# A Realistic Personnel Selection Tool Based on Fuzzy Data Mining Method

**Wei-Shen Tai      Chung-Chian Hsu**

Department of Information Management, National Yunlin University of Science and Technology

## Abstract

Recruitment is an important issue of human resource management (HRM). The main purpose of recruitment is to obtain talent from internal and external of organization. However, excellent talent maybe will be missed at the pre-screening process under lack of complete information for managers to judge. In this paper, a personnel selection tool based on fuzzy data mining method is proposed to assist business managers to find eligible talent more efficiently.

**Keywords**: Human Resource Management (HRM), Recruitment, Personnel selection, Fuzzy Data mining.

## 1. Introduction

In the Knowledge Era, more and more business managers have considered transforming into a learning organization will be a crucial issue for retaining competitive advantage. Besides building a robust organizational system, acquiring excellent knowledge workers should be the other essential in a learning organization [7, 8]. Pre-screening is a familiar procedure before the formal personal selection as usual. It is also regarded as an efficient method for saving the cost and time of personnel selection. However, businesses must spend a great number of labors and time to review applicants' personal data to determine whether a job applicant is eligible or not [7]. Although business mangers expect to filter out ineligible applicants through pre-screening, it may also cause excellent talent will be ignored or missed unfortunately because lack of complete information for managers to select the suitable staff. Due to this mistake of pre-screening, organization will lose numerous opportunities to acquire excellent and potential talent in the processes of recruitment.

In this paper, a personnel selection tool based on fuzzy data mining method is proposed to be a feasible solution for the foregoing problems; it is feasible to assist businesses in finding eligible applicants through reliable information effectively and efficiently. This paper is constructed as follows. Section 2 introduces the concept and application of personnel selection methods. In Section 3, fuzzy partition and fuzzy association rule are elaborated. In Section 4, the proposed system is demonstrated by means of a real case. Finally, some conclusions are stated at the end of this paper.

## 2. Personnel selection methods

Personnel selection often acts the most important role for controlling the human source and quality in HRM [12, 13]. It is necessary for a well-functioning organization to possess three types of behavior (i) people must be induced to enter and remain with the organization; (ii) they must reliably carry out specific role or job requirements; and (iii) there also needs to be innovative and spontaneous activity that goes beyond role prescription [9, pp.4]. Therefore, the goal of personnel selection is applying valid and effective method to reduce the risks of hiring an unsuitable employee, and increase the opportunities to find an eligible employee who can enhance the productivity of organization [3, 9]. In other words, businesses find eligible employees meet the requirements of organization and occupation from mass job applicants through effective personnel selection methods.

The traditional personnel selection methods such as personality, interview, assessment center, and bio-data have applied in businesses for a long time [7, pp.454]. Numerous studies have demonstrated that the relationship is existed between the attributes of job applicants' owned and their future organizational behavior (e.g. job performance, absenteeism and turnover). Those predictors which as the valid determinant of predicting employees' behavior have be proposed and validated in accordance with the results of personnel selection researches [7]. It also provides a feasible solution for businesses to pre-screen job applicants by means of predictors of organizational behavior before the processes of formal personnel selection. In this paper, the fuzzy mining data method is applied to discover those predictors or attributes which are valid to predict the organizational

behavior of applicants in the future. It is helpful for business to pre-screen applicants in the processes of personnel selection.

# 3. Proposed method

There are numerous attributes which are job applicants owned can be as the predictors of organizational behavior. However those attributes maybe possess different data type and unit to represent the diversity of degree, respectively. Additionally, attributes with traditional crisp value will increase the computational complication during the processes of data mining. Therefore, the linguistic variables and fuzzy membership function are applied to integrate heterogeneous data and simplify the computational complication during the processes of data mining. In this paper, the fuzzy data mining method is applied to obtain valid fuzzy association rules from existing transaction database. About the fuzzy data mining technology based on Apriori algorithm [14] is elaborated as follows.

## 3.1. Fuzzy partition

The concept of fuzzy set and linguistic variable was proposed by Zadeh [10, 11]. It is feasible to represent the diversity of degree of the feature with different linguistic value in many applications [14]. A fuzzy set $\tilde{A}$ in a universe of discourse $X$ is characterized by a membership function $\mu_{\tilde{A}}(x)$, which associates with each element $x$ in $X$ a real number in the interval [0,1] [10]. A fuzzy number is a fuzzy subset in the universe of discourse $X$ that is both convex and normal [6].

**Definition 1.** A positive triangular fuzzy number (PTFN) $\tilde{T}$ can be defined as $\tilde{T} = (l, m, u)$, where $l \leq m \leq u$ and $l > 0$. The membership function, $\mu_{\tilde{T}}(x)$, is defined as [6]

$$\mu_{\tilde{T}}(x) = \begin{cases} \dfrac{x-l}{m-l} & ,l < x < m \\[2mm] \dfrac{u-x}{u-m} & ,m < x < u \\[2mm] 0 & ,otherwise \end{cases} \tag{1}$$

**Definition 2.** A linguistic variable is a variable whose values are expressed in linguistic terms. In other words, variable whose values are not numbers but words or sentences in a nature or artificial language [2]. For example, "Age" is a linguistic variable whose values are young, medium, old, etc. These linguistic values

can also be represented by fuzzy numbers. It is suitable to represent the degree of subjective judgment in qualitative aspect than crisp value.

In this paper, each attribute is partitioned by fuzzy partition methods with triangle membership function. For example, the attribute "Age of taking job" is divided into 3 linguistic values (young, medium, old) and its membership function is predefined as follows(shown as Fig. 1). These linguistic values are regarded as initial candidate fuzzy grids to generate the other large fuzzy grids and fuzzy association rule in following tasks.

$$\mu_{low}^{AgeTJ}(x) = \begin{cases} 1 & ,x < l \\[2mm] \dfrac{m-x}{m-l} & ,l \leq x \leq m \\[2mm] 0 & ,x > m \end{cases} \tag{2}$$

$$\mu_{med}^{AgeTJ}(x) = \begin{cases} \dfrac{x-l}{m-l} & ,l < x \leq m \\[2mm] \dfrac{u-x}{u-m} & ,m < x < u \\[2mm] 0 & ,otherwise \end{cases} \tag{3}$$

$$\mu_{high}^{AgeTJ}(x) = \begin{cases} 0 & ,x < m \\[2mm] \dfrac{u-x}{u-m} & ,m \leq x \leq u \\[2mm] 1 & ,x > u \end{cases} \tag{4}$$
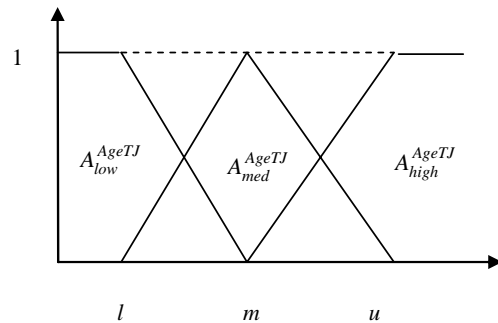


Fig. 1 Linguistic values of "Age of taking job".

## 3.2. Fuzzy association rule

In order to discover valid fuzzy association rules for pre-screening job applicants in the processes of personnel selection, the large fuzzy grid concept is applied in this paper. If a database possesses $k$ attributes and $n$ tuples of employee data, the fuzzy association rules are generated as follows.

Step 1. Perform the fuzzy partition to generate initial candidate fuzzy grids.

Step 2. Scan the database and fuzzy support (FS) value is calculated as follows [4, 5].

$$\mathrm{FS}(A_{s_i}^{x_i}) = \frac{1}{n} \sum_{j=1}^{n} \mu_{s_i}^{x_i}(d_j^{x_i}) \qquad (5)$$

where, $x_i$ is the $i$th attribute, $s_i$ is the term of degree in $x_i$, $A_{s_i}^{x_i}$ is the linguistic value of $s_i$ in the $i$th attribute, $\mu_{s_i}^{x_i}$ is the membership function of $s_i$ in the $i$th attribute, $d_j^{x_i}$ is the original value of the $i$th attribute in the $j$th tuple.

Step 3. If the FS value of linguistic value is larger than or equal to the user-specified min FS, it will be regarded as an eligible fuzzy grid.

Step 4. Those eligible fuzzy grids are combined to calculate their FS value as follows.

$$\mathrm{FS}(A_{s_1}^{x_1}, A_{s_2}^{x_2}, ..., A_{s_k}^{x_k})$$
$$= \frac{1}{n} \sum_{j=1}^{n} [\mu_{s_1}^{x_1}(d_j^{x_1}) \bullet \mu_{s_2}^{x_2}(d_j^{x_2}) \bullet ... \bullet \mu_{s_k}^{x_k}(d_j^{x_k})] \qquad (6)$$

Step 5. Go to step 3 to check the FS value of each fuzzy association rule. It will stop when only one rule is larger than or equal to the user-specified min FS. Otherwise, the processes between steps 3 to 5 will repeat again.

Step 6. Filter out ineligible fuzzy association rules with fuzzy confidence. The general fuzzy classification rule denoted by $R$ is stated as

$$R: A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times ... \times A_{s_k}^{x_k} \Rightarrow A_{s_\alpha}^{x_\alpha} \quad \text{with} \, \mathrm{CF}(R) \qquad (7)$$

where, $x_\alpha (1 \le \alpha \le k)$ is the class label and $\mathrm{CF}(R)$ is the certainty grade of $R$. It means if $x_1$ is $A_{s_1}^{x_1}$ and $x_2$ is $A_{s_2}^{x_2}$ and …and $x_k$ is $A_{s_k}^{x_k}$, then $x_\alpha$ is $A_{s_\alpha}^{x_\alpha}$ with certainty grade $\mathrm{CF}(R)$. The fuzzy confidence of $R$ is defined as follow [4, 5]:

$$\mathrm{FC}(R) = \frac{\mathrm{FS}(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times ... \times A_{s_k}^{x_k} \times A_{s_\alpha}^{x_\alpha})}{\mathrm{FS}(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times ... \times A_{s_k}^{x_k})} \qquad (8)$$

When the $\mathrm{FC}(R)$ of rule $R$ is not larger than or equal to the user-specified min FC, the rule $R$ will be regarded as ineffective fuzzy association rule and filtered out.

Step 7. Prune the redundant fuzzy association rules. If there exist two rules, denoted by $R$ and $S$, possessing same consequence, and the antecedence of $R$ is contained in $S$, then $R$ is redundant and can be discarded.

# 4. Experiment and analysis

In order to demonstrate the proposed method, a realistic fuzzy data mining system is built and an experiment is implemented in a real case. The experimental procedure (shown in Fig. 2) and analysis are described as follows.

## 4.1. Experiment

Step 1. *Data aggregation*. In this experiment, the real employee data is provided by a car agent.

Step 2. *Data transformation*. The essential attributes will be transformed into appropriate format to meet the analytic requirements of this proposed system.

Step 3. *Attributes' level maintenance*. According to the number of fuzzy partitioned of each attribute, the information of each attribute's level is saved into database (shown in Table 1).

Step 4. *Executing fuzzy data mining*. We set "Duration of staying job" to represent the degree of "salesman stability". Additionally, the Max Scan Time is set five in this experiment.

## 4.2. Result analysis

The 10-fold cross validation is implemented to acquire the correct percent (CP) with different combinations of the min FS and min FC setting in this experiment. For each 10-fold cross validation, the data set was first partitioned into 10 equal sized sets and each set was then in turned used as the test set while the system acquires fuzzy association rules from the other nine sets. The membership degree of the consequence is obtained by the Generalized Modus Ponens (GMP) inference method and compared with the raw data [1].

According to the comparison result, we can find the most of the highest CP 61.25 appear when the min FS set 0.2 and the error average is 0.187 when the min FC set 0.3 (shown in Fig. 3 )
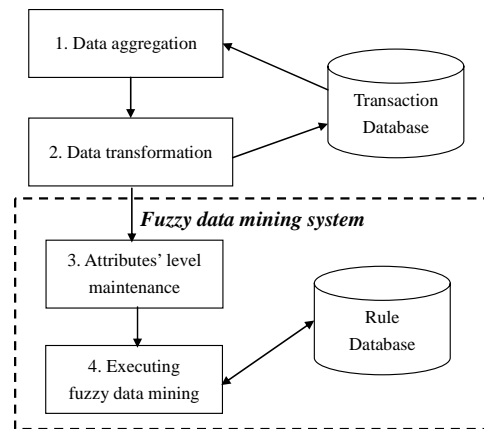


Fig. 2 The procedure of fuzzy data mining in this experiment

Table 1 Attributes and their information

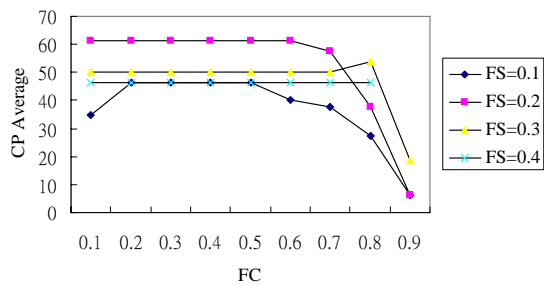| Attribute | Linguistic variable | Fuzzy number |
|---|---|---|
| Duration of staying job (unit: month) | Short | (0, 3, 7) |
| | Medium | (3, 7, 10) |
| | Long | (7, 10 ,10) |
| Age of taking job | Young | (0, 25, 35) |
| | Medium | (25, 35, 45) |
| | Old | (35, 45, 45) |
| Duration of past job (unit: month) | Short | (0, 9 , 13) |
| | Medium | (9, 13, 42) |
| | Long | (13, 42, 42) |
| Duration of past sales job (unit: month) | Short | (0, 0, 0) |
| | Medium | (0, 0, 7) |
| | Long | (0, 7, 7) |
| Education degree | Low | (0, 0.9, 1.6) |
| | Medium | (0.9, 1.6, 3.3) |
| | High | (1.6, 3.3, 3.3) |



Fig. 3 The CP average of each min FS and min FC setting

# 5. Conclusion

Besides the fairness and objectivism are considered, the prediction of the future organizational behavior of employees should be the other focal point in the processes of personnel selection. Therefore numerous predictors are proposed and validated to attempt finding the relationship between the attributes applicants owned and organizational behaviors.

In this paper, a personnel selection tool is proposed to discover those valid relationships between attributes and organizational behaviors. It can assist business manager to find eligible talent more efficiently.

# 6. References

[1] C. Carlsson and R. Fuller, *Fuzzy reasoning in decision making and optimization*, New York: Physica-Verl., 2002.

[2] E. Herrera-Viedma and E. Peis, "Evaluating the informative quality of documents in SGML format from judgments by means of fuzzy linguistic techniques based on computing with words," *Information Processing and Management*, Vol. 39, Issue 2, pp. 233-249, 2003.

[3] F. Herrera, E. López, C. Mendaña and M. A., Rodríguez, "A linguistic decision model for personnel management solved with a linguistic biobjective genetic algorithm," *Fuzzy Sets and Systems*, Vol. 118, Issue 1, pp. 47-64 , 2001.

[4] H. Ishibuchi, T. Nakashima and T. Yamamoto. "Fuzzy association rules for handling continuous attributes," *Proceedings of IEEE International Symposium on Industrial Electronics*, Pusan, Korea, pp. 118-121, 2001.

[5] H. Ishibuchi, T. Yamamoto and T. Nakashima, "Fuzzy data mining: effect of fuzzy discretization," *Proceedings of the 1st IEEE International Conference on Data Mining*, San Jose, USA, pp. 241-248, 2001.

[6] H.J. Zimmermann, *Fuzzy Sets, Decision Making, and Expert Systems*, Boston: Kluwer Academic Publisher, 1991.

[7] I.T. Robertson and M. Smith., "Personnel selection," *Journal of Occupational and Organizational Psychology*, Vol. 74, pp. 441-472, 2001.

[8] J.C. Hayton, "Promoting corporate entrepreneurship through human resource management practices: A review of empirical research," *Human Resource Management Review*, Vol. 15, Issue 1, pp. 21-41, 2005.

[9] J.M. Werner, "Implications of OCB and Contextual Performance for Human Resource Management," *Human Resource Management Review*, Vol. 10, Issue 1, pp. 3-24, 2000.

[10] L.A. Zadeh, "Fuzzy sets," *Inform. and Control*, Vol. 8, pp. 338-353, 1965.

[11] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 8, Issue 3, pp. 199-249, Issue 4, pp. 301-357, 1975; Vol. 9, Issue 1, pp. 43-80, 1976.

[12] M. Nussbaum, M. Singer, R. Rosas, M. Castillo, E. Flies, R. Lara and R. Sommers, "Decision support system for conflict diagnosis in personnel selection," *Information & Management*, Vol. 36, Issue 1, pp. 55-62, 1999.

[13] R. Storey Hooper, T.P. Galvin, R.A. Kilmer and J. Liebowitz, "Use of an expert system in a personnel selection process," *Expert Systems with Applications*, Vol. 14, Issue 4, pp. 425-432, 1998.

[14] Y.C. Hu, J.S. Hu, R.S. Chen and G.H. Tzeng, "Assessing weights of product attributes from fuzzy knowledge in a dynamic environment," *European Journal of Operational Research*, Vol. 154, Issue 1, pp. 125-143, 2004.