The two sources of paragraph text are firstly divided into words by CRF segmentation method. Since sentences often contain "stop words" that appears frequently but not semantically relevant (e.g. is, this, etc.), in this work we remove such words. In addition, Chinese expression is abundant, and synonyms are often used to describe the same thing. We use a Chinese synonym dictionary, and transform semantically similar words into the same form. Finally, we identify similar documents based on a set of common keywords. We employ cosine similarity [12,13] commonly used in text analysis, to characterize the correlation between two segments of text. The formula for calculating post value based on cosine similarity is as follows.

$$Position\_Score=sim(v_1,v_2)=\frac{v_1\cdot v_2}{|v_1||v_2|} \tag{1}$$

Where $v_1\cdot v_2=\sum_{i=1}^{t}v_{1i}v_{2i}$, $|v_1|=\sqrt{v_1\cdot v_1}$, v is a word vector used to describe the content of a passage by word segmentation and removal of stop words. The higher the value of Position_Score, the higher the correlation between the two paragraphs.

3.1.2 Execution Score

From the managers' perspective, their most important concern is the ability of their employees to perform their work. The stronger the execution, the better the employees are considered to be. Therefore, the execution ability is also an important evaluation index in performance appraisal. In our work, the performance of each employee is automatically measured by analyzing the matching degree of the work plan in the employee's work report and his actual work summary. First of all, similar to the above method, we divide the employees' plans and summaries into participles, remove the stop words, and then get the key vectors of the original sentences.

$$Execution\_Score=\frac{\sum_{i=1}^{t}F(i)}{m} \tag{2}$$

Here $F(i)$ is the completion of each plan. Based on the different degree adverbs identified in the summary, each program is assigned a discount ratio for varying degrees, which is provided by the domain experts. The detailed scores are shown in Table 1, where m is the total number of plans listed by the employees.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

Table 1. Discount ratio of different adverbs of degree comparison table

Adverbs of degree	Discount ratio
{基本完成,初步完成,大体上,几乎完成} (almost done)	0.8
{未完成,尚未,没有完成,有待完成} (Not yet)	0.6

3.1.3 Basis Score

In addition to the two aspects of the above assessment, the quality of the employee's report should also be evaluated. Through analysis of the employees' plan and summary after the participle, the sentence that lacks predicate is regarded as the residual sentence, and we use the total number of the residual sentences in the report to evaluate the employee. Employees who have few words or who copy the same content from the plan are assigned lower scores.

3.1.4 Total Score of Work Performance

The score of the above parts are summed up the following formula (3):

$$Work\_Score = \alpha \cdot Position\_Score + \beta \cdot Execution\_Score + (1-\alpha-\beta) \cdot Basis\_Score \quad (3)$$

The values of α and β denote the weights of the position value and the execution scores respectively. The values of α and β are set according to the actual situation of different companies, which are company-specific. For example, Company J wants to assess the ability of employees, but also encourages employees to better complete tasks in line with the strategic objectives of the enterprise, so the value of α and β will both be set to high values of 0.4.

3.2 Assessment of Employee Job Competency

As globalization and technology advance, the working procedures in companies are becoming diversified and complicated, and cross-functional tasks are also increased while new jobs are still constantly created. For employees, the ability of self-improvement is especially important. Therefore, based on position characteristics and requirements of employees, our work selects the most suitable data from the internal databases and external data sources for employees to meet their job requirements. Through analysis of the learning behavior of employees, we evaluate the employees' job competency.

3.2.1 Automatic Multi-source-data Core Concept Extraction

In order to improve the ability to work, and face the complex tasks, employees have to continuously learn knowledge from internal databases and external data sources. It is

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

very important to obtain the core content of each material and generate a reasonable summary for each source quickly and efficiently, for the growth and progress of employees. Here, we employ a combination of TF-IDF algorithm and TextRank algorithm (based on graph model) to automatically extract data [14]. The algorithm can be described as a three-step process including sentence representation, ranking, and selection. The following paragraphs will describe each of the steps [15,16].

Sentence representation

In the TextRank algorithm, it is impossible to process plain text information directly. Therefore, each sentence must be transformed into the weight vector of the word, and then TextRank could be carried out by the similarity between each sentence vector. When converting to sentence weight vector, one possible approach would be to only count the number of occurrences of the term in the sentence, but that will give usual term preference over unusual terms, even if unusual terms often defines a text better than the usual terms that most text contains. To account for this, the frequency of a term is weighted with the inverse document frequency (IDF). The purpose of IDF is to boost the value of rare terms[17]. This is done by taking the logarithm of the number of documents N in the given corpus divided by the number of documents that contains a given term nt.

$$\log \frac{N}{n_t} \tag{4}$$

The IDF-score will be high for a term if it is only present in a small number of documents in the corpus. The IDF-score is combined with the term frequency (TF) to give the so-called TF-IDF score. The TF-IDF for a given term t, document d and corpus D, is defined as:

$$tf - idf(t, d, D) = tf(t, d) \cdot idf(t, D) \tag{5}$$

Through the calculation of TF-IDF, we attach an initial weight to each term in the sentence. So the input text is represented as a graph, where each sentence is converted to a node where an edge between two nodes denotes the similarity between the two sentences.

Sentence ranking

After the sentence weight initialization, we proceed to calculate the importance of each sentence in the whole text through an iterative way [18,19]. The specific iterative process is shown as follows in (6):

$$WS(V_i) = \frac{1-d}{n} + d * \sum_{V_j \in In(V_i)} \frac{w_{ij}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \tag{6}$$

Here, $WS(V_i)$ denotes the weight of sentence i, $\sum_{V_k \in Out(V_j)} w_{jk}$ denotes the contribution of each adjacent sentence. w_{ij} denotes the similarity between sentence i and sentence j, while $WS(V_j)$ denotes the weight of sentence j in the last iteration. The initial weight

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

of array WS is 1/n, where n is the total number of sentences in the passage. d is a damping coefficient in a range of 0 to 1, denoting a probability of pointing to other arbitrary points from a particular point in the graph, and the general value is set at 0.85.

Sentence selection

The last step is to select which sentences to be extracted as the summary. In this case, we select N sentences with the highest scores. The specific value of N is selected in section 4 through specific experimental results.

Also, as books are more structured than plain text, the title of each chapter is often closer to the subject of the paragraph than other sentence. Therefore, we enhance the weight of different sentences based on the title of the book when initializing the weight of each sentence, so as to achieve the purpose of highlighting the topic. The specific lifting effect will be shown in the section 4.

In addition, external data sources and internal databases contain a large number of images, video and other information. We extract metadata to obtain the text description, and then use the same way to process the multi-source-data core concept extraction.

3.2.2 Intelligent Matching of Job Requirements and Learning Materials

After extracting core concepts of multi-source-data, we next consider how to recommend the most suitable learning materials for employees in different positions. First of all, through the analysis of position requirements of our competency model, a set of widely recognized job function requirements in the field of human resources is described, and the key words of quality requirements of different positions are obtained. Here we use the BM25 information retrieval model[20], with the formula (7).

$$RSV_d = \sum_{t \in q} \log\left[\frac{N}{\mathrm{df}_t}\right] \cdot \frac{(k_1+1)\mathrm{tf}_{td}}{k_1\left[(1-b)+b\times(L_d/L_{ave})\right]+\mathrm{tf}_{td}} \tag{7}$$

RSV_d denotes the weight of term t in the document d, L_d and L_{ave} denotes the length of document d and the average length of the entire document. k_1 and b are two free variables, usually $k_1 \in [1.2, 2.0]$, $b = 0.75$.

The keywords of quality requirements are used as query morphemes, and the core concept set of extracted data is used as a set of retrieved documents. The retrieval results of core qualities are arranged according to the order of matching score varying from large to small. This is the order in which learning materials are recommended for the employee.

3.2.3 Employee Competency Evaluation

Using the above methods, we choose the most suitable learning materials for different positions of employees, and then evaluate the learning effect of each employee to

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

get the job competency of employees for that position. Based on the above process, we have developed a program to record the behavior information of employees in the process of material learning. By calculating Pearson correlation coefficient, sensitive data including employee name, personnel code and irrelevant attributes are deleted.

Since it is a classification problem, we use the decision tree model. The final test result is used as the prediction target, and other attributes are used as input. We construct a learning effect evaluation model based on employee learning behavior, and the results of the model are used to evaluate the job competency of employees in this position. The results of the model and the analysis are described in detail in section 4.

3.3 Employee Comprehensive Performance Appraisal

Through the above two modules, we automatically evaluate employees' work performance and job competency respectively, and the final assessment scores are as shown in (8):

$$PAScore = \alpha_1 \cdot Work\_Score + \alpha_2 \cdot Competency\_Score \tag{8}$$

$Work\_Score$ denotes the work performance of employees, and $Competency\_Score$ denotes the job competency. These two parts reflect the employees' current competence and the future growth potential. These two parts are very important indicators for the development of an enterprise. Different companies have different levels of concern for these two indicators. Therefore, enterprises can adjust the weights of the two parts according to their actual situations, and get the comprehensive performance appraisal results that meet their own business needs. For example, Company H, which is one of the largest high-tech companies in China, has intensively employed our model to evaluate their employees. Positive feedbacks are obtained from Company H.

4 Experiment

4.1 Textual Core Concept Extraction Based on Graph Model

In our textual core concept extraction experiment, we employ the famous "principle of salary management" in the field of human compensation. The book contains about 4.65 million Chinese characters. It is the latest textbook of original salary management in China. It is very suitable for the employees' self-learning scene in the assessment of competency. We compare the key sentence proposed by the author with the core concepts extracted by the TextRank graph model algorithm, to verify whether the core concept extraction method based on TF-IDF and TextRank is suitable for this scenario. Then, according to the results, we choose the most appropriate number of core concept sentences. Here, we introduce the precision and NDCG [21] as the evaluation indexes. These two evaluation criteria are shown in (9) and (10):

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

$$P = \frac{x_i \cap y_i}{n} \tag{9}$$

$$NDCG = Z\sum_{p=1}^{n} \frac{2^{r_p}-1}{\log(1+p)} \tag{10}$$

In the formula of precision, x_i denotes the set of extracted sentences, y_i denotes the set of author's intention, *n* denotes the number of extracted sentences. In the formula of NDCG, Z is a regularization term, r_p denotes the score of the sentence p. Accuracy is used to evaluate the degree of matching between the extraction result and the author's intention. The higher the accuracy is, the more representative the author's intention is.

The NDCG value is used to evaluate the difference between the weight ranking of the core concepts and the key sentence ranking of the author's intention. The higher the value is, the more accurate the sentence ranking is. Because of the structure of the article, we can enhance the weight of the key information based on its title information. The results of the experiment in the "*concept of compensation*" is presented in Table 2:

Table 2. Result of core concept extraction experiment

Number of test groups	Number of sentences extracted	Whether or not to optimize based on title	Precision	NDCG
1	10	No	1.0	0.6776
	10	Yes	1.0	0.739
2	20	No	0.9	0.6426
	20	Yes	1.0	0.7445
3	30	No	0.8333	0.6702
	30	Yes	1.0	0.7408

Through our experiments, it is evident that the improvement based on the title has a significant effect on the extraction of the core concept, and the effect is best when the number of sentences is 20.

Therefore, in actual use, we select 20 sentences with the title enhancement, we can automatically get very accurate core concepts. It provides a reliable basis for personalized recommendation based on the characteristics of employee quality.

4.2 Employee Competency Evaluation Based on Decision Tree

In this part of the experiment, we use the learning behavior data from 1735 employees of Company H to build a decision tree model. These data are valid data obtained through the background when employees use the learning program. 1132 pieces of data are used as training sets and 603 are used is test sets. Three decision tree models, C & RT, CHAID and C5.0, are used to construct the model. Here, we define the precision in (11):

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

$$P = \frac{n_t}{n} \tag{11}$$

n_t denotes the number of correctly classified samples, and n denotes the number of total samples.

The outcome shown in Table 3:

Table 3. Outcome of different decision Tree models

Decision tree model types	Number of correctly classified samples	Number of wrongly classified samples	Precision
C&RT	599	4	99.34%
CHAID	599	4	99.34%
C5.0	601	2	99.67%

The classification accuracy obtained by C5.0 model is the highest. The decision tree model using C5.0 is shown in Figure 2:

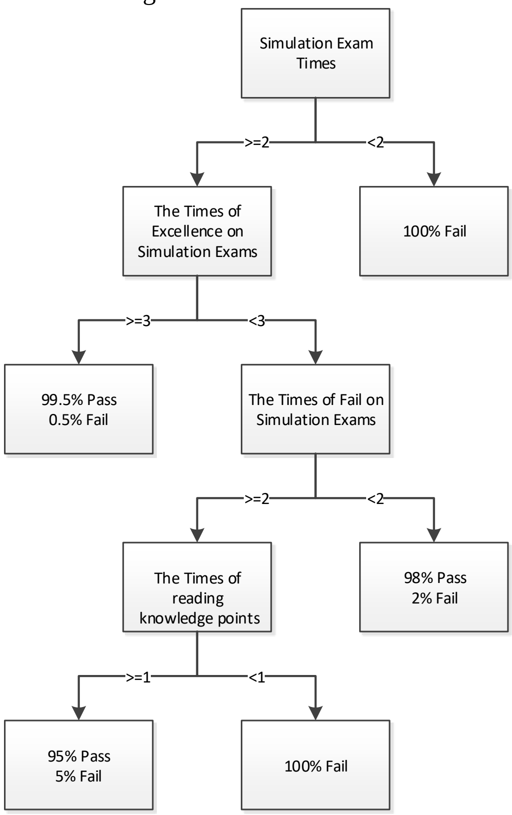

Fig. 2. Decision Tree Model based on C5.0

With the above decision tree model, we get job competency evaluation model based on employee learning behavior. The indexes that can best reflect the learning ability of

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

the employee include: the number of times to participate in the simulated examinations, the excellent times of the test results, the number of times of unqualified examination results and the number of times of reading the key knowledge points. Our model results can also be used to guide the employee learning and skill-set building. For example, one implication is that learning should be accompanied by taking simulation tests and reading core knowledge points.

5 Conclusion

Human resource performance appraisal index system has great application values in enterprise management. It is of great significance to tap the potential of employees, motivate the enthusiasm of employees, and to ensure the overall performance of enterprises. In this paper, we developed a comprehensive employee performance appraisal framework based on data mining and text analysis. Our framework has been successfully applied in Company H and Company J. It effectively improves the fairness in the performance appraisal procedures and fits the latest strategies of the enterprises. It also evaluates the adaptability of employees and obtains a more complete performance appraisal method. Our work can help enterprises to rationally allocate and adjust their positions. Based on the requirements, they can formulate corresponding growth plans for their employees, motivate their work enthusiasm, and enhance their working ability and efficiency.

Acknowledgement

This project was partially supported by Grants from Natural Science Foundation of China #71671178/#91546201/#61202321, and the open project of the Key Lab of Big Data Mining and Knowledge Management. It was also supported by Hainan Provincial Department of Science and Technology under Grant No. ZDKJ2016021, and by Guangdong Provincial Science and Technology Project 2016B010127004.

References

1. Ding, Xiao Rong, X. M. Deng, and J. Luo. "The development of key performance evaluation system of nurse staff." Chinese Nursing Management (2009).
2. Abdulqadir, Ahmad. "A New Model of Information Systems Efficiency based on Key Performance Indicator (KPI)." International Journal of Advanced Computer Science & Applications (2017).
3. Hardison, Chaitra M, et al. "360-Degree Assessments." (2015).
4. Han, Jiawei, and M. Kamber. "Data Mining: Concepts and Techniques." Data Mining Concepts Models Methods & Algorithms Second Edition 5.4(2011):1 - 18.
5. Shmueli, Galit, P. C. Bruce, and N. R. Patel. "Data Mining for Business Analytics: Concepts, Techniques, and Applications in XLMiner, 3rd Edition." (2016).

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37

6. Olson, David L., and D. D. Wu. "Data Mining Models and Enterprise Risk Management." (2017).
7. Li, Liling, S. Shrestha, and G. Hu. "Analysis of road traffic fatal accidents using data mining techniques." IEEE, International Conference on Software Engineering Research, Management and Applications IEEE, 2017:363-370.
8. Chaurasia, Vikas. "Early Prediction of Heart Diseases Using Data Mining Techniques." Social Science Electronic Publishing vol.1.2013(2013):208-217.
9. Rygielski, Chris, J. C. Wang, and D. C. Yen. "Data mining techniques for customer relationship management." Technology in Society 24.4(2002):483-502.
10. Slater, Stefan|Joksimovic, Srecko|Kovanovic, Vitomir|Baker, Ryan S.|Gasevic, Dragan. "Tools for Educational Data Mining: A Review. " Journal of Educational & Behavioral Statistics 42.1(2017):85-106.
11. Chien, Chen Fu, and L. F. Chen. "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry." Expert Systems with Applications 34.1(2008):280-290.
12. Li, Baoli, and L. Han. "Distance Weighted Cosine Similarity Measure for Text Classification." International Conference on Intelligent Data Engineering and Automated Learning --- Ideal Springer-Verlag New York, Inc. 2013:611-618.
13. Shen, Yung-Chi, et al. "A cross-database comparison to discover potential product opportunities using text mining and cosine similarity." (2017).
14. Shouzhong, Huang, and Minlie. "Mining microblog user interests based on TextRank with TF-IDF factor." Journal of China Universities of Posts & Telecommunications 23.5(2016):40-46.
15. Audich, Dhiren A., R. Dara, and B. Nonnecke. "Extracting keyword and keyphrase from online privacy policies." Eleventh International Conference on Digital Information Management IEEE, 2017:127-132.
16. Karlsson, Simon. "Using semantic folding with TextRank for automatic summarization." (2017).
17. Ibrahim, M., and M. Murshed. "From Tf-Idf to learning-to-rank: An overview." (2016).
18. Barrios, Federico, et al. "Variations of the Similarity Function of TextRank for Automated Summarization." (2016).
19. Page, L. "The PageRank citation ranking : Bringing order to the web." Stanford Digital Libraries Working Paper 9.1(1998):1-14.
20. Niu, Jianwei, et al. "OnSeS: A Novel Online Short Text Summarization Based on BM25 and Neural Network." Global Communications Conference IEEE, 2017:1-6.
21. Wang, Yining, et al. "A theoretical analysis of NDCG ranking measures." Proceedings of the 26th Annual Conference on Learning Theory (COLT 2013). 2013.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93701-4_37