Data mining Approaches in Manpower Evaluation

Jun Meng, Xiao Chen, Tianyu Zhu, Yangyang Pan
College of Electrical Engineering
Zhejiang University
Hangzhou, China
junmeng@zju.edu.cn

***Abstract*—Manpower allocation determines the competence of the companies. Scientific manpower allocation calls for accurate evaluation on the abilities that the employees have for the posts. In this paper, we first present a general fuzzy clustering model for manpower evaluation for companies. To verify the approach, a new distance-based evaluation model is also presented. Simulation results demonstrated the accuracy of our research.**

***Keywords*-manpower evaluation, fuzzy clustering, distance-based, ability, post**

I. Introduction

Manpower allocation is one of the managerial problems that most companies have to deal with. [1].Depending on the company and the purposes of the allocation, the problem definition and the constraints would vary. Whether a company can identify its employees' special abilities and talents, and whether the employees can be installed into the posts that they fit best, largely determines the future development of the company. In order to build proper and effective structure, the personnel placement should be based on the comprehensive analysis of all the abilities of the employee. Manpower allocation is a complicated problem for many different factors [2].The demand of discovering employees' abilities calls for the help of a scientific evaluation system. The research of this problem in the traditional management theory is qualitative.

In relate woks, many kinds of model are developed for manpower allocation. The author a Basic distance/time models to allocate the manpower for job specification in [3].An utility function is developed for manpower allocation to new drug screening programs in [4]. Hungarian Algorithm is used for manpower allocation [5]; it is confined by the condition that the efficiency of the assignment problem is definite. A fuzzy model based on the Hungarian Algorithm is presented in the allocation of manpower in Logistic Corporations [6]. Also, a met heuristic model is also used in the allocation of salesmen force in a company [7].All these researches only focus the force evaluation of one calling.

Appropriate allocation of manpower calls for accurate evaluation of employee abilities such as horizon, specialization, creativity and age. Age distribution of manpower age model were proposed in [8] based on the condition that employees with the same age are on the same skill level; however, it provided little information in manpower evaluation with other abilities. A real size manpower allocation problem was modeled after a real world problem of distributing the salesmen force over the branches of a company [9]. The problem includes multiple objectives and the number of salesmen at each branch is unspecified; however, it focused on salesmen only and did not provided general evaluation standards.

In this paper, we present a general and applicable manpower-allocation model involving many abilities of employees for all kinds of companies. We first present a fuzzy clustering algorithm for the manpower allocation, then, to verify our approach, a new distance-based manpower evaluation method is developed. The results of the former two different models demonstrated the accuracy of our approach.

This paper is organized as follows: Section 2 provides a fuzzy clustering model and a distance-based evaluation model is presented in Section 3; simulation results are given in Section 4 and Section 5 provides concluding remarks.

II. The Algorithm of the Fuzzy Clustering

A. *Construction of sample set*

In our research, an employee is denotes a sample and the abilities of an employee is denotes as the attributes of the sample. We assume that the sample is denoted as $c_1,c_2,c_3......c_n$, which has m attributes denoted as $c_i=\{c_{i1},c_{i2},c_{i3}......c_{im}\}$; Let $w_1, w_2, w_3......w_m$ be the weight of the m attribute which can be determined by the company according to the requirements of the post. （$w_1+w_2+w_3+......w_m=1$）.We assume $U=\{u_1,u_2,u_3......u_n\}$ be the sample set, where $u_i=\{w_1c_{i1},w_2c_{i2},w_3c_{i3}......w_4c_{im}\}$ .For the reasons that $U_{ij}\,(1\le i\le n,1\le j\le m)$ may not belong to [0, 1] in actual evaluation system，we first deal with the data for standardization as follows:

$$U_k{}'=U_{1k}+U_{2k}+U_{3k}+.....+U_{nk}=\frac{1}{n}\sum_{i=1}^{n}U_{ik} \qquad (1\le k\le m) \tag{1}$$

The standard deviation of the original data

$$s_k=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(U_{ik}-U_k{}'\right)^2} \qquad (1\le k\le m) \tag{2}$$

Then we used the extremum standardization formula for the calculation as follows:

$$U_{ik}' = \left| \frac{U_{ik} - U_k'}{S_k} \right| \quad (3) \qquad T_{ik} = \frac{U_{ik}' - U_{\min k}'}{U_{\max k}' - U_{ik}'} \quad (4),$$

where $U_{\min k}' = \min\{U_{1k}', U_{2k}', U_{3k}', U_{4k}' \ldots U_{nk}'\}$, and $U_{\max k}' = \max\{U_{1k}', U_{2k}', U_{3k}', U_{4k}' \ldots U_{nk}'\}$

Following these steps, we have $T = (T_{ik})_{n \times m}$ which is the final sample set of our research.

B. Correlation matrix R

R is described by the similarity matrix as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1n} \\ r_{21} & r_{22} & \Lambda & r_{2n} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{n1} & r_{n2} & \Lambda & r_{nn} \end{bmatrix} 0 \le r_{ij} \le 1, (0 \le i, j \le n) \quad (5)$$

$$r_{ij} = \sum_{k=1}^{m} \min(T_{ik}, T_{jk}) / \sum_{k=1}^{m} \max(T_{ik}, T_{jk}) (0 \le i, j \le n) \quad (6)$$

C. clustering analysis

For the clustering, we use Maximal Tree Method. We construct a graph whose nodes are the samples and the nodes i and j can be connected with each other if $r_{ij} \ne 0$. We construct the graph as follows: first we describe a node i which is more concentrated, then under the condition that no loop is brought, we connect the nodes according to sequence of the values of r_{ij}, we can have a maximal tree. For different methods of connection, different kinds of tree can be brought. However, the results of all the trees are the same.

Then, we get the cut set of the maximal tree with the parameter of λ, the sides are deleted if $r_{ij} < \lambda$, then we have a few sub trees . In the model λ can be determined by the company according the requirements of the post. For different kinds of maximal develop in the anterior Section, the sub tree are the same which can bring the same results of the model. The sub trees can be used as modes discovered form the data warehouse.

D. Evaluation of the manpower

For the modes developed in Section2.2, we have

$$Mode_{ij} = \sum T_{kj} / p, (i = 1,2,\ldots s; j = 1,2,\ldots m) \quad (7)$$

Where s is the number of modes, k is the k th data record among p data records that brought the i th mode.

Then we have the correlation factors by comparing X (where $X_1, X_2, \ldots X_n$ denote the n sub set of X) with the modes:

$$(X, Mode_i) = 1/2 [X \cdot Mode_i + (1 - X \odot Mode_i)] \quad (8)$$

According to the proximity principle, we have

$$(X, Mode_i) = \max\{(X, Mode_1), \ldots (X, Mode_s)\} \quad (9)$$

In our approach, if the value of a correlation factor to one mode is the biggest of the correlation factors with all the modes, it belongs to the modes. Then we can evaluate the sample by judge which mode it is close to from its correlation factors with the modes. Different modes means different abilities scales for the post.

III. DISTANCE-BASED EVALUATION

A. Construction of evaluation matrix

We use the sample matrix $X = (X_{ij})_{n \times m} = U$ in Section 2.1. Then we construct the evaluation standard matrix $Y = (Y_{ij})_{m \times q}$ where q is the scales of the evaluations. In the model, if an employee is in the first scale, it means that he is the most competent employee for the post. Y_{ij} denotes that the expected values of the j th scales of the i th attribute. The matrix can be determined by the company according to the requirements of the post.

B. Dimensionless matrixes

In actual evaluation system, the data are usually out of standardization, we first convert the data to dimensionless data. Then we have the following matrixes:

$$A = (a_{ij})_{(i=1,2,3,4,5 \ldots m; j=1,2,3,n)} \quad B = (b_{kt})_{(k=1,2,3,\ldots p; t=1,2,3,4,\ldots q)}$$

Where $a_{ij} = \begin{cases} X_{ij} / \max_j X_{ij} & j \ne m-1 \\ \min_j X_{ij} / X_{ij} & j = m-1 \end{cases}$ and $b_{ij} = \begin{cases} y_{kt} / \max_k y_{kt} & k \ne m-1 \\ \min_k y_{kt} / y & k = m-1 \end{cases}$

C. Evaluation based on the distance

1) Euclidean distance

Distance from vectors in A to vectors in B:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (a_{ik} - b_{kj})^2}, (j = 1,2,3,4,5 \ldots q) \quad (10)$$

If $d_{ik} = \min_{1 \le j \le q} \{d_{ij}\}$, then the i th employee belongs to the k th scale $(i = 1,2,3,4,5 \ldots q)$

2) Absolute distance

Distance from vectors in A to vectors in B:

$$D_{ij} = \sum_{K=1}^{p} |a_{ij} - b_{ij}| \quad (11)$$

if $D_{ik} = \min_{1 \le j \le q} \{D_{ij}\}$, then the i th employee belongs to k th scale. $(i = 1,2,3,4,5 \ldots q)$

IV. SIMULATION RESULTS

In our approach, we establish twenty samples with five attributes (horizon, specialization, creativity, responsibility, and confidence) for the posts of research.

In the simulation of the model presented in Section 3.3, we have six modes with $\lambda = 0.63$ and $w_1 = w_2 = w_3 = w_4 = w_5 = 0.2$ and all the correlation factors of the samples to the six modes. From the correlation factors, we divide the employees into six sets with different ability scales where the first scale means the most competent for the posts as follows:

Scale 1: $\{u_7\}$

Scale 2: $\{u_2, u_6, u_{13}, u_{18}\}$

Scale 3: $\{u_1, u_4, u_9\}$

Scale 4: $\{u_3, u_8, u_{11}, u_{12}, u_{16}, u_{20}\}$

Scale 5: $\{u_5, u_{10}, u_{17}\}$

Scale 6: $\{u_{14}, u_{15}, u_{19}\}$

In the simulation of model presented in Section 3.3.1, we construct an evaluation standard matrix $Y=(Y_{ij})_{m\times q}$ with $m=5, q=6$, and then we have the following results:

Scale 1: $\{u_7\}$

Scale 2: $\{u_2, u_6, u_{13}\}$

Scale 3: $\{u_1, u_4, u_9, u_{18}\}$

Scale 4: $\{u_3, u_8, u_{11}, u_{12}, u_{16}, u_{20}\}$

Scale 5: $\{u_5, u_{10}, u_{17}\}$

Scale 6: $\{u_{14}, u_{15}, u_{19}\}$

Also, we have the following results in the simulation of the model presented in Section 3.3.2

Scale 1: $\{u_7\}$

Scale 2: $\{u_2, u_6, u_{13}, u_{18}\}$

Scale 3: $\{u_1, u_4, u_9\}$

Scale 4: $\{u_3, u_8, u_{11}, u_{12}, u_{16}, u_{20}\}$

Scale 5: $\{u_5, u_{10}\}$

Scale 6: $\{u_{14}, u_{15}, u_{19}, u_{17}\}$

The results of the simulation with small differences demonstrate our approach accuracy. In actual evaluation system, the company can determine the standard for the requirements of the posts. And with the results of the evaluation, the company can allocate the employees to appropriate posts. Also, the company can determine different attributes for different posts.

V. CONCLUSIONS

In this paper, the accuracy of the general fuzzy clustering model and the distance-based evaluation model are demonstrated basically by case study. And this method may improve the operation efficiency of the companies by obtaining the appropriate matching between posts and employees. In the future, we will have further study on the issue that the optimal manpower allocation of a company with different posts and given employees with the method proposed in this paper , and it would work better in enhancing the efficiency of organization system of a company.

ACKNOWLEDGMENT

This work is supported in part by National Natural Science Foundation of China(60574079); Zhejiang province science and technology plan by the funded projects (2011 c23097); Zhejiang fund projects (LY12F03023); Zhejiang provincial projects (Y20107866)

REFERENCES

[1] HESHAM K. ALFARES and JAMES E. BAILEY, “Integrated project task and manpower scheduling”, IIE Transactions Design&Manufacturing, Vol.29, No.9 ,1997.

[2] Michael E. Borus; Charles G. Buntz, “Problems and Issues in the Evaluation of Manpower Programs”, Industrial and Labor Relations Review, Vol. 25, No. 2, pp. 234-245, 1972.

[3] SYDNEY C. K. CHU' and CARRIE K. Y. LIN2, “A Manpower Allocation Model of Job Specialization”, Journal of the Operational Research Society, Vol. 44, No. 10,pp. 983-989, 1993

[4] E. B. Pyle Ⅲ, B. Douglas, G. W. Ebright, W. J. Westlake and A. D. Bender, “Scientific manpower allocation to new drug screening programs”, Management Science, Vol. 19, No. 12, pp. 1433-1443, 1973.

[5] Kohlas,S, “Stochastic Methods of operations Research Cambridge”, Cambridge University Press 1982

[6] Jia Jiang, Zhan-jing Wang, Le-le Qin, “On the manpower dispatch in logistics enterprises by generalized fuzzy assignment”, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006

[7] Nicolas Abboud, Masatoshi Sakawa, “Fuzzy Manpower Allocation Using the Mutate and Spread Metaheuristic”, Second International Conference on Knowledge-Based Intelligent Electronic Systems, 21-23 April 1998.

[8] Chirumamilla Shivanagaraju, B. Mahanty, K. Vizayakumar, Pratap K.J. Mohapatra, “Beta-distributed age in manpower planning models”, Applied Mathematical Modeling ,pp.23-37, 1998.

[9] Abboud N, Inuiguchi M, Sakawa M, Uemura Y, “Manpower allocation using genetic annealing ", European Journal of Operational Research, Volume 111, Number 2, pp. 405-420,1998