

Advanced in Control Engineering and Information Science

Data mining of graduation project selection database

Wu Xie^a, Huimin Zhang^b, Bizhong Wei^a, Guanghai Fang^{a*}

^a*School of Computer Science & Engineering, Guilin University of Electronic Technology, Guilin 541004, China*

^b*School of Mathematics & Computing Science, Guilin University of Electronic, Guilin 541004, China*

Abstract

In order to improve the quality of graduate project, a database system of hundreds of graduation project selection results was established in C# language of Visual Studio. The information system was data mined with ID3 algorithm, and a decision tree is gained for researching these graduation projects choices. The mining results of software testing demonstrate that the quality of graduation project selection is associated with the difficulty, project direction and major direction mostly, which guilds the students to choose their suitable graduation projects in time, improving the efficiency and quality of graduation project selection greatly.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS 2011]

Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Data mining; database system; decision tree; classification; graduation; selection;

1. Introduction

In recent years, the quality of graduate design tends to decline in many colleges in China. The title of graduation project involves in the work of all graduates for a semester. Successful completion of graduate design with high quality depends on the important factor of project selection. The projects usually stimulate creativity and enthusiasm of the students greatly. Science graduation project topic is the primary task for most undergraduates. From the perspective of teaching reform, some literatures state the quality of graduation subject choosing, while qualitative analysis is too much with few quantitative studies [1-7]. Especially, it is very difficult to reflect the importance degree and deeply interdependent relationships of those factors that affect hundreds of students to choose their graduation projects at the same time,

* Corresponding author. Tel.: +38 135 5713 4096.

E-mail address: xiewu588@126.com

leading to little effect during improving graduate design management with some problems recurring year by year. Students often have to choose their projects in a specified time after they only rush to catch the topic information, lacking talking with their instructors. At the beginning, they are not familiar with the issues of the selected projects, resulting in the low score of the graduate design. Some topics absorb many students' elections, while others have few candidates. The poor quality of all selections influences the overall effect of graduation greatly.

Therefore, a quantitative method by data mining of graduation project selection database is proposed to improve the quality for facing a lot of graduation choices. Before data mining, Section 2 will describe data preparation work through constructing a database system from graduation project selection results within five years. The data mining method for graduation project selection results database is followed in Section 3, while Section 4 shows the data mining results and discussion. Finally, in Section 5, an indication of some rules or conclusion from this paper is demonstrated.

2. Data Preparation of Graduation Project Selection Database

A database of graduation project selection results is needed establishing to analyze the main factors that affect the quality of scores before data mining. Therefore, some important attributes are investigated to construct the relational tables of the database.

The directions of graduation selection are often connected with education scheme of undergraduate courses. Professional direction is corresponding to different training objectives and curricula. All graduates need to use comprehensively those theoretical knowledge and major skills during graduation project. So, the titles of graduate project should be consistent with major direction. For example, for the major of computer science and technology, the direction includes database, network and hardware, etc.

Graduation projects are usually from research projects, optional subjects, courses construction with diverse difficulty. The easy titles gain more choices, while the difficult have few candidates, especially for some projects from actual engineering. For example, projects directions of computer major are database, network and hardware, etc, and the difficulty of the subjects is divided into high, medium and low level.

Easy question often gets more choice, but the students have fewer opportunities to exercise from the engineering design and practice. Graduates usually defuse the difficult titles, missing the opportunity to work in company. As the computer science major, the direction of the subjects also divided into the database, network and hardware. The difficulty of the subject is divided into high, medium and low.

The quality of graduation design is related with results of choosing titles, and the grade is good, medium or poor. After projects selection, the speed to finish all the work is associated with the attitude, difficulty, interest, employment, supervisor, graduate retest, civil service exam, which interfere the progress greatly. Here, the speed of graduation procedure is divided into the fast, medium, slow.

The graduation selection quality results are divided two kinds, which P stands for the good, while N expresses the poor.

Therefore, the process of establishing a database, select the school number, name, professional direction, the direction of topics, difficult topics, completing quality, completing speed and grade of selection as the relational schema.

The students are assigned a number No as the primary key during designing the relational table of graduation selection results. According to the database principle, all of the attributes above can not be divided into other column data; therefore, this table meets the first normal form (1NF).

As the major direction, title direction, difficult, quality of completion, speed of completion, grade of graduation selection are fully dependent on the number No, the table of selection results meets the second normal form (2NF).

According to database principle, the relational table of graduation selection results does not meet the third normal form (3NF). There are transmission dependences in this table, which is not Boyce-Codd normal form (BCNF).

3. Data Mining of Graduation Project Selection Results Database

The key to analyze the relationship between of the project selection results and the graduation quality is to find which attributes play important role on the dependence relationship. The data mining is one of the best ways to solve the problem. To improve the work of selecting the graduation projects, the results of choosing the titles need classification. Here, the ID3 decision tree algorithm is adopted, and the database system of graduation selection results are data mined with five steps.

- (i) The mutual information values are calculated from all the attributes of the relational table of the graduation projects selection results.
- (ii) The maximum mutual information value from all the table attributes is chose.
- (iii) The attributes of the relational table of the graduation projects selection results are classified by the maximum mutual information value, and a decision tree is gained.
- (iv) The ID3 algorithm is recursively called for the attributes of the relational table of the graduation projects selection results excluding the attribute of the maximum mutual information value in the above (ii), and some sub-trees are obtained.
- (v) The leaf nodes of the final decision tree are indicates the grade type of graduation selection results. The numbers of students are shown in these leaf nodes.

4. Data Mining Results and Discussion

To calculate large-scale data records of graduation projects automatically, the software of ID3 decision tree algorithm are developed with Visual Studio 2010 in C# programming language. The database system of the graduation projects selection results is data mined with ID3 algorithm by the software, and a decision tree are gained, as shown in Fig. 1.

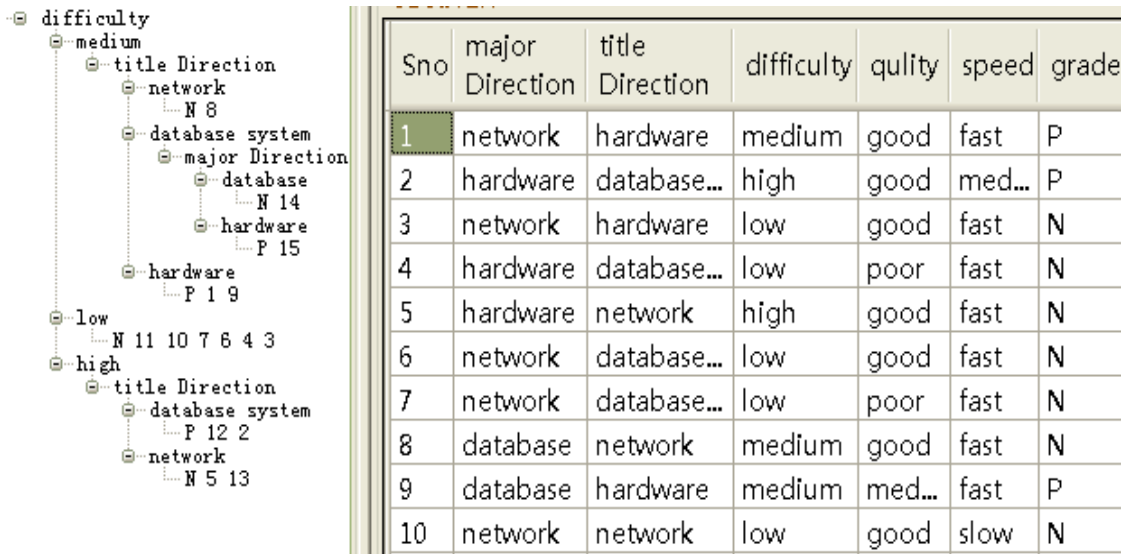


Fig. 1. Data mining results of graduation project selection database

From the mining results of graduation projects selection on the left side of Fig.1, the attribute of difficulty of the decision tree is the main factors that influence students to choose their project titles, and then they consider title direction and major directions, which are actually consistent with the case of graduation.

In order to facilitate understanding and application of the data mining results, the decision tree can also be transformed into associated rules:

(i) If the difficulty is low, then the students are many.

(ii) If the difficulty is high, then the students choose the projects by title direction, the database and network are more than the hardware.

(iii) If the difficulty is medium, then the title direction affect mostly student to choose, and the second is major direction. The key graduation selection factors are difficulty, title direction and major direction. The quality of the final graduation design is associated with the grade of projects selection results, while the graduation speed affects little.

5. Conclusion

A database system of graduation project selection results is created as a data resource, which is data mined with ID3 algorithm. The decision tree is gained, finding out the main factors that mainly affect the selection of graduation titles. Then the work of graduation project selection can be improved with the decision tree in the future. The quantitative research and analysis in this paper can reduce the probability of change projects during graduation, leading the graduates to select the graduation projects by their major direction, expertise, interests, abilities, knowledge. So, the efficiency and quality of choose the title of graduation can be improved greatly, strengthening the teaching management of graduation from the source fundamentally.

Acknowledgements

The authors would like to thank Guangxi Zhuang Autonomous Region Education Department project (No.2010JGB035, No.2010C059) and the teaching reform of Project from Guilin University of Electronic Technology (No. JGZ201009).

References

- [1] Akorede, M. Guidelines for Writing an Undergraduate Engineering Project. *IEEE Potentials* 2009; 28, 10–15.
- [2] Neish G, Collins L. If you ask me. *Engineering & Technology* 2009; 4, 20
- [3] Goldberg. Capstone Projects and National Student Design Competitions. *IEEE Pulse* 2011; 2,7
- [4] Huawei Z, Yu Z. The Quality Control in Subject Choosing of the Graduation Design of Electronic and Information Engineering. *Journal of Higher Education in Science & Technology* 2010; 29, 74–77.
- [5] Yawu M, Fangping X, Wen-kai C. Start from the scientific topics designed to improve the quality of graduate. *Hunan Agricultural Machinery*, 2010; 37, 162–163.
- [6] Zhi-jun Y, XU Ai-jiang X. Quality Control in Topic Selection for Undergraduate Graduation Design and Thesis. *Journal of Ningbo University of Technology* 2008; 20, 108–111.
- [7] Zhao Jinghua Z, Liang M, Yingjie M, Zusen C, Pang Yi P. Study on professional graduation project mode for agriculture and water conservation specialty. *Experimental Technology and Management* 2010; 27, 133–135.