

Students' comprehensive evaluation system based on data mining method

He yongrong

School of Computer Science
Zhejiang International Studies University
Hangzhou, China
hangdianproe@163.com

Bian xiangjuan

School of Computer Science
Zhejiang International Studies University
Hangzhou, China
bianxiangjuan@163.com

Abstract—the paper researched comprehensive quality evaluation system of college student, and introduces data mining method to solve mess students data. Firstly, the paper construct students' data warehouse, then construct comprehensive multi-dimensions OLAP (Online Analysis Process) model, and through classification deterministic method, we can get useful message from the system to guide college manage students, and some Employers can find their satisfied employees.

Keywords-Data mining, comprehensive evaluation, OLAP

I. INTRODUCTION

With the higher education change from elite education to popular education, there are more and more student's management problems. The students' data is huge and complicated, and students' state and development is also can't forecast. So the college managers hope to get a students' management system which has auxiliary deterministic ability and data mining technology just provide the effective method to solve the problem. Data mining method try to find useful information from huge data, this is the processes which find relationship between model and data from huge data^[1], this model and relationship can use to forecast. Under this background, utilize data mining technology in students' management system, construct perfect students' evaluation system, which can improve the students management level and speed up students' management work specialized^[2]. The paper firstly design students data warehouse, then construct students' data mining model which include: students information data mining, students' course selection data mining, students' obtain employee chance data mining, at last, we can obtain data mining result to get some reason about: the factors influencing student achievement, the factors influencing students course selection and factors influencing students obtain work chance, The whole process is just like figure 1.

II. STUDENTS COMPREHENSIVE EVALUATION DATA WAREHOUSE CONSTRUCTION

To finish data mining work, the first thing is prepare data, according to different department and spatiality's students' information data, so all collective database should be reorganized and classified, systematized, to finish these work, the only thing to do is transfer these data to data warehouse. As teaching manager, their work often face deterministic analysis, their most attractive information is student's

graduation information and the information of the students adapt the society need. So, we use recent 5 years different department's students graduation situation, course arrangement, employment tendency as basic data, the data warehouse structure is like figure 2.

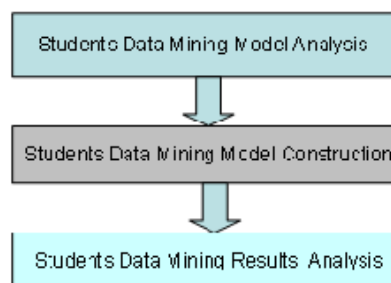


Figure 1: The data mining process in the system

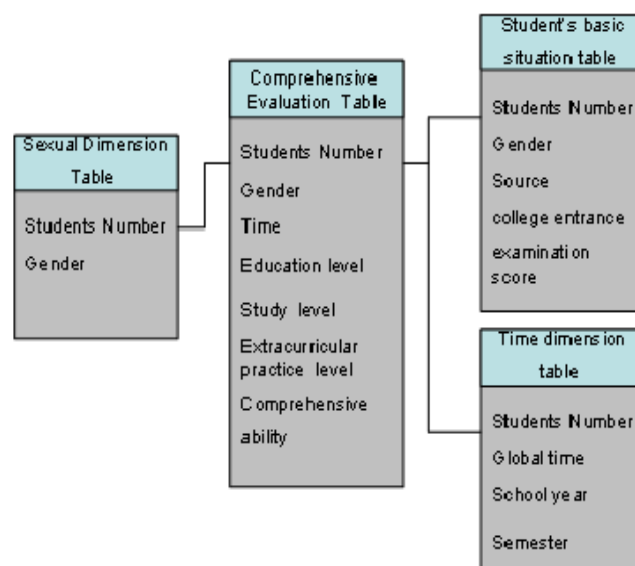


Figure 2: Students comprehensive evaluation data warehouse structure table

The student's comprehensive data warehouse was divided as computer department database, english department etc, then divided as thirteen data mart, the whole process:

Step 1 : construct data warehouse model function: determine system main body, and relationship with these bodies, refine the body's every lever, such as , educational administration management system mainly divided as

students' achievement analysis theme, course arrangement theme, students' arranged employment situation;

Step 2: data warehouse physical database's construction: define the system physical database model, not only record's type, default value, and constraints relationship, but also some indexes and physical view;

Step 3: data extraction, transformation, and integration: adopt Microsoft SQLSever tools extracting appoint record, delete unqualified data, and processing preliminary data integration.

Step 4: data import: when creating data warehouse, data transforming function of Microsoft SQLSever is necessary, because data from other database should be selected, processing, and loading to data warehouse.

III. CONSTRUCTION THE DETERMINISTIC TREE OF STUDENT'S COMPREHENSIVE EVALUATION SYSTEM

The paper utilize deterministic classified theory to construct deterministic tree of student's comprehensive evaluation system, to realize qualitative analysis, the whole data classified process is like figure 3:

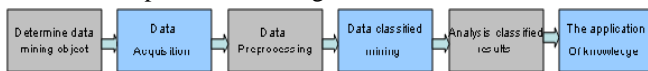


Figure 3: The whole process of data mining

Data mining classification is the key step of data mining application, to realize it, a suitable algorithm should firstly be selected, and also a proper program must be found to realize this algorithm. The paper adopt the famous deterministic classified method: ID3 algorithm^[3], the algorithm trains all samples from root node, select an attribute to divide these samples, every value of attribute generates a branch, and then transforms branch attribute value of samples subset to a new node. This is a recursive process which has been used at every node, until every sample at node has been divided to a type. The figure 4 is showing the generating process of deterministic tree. ID3 algorithm is most typical deterministic tree algorithm; it was first put forward by J.R.Quinlan. The main core ideology of the algorithm is the greedy search algorithm which selected maximum information gain as current deterministic attribute. ID3 algorithm can keep depth of every branch of the deterministic tree is least, here give the ID3 algorithm detailed description:

```

ID3(table, attributes)
//input: training set table, attributes;
//output: Deterministic tree;
{if(table is empty)then Return(null);
N=a new node;//create node;
If(there are no predictive attributes in table)//the first situation
then label N with most common value of attributes in table(deterministic tree)or with frequencies of attributes in table(probabilistic tree);
else If(all instances in table have the same value V of attributes)then/the second situation

```

```

label N,"X.attributes=V with probability 1" else For each
attribute A in table compute AVG
ENTROPY(A,attributes,table);
AS=the attributes for which AVG
ENTROPY(A,attributes,table)is minimal;
If (AVG ENTROPY (A, attributes, table) is not substantially
smaller than ENTROPY (attributes, table)) then Label N
with most common value of attributes in table (deterministic
tree) r with frequencies of attributes in table (probabilistic
tree) ;
Else label N with AS;
FOR EACH VALUE V of AS DO
N1=ID3(SUBTABLE(table,A,V),attributes));
IF(N1!=null)then make an arc from N to N1 labeled
V;
End
End
End
Return N;
End

```

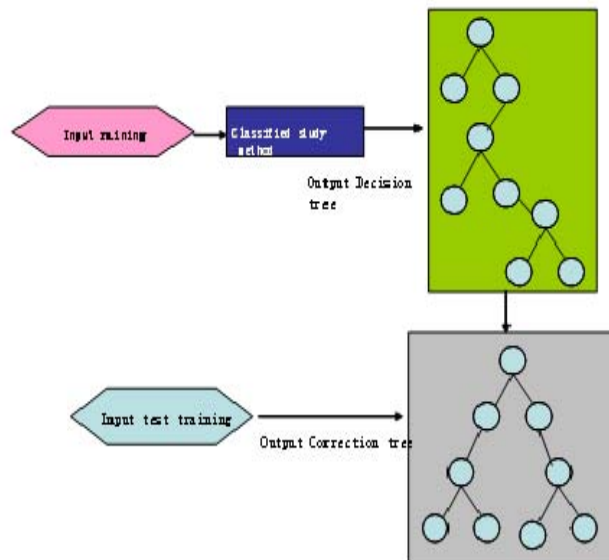


Figure 4 the generating process of deterministic tree.

The deterministic tree's generation passed two stages: studying and testing. At studying stage, deterministic tree adopt top-down way's to finish its recursive process, when the tree began to generate, all data is in root node, then divide by recursive way, until branch node generated. The second stage is testing stage; this stage's main aim is to delete some noise and unmoral data. To stop the recursive process, there are must be satisfied a condition: the data at same node should must be same type; no other attribute can be divided by data.

Students comprehensive evolution's class was divided by three areas: moral level, academic level, extracurricular practice level^[4], the order is: moral level, academic level, extracurricular practice level; all the rules are constructed by if/then language:

```

If (moral level1) then

```

(if academic level1 or 2) and (extracurricular practice level1) then (comprehensive quality1)
 (if academic level1 or 2) and (extracurricular practice level2) then (comprehensive quality2)
 (if academic level1 or 3) and (extracurricular practice level1) then (comprehensive quality2)
 else(comprehensive quality3)
 If (moral level2) then
 (if academic level1) and (extracurricular practice level1) then (comprehensive quality1)
 (if academic level1) and (extracurricular practice level2) then (comprehensive quality2)
 (if academic level2) and (extracurricular practice level1) then (comprehensive quality2)
 (if academic level2) and (extracurricular practice level2) then (comprehensive quality3)
 (if academic level3 or 4) and (extracurricular practice level1) then (comprehensive quality3)
 else (comprehensive quality4)
 If (moral level3) then
 (if academic level1) and (extracurricular practice level1) then (comprehensive quality3)
 Else
 else (comprehensive quality4)
 According to classified rules, we can get deterministic tree like figure 5:

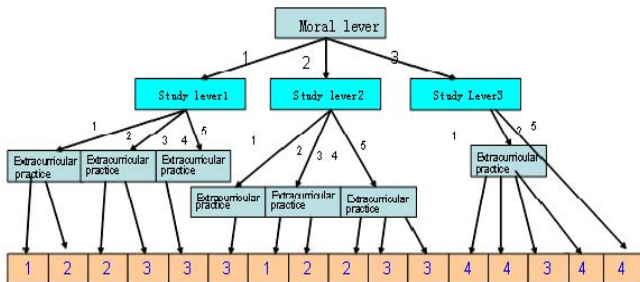


Figure 5 Students comprehensive evolution's deterministic tree

IV. STUDENTS COMPREHENSIVE EVOLUTION PROTOTYPE SYSTEM DEVELOPMENT

The system adopt multi-dimension model , the model can be seen as three-dimensional model, database model was adopted, the processed data can make following online analyzing easily, the system can be searched by student number, scheme, school year. The system was divided as three isolated aero: moral, academic, extracurricular practice. The moral data often come from themselves' evaluation and evaluated by instructor; The academic data includes core courses, required courses, and elective courses, when computing scores, can give different weight to Corresponding course, and standardized all the score, at last give the result by orders; extracurricular practice adopt bonus system, divide some little project, every project has its upper limit, then standardized the scores. The whole system functional model is like figure 6: The Fig 7 and Fig 8 are the system working interface.

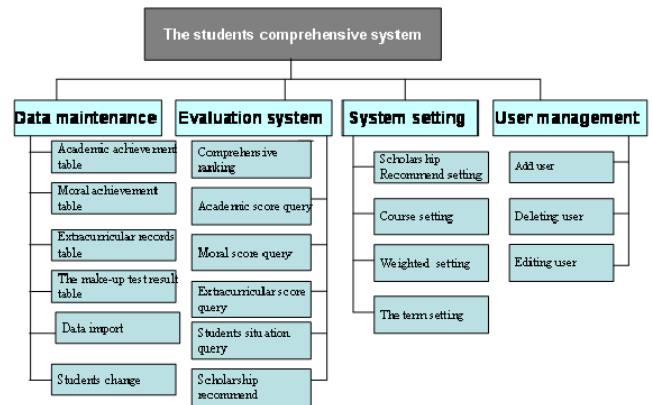


Figure 6 the students comprehensive system functions



Fig 7 Integral rendering of The System

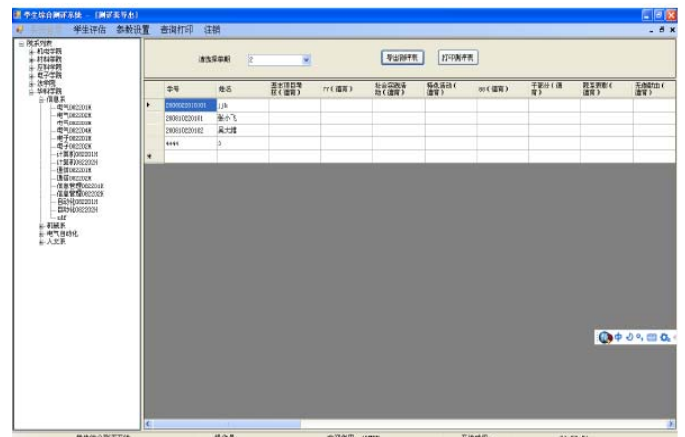


Fig 8 Rendering of Evaluation Result

V. CONCLUSION

Blend the existing management system in a suit perfect students management information system and adopt data mining technology to get some useful information, which can widely used in students comprehensive evaluation, graduate interview recommendation and enrollment analysis etc, it can improve students management level and speed the students specialization .

ACKNOWLEDGMENT

The paper was support by Zhejiang new talent program project: “Student integration capability evaluating system based on data mining technology”, The National Natural Science Foundation of China “MEMS Multi-fields Uniform Simulation Model Construction and Optimism Research” (61100101), the paper was also supported by Key Discipline of The Ocean Mechatronic Equipments Technology.

REFERENCES

- [1] Bodea, Constanta-Nicoleta, Bodea etc. Student performance in online project management courses: A data mining approach: 3rd World Summit on the Knowledge Society, WSKS 2010, 2010, 470-479
- [2] Ogor, Emmanuel N. Student academic performance monitoring and evaluation using data mining techniques . Electronics, Robotics and Automotive Mechanics Conference, CERMA 2007,2007,354-359
- [3] Vialardi, César ; Chue, Jorge; Peche, Juan Pablo etc. A data mining approach to guide students through the enrollment process based on academic performance, User Modeling and User-Adapted Interaction, v 21, n 1-2, p 217-248, April 2011
- [4] Zhang, Zhiyu1 . Study and analysis of data mining technology in college courses students failed, 2010 IEEE International Conference on Intelligent Computing and Integrated Systems, ICISS2010,2010. 800-802
- [5] YANG Hong-ying, LUO Huan, HE Qiang. Data Mining in Network Education Based on Decision Tree. Computer Knowledge and Technology, Vol.6,No.10, April 2010, pp.2313-2314