

Research on the College Graduate Employment Education Based on Data Mining Technology

Jianjun Wang¹ and Li Li²

¹*North China Electric Power University, School of Economic and Management
Administration, Beijing, China*

²*Beijing Information Science and Technology University, School of Economics and Business
Administration, Beijing, China
E-mail: <wangjianjunhd@126.com>*

KEYWORDS Career Plan. Decision Tree. Employment factors. Rough Set. Vocational Education Guidance

ABSTRACT With the weak global economic process, the unemployment rate of the college graduate has become a hot issue of all the countries. How to enhance the employment competitiveness of the college graduate is a main problem of universities. It is an effective way to find the important factors and the relationship of graduate college career match. This paper applies a data mining technology with decision tree technology to make the students know the key factors for their employment, and rough set is employed in data preprocessing to reduce unnecessary information from the given database. From a case study, the results show that personality and professional skill have a significant impact on college graduate career decision making. It also points out that male college graduates find it easier to get a job than females with a good professional skill.

INTRODUCTION

Nowadays, the employment of college graduates has become one of a main society problem of all of the countries, especially of China. In the summer of 2013, Chan (2015) points out that there are more than 2,000,000 Chinese university graduates waiting to be employed, while the demand for talents in enterprises is growing relatively slow. The contradiction between the number of college graduates and employer demand is more and more obvious, reveals Wei et al. (2014). Moreover, Chinese government work report pointed out there are a large number of laid off workers needing to be reemployed, which make a higher competition of the college graduates, it has been form a huge society problem. How to enhance the employment competitiveness of the college graduate is a main problem of universities, a good way is finding the factors and using employment education to give a advising for making a suitable career decision of college graduates, which presents by Jun (2015).

Employment education of college graduate can not only help graduates to make satisfactory career decisions, but also lay a solid foundation for better realization of graduates' value. Lee et al. (2015) studies the link of the teaching and the employment problem and point out that employment education can also provide the theoretical and practical support to the career education in different level of schools, Five-Factor, SWOT

method and so on are widely used in education research such as the studies by Martincin et al. (2015) and Thomas et al. (2014). These methods analyze the characteristics of graduates and match the characteristics with the occupations, but these methods ignore the factors except graduate characteristics. Recently, several researchers have applied the latest intelligent algorithms and data mining methods to study the graduate career decision. Zhang et al. (2010) apply SVM method to predict the employment of graduates. The studies by Wu (2010) reveals factor-cluster analysis to predict the satisfaction of graduate employment. These empirical employ data mining methods in graduate career decision research. However, these studies can only predict the situation of graduate career decision, but not give the graduates some related recommendations according to their characteristics. This paper applies decision tree method, a kind of data mining technology, to analysis the factors affecting the graduate employment decision-making. The decision tree is built based on those factors and the values of each factor. This research can guide the graduates to make appropriate career decisions better.

METHODOLOGY

Classification, applied to form the decision tree, is a new approach of data mining to find the hidden relations of the data. The decision tree

can be given by using data mining to dig the information received from searching. The classical calculation method in data mining is the ID3 method which is based on dividing the minimum attributes in the information theory. The method is mainly as the described by Eesa et al. (2015): S is supposed to be the training set, there are m independent values in the classification attributes. In other words, the attributes are with value C_i ($i=1\dots m$), R_i is the subclass which attributes to C_i in the S , r_i is the total number of the R_i . Then the expectation of the S can be calculated by the equation (1).

$$I(r_1, r_2, \dots, r_m) = -\sum P_i \log_2 P_i \quad P_i = r_i / |S| \quad (1)$$

There are v different values in A , and then S can be divided into v subclasses according to different values. S_j is a subclass in S of which the value is equal to a_j ($j=1\dots v$). If S_{ij} is the number in the C_i , then the expectation of A to the C_i ($i=1\dots m$) is in equation (2):

$$E(A) = -\sum W_j I(s_{1j}, s_{2j}, \dots, s_{mj}) \quad (2)$$

In which: $W_j = (s_{1j} + s_{2j} + \dots + s_{mj}) / |S|$, $I(s_{1j}, s_{2j}, \dots, s_{mj}) = -\sum P_{ij} \log_2 P_{ij}$, $P_{ij} = s_{ij} / |S_j|$.

The information increase of A regarded as a decision attribute is in equation (3):

$$Gain(A) = (r_1, r_2, \dots, r_m) - E(A) \quad (3)$$

Every information increase of each decision attribute needs to be calculated, and the maximum one will be chosen as the decision attribute node for data set S , and then the branch of the node is constructed with every values of the node. Continue to calculate the information increase of sub class attribute of the branch introduced from this node except the node which has been calculated already. Finally, the decision tree can be constructed. The quality of the decision tree can affect not only the efficiency but also the accuracy of the classification. Many scholars try their best to find the better heuristic functions and evaluation functions. Some scholar prove that to find the ideal decision tree is a NP difficult problem.

The ID3 algorithm needs calculate the information increasing of each decision attribute in the information set, so when the number of the attributes is large, the computation of the algorithm is larger, which is a challenge for producing the ideal decision tree. If the attributes can be reduced before calculating, the computation of the algorithm would be reduced. Rough set defines the knowledge from a new perspective, in which the knowledge is considered as the divided fields of the research of the studies by

Zheng et al. (2015). The knowledge is considered as some granularity, and the theory introduces equivalence relationship of algebra to study the knowledge. This theory is mainly about analysing the reduction of the knowledge and the relationship of the knowledge, which can be applied to computer learning and complex data analysing. Thus, the attributes which is in the risk information sets can be reduced before the data mining calculation base on the theory, then some attributes that is inessential for the decision can be eliminated, and then the computation of the ID3 algorithm can be reduced by this process.

The attributes in the database is equal to the equivalence relationship in the rough set, and this paper will not distinguish the relationship between attribute and equivalence. Supposing U is a finite set which is composed by the interesting elements, and R is one equivalence relationship in U . U/R is the set from which R divides U . $[x]_R$ is the elements whose values are belong to R (x belongs to U). (U, R) is the approximate space in rough set. Any subset which belongs to U is called a concept. The lower approximation and the upper approximation of X can be defined as equation (4). The lower approximation is the set of the elements which are ensured to belong to X of the R in U . The upper approximation is the set of the elements which is likely to belong to X .

$$\begin{aligned} \underline{RX} &= \{x \in U : [x]_R \subseteq X\} \\ \overline{RX} &= \{x \in U : [x]_R \cap X \neq \emptyset\} \end{aligned} \quad (4)$$

$POS_P(Q)$ is the set of the elements which belongs to U/Q by the knowledge P , it defines in equation (5).

$$POS_P(Q) = \bigcup \underline{PX} \quad X \in U/Q \quad (5)$$

P and Q are the two equivalence relationships in U . If it is according with equation (6) as follows, then R belongs to P is called Q -unnecessary, otherwise R is called Q -necessary.

$$POS_{IND(P)}(IND(Q)) = POS_{IND(P-\{R\})}(IND(Q)) \quad (6)$$

$IND(Q)$ is the intersection of all the equivalence relationships belonging to P that is also a equivalence relationship, and called an indiscernibility relation of P . The intersection of all reduction attribute sets is the core attribute set, representing $CORD_Q(P)$. According to the theory, when the rough set is applied to data mining, P can be considered as condition attributes, and Q is the decision attributes. If R belongs to P is also belongs to Q -unnecessary, then eliminating R from the P will not change the decision in the

information sets, otherwise change. Jiang et al (2013) reveals the theory to divide the information data and eliminate the attributes of the Q-unnecessary can reduce the condition attributes.

RESULTS

After data analysis on the Vocational Counsel Centre Database of a university in 2013, we pick out the basic information and employment status data from the database of 126 representative graduates.

According to the above introduction, first, the attributes of the information sets can be reduced by rough sets theory. Second, a decision tree can be acquired from the data mining algorithm in the Table 1, in which, the employment status is a decision variable, denoted by D . The else attributes are all condition attributes, denoted by $A1$ to $A7$ respectively. The researchers take an attribute reduction on these condition attributes by using Matlab 2012b, and the result of attribute analysis is $Core_o(P) = \{A1, A2, A4, A5, A6\}$. That is, gender, personality, personal wishes, family wishes and professional skill are the final decision attributes. We transform the original data into corresponding attribute values in order to make the right format for ID3 algorithm, the transformation rules are listed in Table 1. The gender of male and female are changed into 0, 1 respectively, the three types of the personality, which are extroverted, intermediate and introverted are transformed to 3,2,1 respectively, and the professional skill, learning ability, English skill are all changed by the similar rules so as to ID3 algorithm can calculate the data quickly.

The researches apply ID3 algorithm to build the final decision tree. The final decision tree is shown in Figure 1. Only the branches which confidence level is higher than 70 percent are shown in Figure 1.

It can be seen from the analysis results of decision tree that personality and professional skill have a significant impact on college gradu-

ate career decision making. Most of the graduates with extroverted personality and excellent professional skill are able to get a good employment. It also can be seen that male college graduates easier to get a job than female with above good professional skill. On the other side, an introverted female college student with excellent professional skill and CET-6 is easy to get a postgraduate degree, however, if an introverted female college student has poor professional skill or learning ability, she will difficult to find a job in the employment market.

DISCUSSION

In general, it can be concluded that the graduates with excellent professional skill have many opportunities to be employed, and part of them who have CET-6 English level may choose to be postgraduates. While the graduates without excellent professional skill make their career decisions difficultly. The analysis results are consistent with the actual situation, and this can lead the employment education of college graduates. As the researches of Lee et al. (2015) presented, the education resources of per college will help the graduates enhancing their professional skill, however, this proposed results are also emphasize the personal learning ability of the college themselves.

Because the personality and professional skill is the most factors of a college to get a job, it should encourage the colleges to study well, the results is similar with the researches of Jun (2015). However, from the education or teaching perspective, the most important things is teaching and leading a college to know how to learn by himself, it depends on self-concept, which is the knowledge of the situation for himself/herself situation, in particular, and a college should give his/her career plan by his/her Self-Concept. The Self-Concept can get by the judge of experience, reflection and self-evaluation and feedback

Table 1: Basic information and employment status of 246 representative graduates

No.	Gender	Personality	Post treatment	Professional skill	Learning ability	English skill	Computer skill status	Employment
	(A1)	(A2)	(A3)	(A4)	(A5)	(A6)	(A7)	(D)
1	Male	Extroverted	Average	Good	Fast	CET-6	NCRE-2	Postgraduate
2	Female	Intermediate	Higher	Just so-so	Slow	CET-6	No	employment
126	Male	Intermediate	Lower	Just so-so	Slow	CET-4	No	Unsigned

of others such as parents. Self-Concept can make a college consolidate gradually self-concept and understanding, including cognitive own looks, recognition of the extent and so on. Self-concept is an organic structure constituted by a complex of attitudes, beliefs, feelings and values, which runs through all of the college's learning life. Recently, more and more colleges will focus on self-concept, they learn through classroom, participate in social activities and networking channels of access to be able to improve their ability and level of information, but also began to explore their interests, abilities and other expertise's project management and professional students as well. However, most of the colleges are not enough understanding of their own comprehensive and in-depth, and the results is that the colleges do not know their purpose of the professional skill training system in the university, so the university should help the colleges to realize their self-concept.

CONCLUSION

Making a career decision is a very important thing in a college graduates. This paper combines an actual situation of universities, and uses an information data table and decision tree data mining technology to build a decision tree to give some advising on employment education. Before building the decision tree, this paper takes an attribute reduction based on rough set theory, which reduces the computational cost of ID3 algorithm. This research can provide some suggestions for employment education of college graduates.

RECOMMENDATIONS

College students are difficult get to work because they lack of the knowledge on how to make a scientific, rational decision-making and show its ability to the employer. The result of the decision tree will help the college students to recognize the key factors such as personality and professional skills. We suggest that college should enhance the career education of colleges, and make the colleges focus on self-concept to improve their comprehensive quality and professional skills, which will make colleges get a job easier.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referees and the editor of this journal. The authors gratefully acknowledge the financial support by the National Natural Science Foundation of China (Grant no. 71401054 and 71403030), Beijing Planning Project of Philosophy and Social Science (Grant no. 14JGC108), the teaching reform research project of North China Electric Power University, the Fundamental Research Funds of the Central Universities of China.

REFERENCES

- Chan WK 2015. Higher education and graduate employment in China: Challenges for Sustainable Development. *High Educ Policy*, 28: 35-53
- Eesa AS, Orman Z, Brifcani AMA 2015. A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Syst Appl*, 42 : 2670-2679.
- Jiang Feng, Sui Yuefei, Cao Cungen 2013. An incremental decision tree algorithm based on rough sets and its application in intrusion detection. *Artif Intell Rev*, 40: 517-530.
- Jun K 2015. Go to graduate school or go to work: what factors determining college students' selections? *Anthropologist*, 19: 773-780.
- Lee KW, Chung M 2015. Enhancing the link between higher education and employment. *Int J Educ Dev*, 40: 19-27.
- Martincin KM, Stead GB 2015. Five-factor model and difficulties in career decision making: A meta-analysis. *J Career Assessment*, 23: 3-19.
- Thomas S, Chie QT, Abraham M, Raj SJ, Beh LS 2014. A qualitative review of literature on peer review of teaching in higher education: An application of the SWOT framework. *Education and Educational Research*, 84:112-159.
- Wei Chi, Bo Li 2014. Trends in China's gender employment and pay gap: Estimating gender pay gaps with employment selection. *J Comp Econ*, 42: 708-725.
- Wu Yajuan 2010. Graduate employment satisfaction statistics and prediction based on factor-cluster analysis. *Journal of Nanjing University of Information Science and Technology: Natural Science Edition*, 2: 510-513.
- Yang Ming, Guo Shuxu, Wang Jun 2011. Generating decision trees method based on improved ID3 algorithm. *China Commu*, 8: 151-156.
- Zhang Fengxia, Fan Yali, Zhang Fanlong 2010. Prediction for graduates to take part in the entrance exams for postgraduate schools based on Support Vector Machine. *Journal of Liaocheng University: Natural Science Edition*, 23: 4-7.
- Zheng Tingting, Zhu Linyun 2015. Uncertainty measures of Neighborhood System-based rough sets. *Knowl-Based Syst*, 86: 57-65.