Research on Employment Prediction and Fine Guidance based on Decision Tree Algorithm under the Background of Big Data

H Y Li¹ and Y Zhang²,*
¹School of Automation, Wuhan University of Technology
²School of Energy and Power Engineering, Wuhan University of Technology
*zhangyant1223@163.com

Abstract. Refinement of college employment guidance is the demand of the current era. Big data thinking and technology have created the conditions for it. In this paper, the factors influencing college students’ employment are discussed and the method of applying decision tree algorithm to predict college students’ employment is studied. On this basis, the paper analyses the effective ways to improve the refinement of college employment guidance based on employment prediction, from the four aspects of improving the scientificity of decision-making, the pertinence of guidance, the accuracy of service, and the timelines of assistance.

1. Introduction
With the rapid development of social economy and the expansion of college enrollment, the number of college graduates has been increasing in recent years, which is expected to reach 8.74 million in 2020. The employment situation is still severe. The demand of enterprises for the employability of the applicants tends to be diversified. Graduates' career choices are also more personalized. The traditional way of employment guidance is difficult to meet the needs of both the supply and demand sides of the employment market. In order to thoroughly implement the spirit of The Opinions of the State Council on Further Improving Employment and Entrepreneurship in the New Situation, in 2016, the Ministry of education required colleges and universities to establish and improve the precise employment service mechanism, so as to promote more full and high-quality employment of graduates. The refinement of employment guidance in colleges is the general trend.

Employment prediction refers to establishing employment model and predicting employment changes and trends in future, based on the historical data of college students' employment. Scholars at home and abroad have studied employment prediction, and the main methods are naive Bayes algorithm [1], natural neighbor classification algorithm [2], decision tree algorithm [3], BP algorithm [4], etc. The existing research on employment prediction mainly focuses on the optimization of prediction algorithm and the employment prediction of specific groups of students. Few have studied the application of prediction results, especially the application of individual prediction results to employment guidance and service.

This paper uses C4.5 decision tree algorithm to predict employment of college students, and analyzes the effective path to improve the refinement of employment guidance in colleges.
2. Refinement of college employment guidance under the background of big data
As an important product of the development of the Internet, big data has greatly changed people's life and work style. In the era of big data, scientific thinking mode and efficient technical means create conditions for improving the refinement of employment guidance in colleges.

2.1. Providing data sources of prediction by collecting massive data
Massive data information is the most basic feature of big data. Only by analysing the massive data of college students, can we provide sufficient data for employment prediction. Therefore, it is necessary to expand the ways of data collection and integrate the resource of different departments in colleges. In addition to the employment information of graduates and recruitment information of employers held by the employment management department of colleges, it also includes student consumption information, academic performance, book borrowing, reward and punishment information, etc.

2.2. Employment prediction by data mining
The production and collection of the mass data results in information explosion, and greatly impairs our ability to find meaning in vast collections of data. Using data mining technology, we can find the law from the massive employment information through fast and efficient calculation, increase the predictability of the trend of college students' employment intention, and understand the development level of college students' employability timely.

2.3. Application of prediction
Most colleges issue employment quality reports every year, but the data is not fully used to improve employment guidance. Through analysis and judgment, the results of employment prediction can provide scientific, readable and available information and reference for school administrators and personnel training program makers to improve the refinement of employment guidance.

3. An analysis of the factors affecting the employment of college students
We need to sort out the factors that affect the employment of college students, which are complex, in order to build a prediction model. Zhang Qingfang applied logit model to verify that the factors of students themselves, school, and society have an impact on the employment of college students[5]. Pan bin used AHP to analyse the internal factors such as students' achievements and political status, which are important factors influencing the employment of college students[6]. Sulastri A. took psychology graduates as the research object, and the results shown that GPA, extra-curricular activities, foreign language skills and some other factors had a significant relationship with finding professional related jobs [7].

Based on the research results of scholars at home and abroad, ensuring the independence and feasibility of the indicators, this paper mainly considers the impact of gender, source of students, political outlook, family economy, student cadres, awards, social practice ability, GPA, English level, computer level, etc. on the employment of college students.

4. Employment prediction based on decision tree algorithm
Decision tree is one of heated fields in data mining. The decision tree algorithm classifies the data based on the attribute value of the data, so as to build a classification model, which can classify the data into given categories. C4.5 algorithm is a classical decision tree algorithm, which is easy to implement, with a high prediction accuracy. It can meet the requirements of this study and will be applied to the employment prediction in this study.

4.1. Research objects and objectives
For research purpose, the employment data of graduates of a university or a major in recent 3-5 years can be selected as the research object. There are two research objectives: one is to predict whether students can be employed, which is the most basic research on the employment of college students, but
also the basis of employment warning and assistance. The second is to predict the nature of the enterprise, the location of employment and other issues, which will provide a basis for accurate employment services.

4.2. Data preprocessing
We should collect the basic information, academic information and employment information of the students from departments, such as Student Affairs and Dean's Office. Integrate multiple independent databases with the "student number" attribute, and delete the attributes that does not affect the employment of college students. According to the previous analysis, the reserved attributes include: gender, source of students, whether be a CPC party member, whether be a student from poor family, whether be a student cadre, award, social practice ability, GPA, English level, computer level, whether obtain employment, enterprise nature and employment region. The data is generalized as shown in table 1.

| Attributes                  | Attribute values                                                                 |
|-----------------------------|-----------------------------------------------------------------------------------|
| Gender                      | Male, Female                                                                      |
| Source of students          | Eastern regions, Central regions, Western regions                                 |
| Whether be a CPC party member | Yes, NO                                                                           |
| Whether be a student from poor family | Yes, NO                                                                              |
| Whether be a student cadre  | Yes, NO                                                                           |
| Award                       | Yes (school-level and above award), NO                                             |
| Social practice ability     | Good, General                                                                     |
| GPA                         | Excellent (GPA>=85), Medium(70<=GPA<85), Poor (GPA<70)                             |
| English level               | Good (passed cet6), General (not passed cet6)                                      |
| Computer level              | Good (passed ncre level 3), General (not passed ncre level 3)                      |
| Whether obtain employment   | Yes, NO                                                                           |
| Enterprise nature           | Category A (government departments, public institutions and state-holding enterprises, etc.), Category B (joint ventures and wholly foreign-owned enterprises, etc.), Category C (private and township enterprises) |
| Employment region           | Eastern regions, Central regions, Western regions                                 |

The error and incomplete records deleted, the sample set can be obtained, and the sample set should be divided into the training sample set $S_1$ and the testing sample set $S_2$.

4.3. Construction of decision tree
Taking gender, source of students, whether be a CPC party member, whether be a student from poor family, whether be a student cadre, award, social practice ability, GPA, English level, computer level as decision-making attributes, taking whether obtain employment, enterprise nature and employment regions as category identification attributes, three decision trees can be constructed according to the training sample set $S_1$. 
4.4. Generating classification rules
By traversing the three decision tree models from the root node to each leaf node in turn, three groups of classification rules can be obtained to predict whether students can be employed, enterprise nature and employment region.

4.5. Test of classification rules
Three groups of classification rules can be used to predict the test sample set \( S_2 \), and the predicted results should be compared with the actual employment situation of test samples. If \( N \) is the total number of samples and \( T \) is the number of samples with correct results, then the test accuracy \( P \) is expressed as \( P = T/N \). Generally speaking, if the test accuracy is above 85%, the model meets the design requirements.

5. Refinement of college employment guidance based on employment prediction using the decision tree algorithm

5.1. Improving the scientificity of decision-making
According to the position of the decision attributes in the decision tree in the employment prediction model, we can find the impact of the attributes on the employment of college students and make targeted employment guidance programs. For example, from the prediction model, we know that the factors such as GPA, social practice ability, and whether be a student cadre have a significant effect on the employment of students, then we should strengthen the construction of study-atmosphere, the improvement of practical ability, the appointment and training of student cadres in the whole process of education, so as to lay the foundation for the future employment of students.

In addition, the prediction of students' employment is conducive to timely monitoring the employment situation, analyzing the development trend of future graduates, and the formation of early warning mechanism for possible employment imbalance and other crises.

5.2. Improving the pertinence of guidance
At present, the employment guidance in most colleges mainly focuses on instilling the employment policy, employment situation and employment skills in all graduates, which ignores the individual needs of students. Some colleges offer one-on-one career counseling and guidance, which requires a high level of student initiative with a low level of coverage. Personalized guidance can be provided according to the needs of different students based on the employment prediction results. The employment guidance plan can be customized for each student. For example, training of foreign language interview skills can be provided for students who are likely to enter foreign companies in the future, while guidance on the examination of public institutions can be provided for students who are likely to enter state-owned enterprises in the future.

5.3. Improving the accuracy of service
The function of employment prediction can be added in the college student employment service system. Students can make employment prediction in advance and clear their personal positioning. If the predictions are consistent with their own employment expectations, it can effectively improve students' confidence in employment. If there is a gap between the prediction and one's own employment expectation, one can reflect on the accuracy of one's own positioning on the one hand, and improve one's employability by referring to the classification law on the other hand.

The college student employment service system can provide students with career reference data and personalized recommendation of recruitment information according to the predictions. The employment reference data refers to the specific employment data of the previous graduates corresponding to the same classification rules as the student himself, including industry information, enterprise information, position, salary, etc. Personalized recommendation of recruitment information refers to adjustment of the order of the recruitment information personalized according to the
employment prediction. Recommend suitable recruitment information for students, avoid the interference of irrelevant information and waste a lot of time, so as to help students to select jobs quickly and accurately.

5.4. Improving and the timeliness of assistance

For students who may be difficult to obtain employment, early warning, necessary assistance and guidance shall be given in time. For students with excellent comprehensive ability who are predicted to be difficult to find employment, we should pay attention to whether they have the idea of "slow employment" or other personal plans, pay attention to the education of employment concept, and help them set reasonable employment goals. For students who may find it difficult to find a job due to their lack of employability, we should help them to analyse their strengths and weaknesses, build up their confidence, and develop a plan to improve employability.

6. Conclusion

This paper studies the method of college student employment prediction using the decision tree algorithm, and analyses the effective ways to improve the refinement of college employment guidance based on employment prediction, from the four aspects of improving the scientificity of decision-making, the pertinence of guidance, the accuracy of service, and the timeliness of assistance.

Acknowledgments

This research was financially supported by the Research Fund on Student Work for Wuhan University of Technology (Grant NO.201807) and the Research Fund on Teaching Reform for Wuhan University of Technology (Grant NO. w2018081) and the Research Fund on Teaching Reform for Hubei (Grant NO. w2017015)

References

[1] Masethe M A, Masethe H D, Prediction of Work Integrated Learning Placement Using Data Mining Algorithms 2014 J. Proceedings of the World Congress on Engineering and Computer Science, vol5.

[2] Zhu Q S, Gao X, Model of College Students' Emolument Prediction Based on the Classification Algorithm with Natural Neighbor 2017 J. Computer Systems & Applications, vol26(08) pp 190-194.

[3] Tang Y, Wang P, Study on employment forecasting of graduates of traditional Chinese medicine based on C4.5 and random forest algorithm 2017 J. China Medical Herald, vol14(24) pp166-169.

[4] Jiang Z T, Yuan Z S, Yan R L, College Students Employment Forecasting Model Based on IAFSA-BP Parallel Integrated Learning Algorithm 2019 J. Value Engineering, vol38(19) pp232-234.

[5] Zhang Q F, On the Comprehensive Factors that Affect College Students' Employment: An Empirical Analysis Based on the Logit Model. 2014 J. Journal of Hunan University of Science &Technology (Social Science Edition), vol17(03) pp175-180.

[6] Pan Bin, Analysis of Influence Factors on Current Employment Ability of Agriculture and Forestry University Students 2012 J. Journal of Anhui Agricultural Sciences, vol40(08) pp5056-5058+5072.

[7] Sulastri A, Handoko M, Janssens J M A M, Grade point average and biographical data in personal resumes: predictors of finding employment 2015 J. International Journal of Adolescence and Youth vol 2.