An Automatic Scholarships Evaluation Method Based on Machine Learning

Ji Xinhao

Zhejiang Business College
East China Normal University
Hangzhou, China
e-mail: jxh@zjbc.edu.cn

***Abstract* — The paper designed an automatic evaluation method with Support Vector Machine, SVM-based scholarships. This method assesses scholarships as a classification issue: whether the scholarship is obtained or not. According to learning from the previous assessment data, the method produces the classification model; then it can be used to automatically assess student scholarships and eliminate the influence of human factors in the evaluation process. The automatic evaluation method deals the student scores with standardization process to eliminate impacts due to inconsistent scores. In order to remove the distinction of specialty training program, the method indicates to classify the scores for unification of feature quantity. As a result, the experiment findings proved the effectiveness of this method.**

Keywords - component; formatting; style; styling; insert

I. INTRODUCTION

Currently, scholarships assessment is an important task of teaching management. scholarship plays a role in promoting the enthusiasm of studying; as well as the schools' recognition and affirmation to students who study hard, which will certainly be a positive feedback their studies; in addition, it also has guide functions, especially making a unique contribution in certain areas, such as the school of computer science awards the students who win the acm programming contest with special scholarship, which can effectively stimulate others for participation. All in all, the scholarship has great meaning for promoting the teaching and scientific research, as well as the physically labor-round development of students.

However, we should also see: an unfair assessment of scholarship will even bring negative effects on students rather than positive impacts. Thus, it will affect the students' enthusiasm for learning and guide them to be opportunistic or even curry favor with teachers. If the evaluation results of scholarships are not accepted by all students in class, it is hard to promote other students to work harder.

At present, scholarship is evaluated based on assessing the weighted summation and ordering scores. Weighted summation assessment method is intuitive and easy to operate. However, there existing some problems, for instance, the determination of weight in each score, which generally requires teachers or specialists to determine and has subjective factors. In addition, the weighted sum method can only reflect the student's overall average level, which is a comprehensive evaluation; while the universities aim to cultivate professional skilled talents in addition to comprehensive talents, for example, students majored in computer science who can obtain the senior programmer certificate are considered as talents in software field development, then, scholarship evaluation rules should provide higher possibility for them to receive scholarships. Nowadays, the method only gives such students extra points in the weighted sum equation, which requires scoring by teachers with various human subjectivity impacts.

To reduce the impacts of human factors in assessing the scholarship, it is a feasible idea to combine the automatic evaluation of machine with human evaluation. Currently there are some relevant scholars discussed and gave out related methods. Chen Hailing [1] proposed a fuzzy mathematical method to establish fuzzy evaluation function through the students, and give the final evaluation on the basis of fuzzy functions. Ma Baoyun [2] proposed postgraduate scholarship evaluation method, which is concerned about the degree of difficulty and score differences in courses. To solve this problem, Ma Baoyun presented the characteristics based on normal distribution and scores correction of different courses. Liao Yiqiang [3] proposed the evaluation project by standard division model, the difficulty factor model, dictionary sort model and normal distribution model. Existing methods ultimately apply the weighted summation to assess scholarships, which focus on standardizing scores and reducing effects caused by teachers, course difficulty and other factors.

For the case of only concerning the comprehensive points, this paper propose an automatic scholarship evaluation method based on machine learning. This method learns laws from existing data in assessing awards and its decision-making regulations, and utilizes new scholarships evaluation to lower down human impacts, and can award talents who have special skills with scholarships in assessment.

With the current advances in artificial intelligence, machine-learning algorithms received wide attention. Many scholars made a lot excellent learning algorithms, such as Adaboost, SVM, deep learning, etc. Deep learning is considered as a popular way to learn, but it requires a huge data sample to make the algorithm effective, while for scholarship evaluation, it is difficult to maintain such a wide range of data. The main problem shows that with the temporal variation and teaching program exchange, the test scores and content for students have also changed greatly. Besides, Adaboost algorithm result form a combination of weak classifiers to strong classifier, and weak classifiers in general is characterized by single classification decisions, therefore, it is easy to filter students who may win awards by single feature. In this paper, we try to use SVM learning methods of scholarship assessment rules. SVM can be used in a smaller sample database, and give a good decision plane so as to classify the sample. After input the student scores, the assessment is devided into two categories: obtained the scholarships or not.

II. SVM CLASSIFICATION METHOD DESCRIPTIONS

SVM classification method is a small sample library classification whose effect is very good classification algorithm, it is based on the theory of VC dimension and structural risk minimization, and it can effectively reduce the problem of over saturation and the need for full sample. The SVM classification method of learning from the classification of the optimal separating hyperplane, requiring hyperplane which can not only separate the two kinds of the sample library, but also make the maximum classification distance. In addition, for a linear non separable sample set, the SVM can be through the adoption of relaxation factor at the same time considering the minimum wrong sample and the maximum classification interval to get the generalized optimal separating hyperplane. Sample points on hyperplane is called support vector. For some nonlinear problem, you can make the linear separable in the high dimensional space by nonlinear transformation.

SVM classification learning methods exhibits great advantage in solving small sample, nonlinear and high dimensional input space classification troubles. The paper applies LIBSVM open-source software for training and classifying the SVM models. LIBSVM is produced by Dr. Chih-Jen Lin Taiwan University which is a simple operated, fast and effective generic SVM software package, can solve the classification and regression problems, provides a linear, polynomial, radial basis function and four Sigmoid common kernels, and cross-validation tool to select parameters and SVM model learning tool.

III. SCHOLARSHIPS EVALUATION METHOD BASED ON SVM

This method is used to obtain the scholarship classification decisions plane by studying existing data and then to input the unknown data for the automatic classification, therefore, the entire learning and classification process can be divided into two parts. The flow chart of the overall method is given in Figure 1.

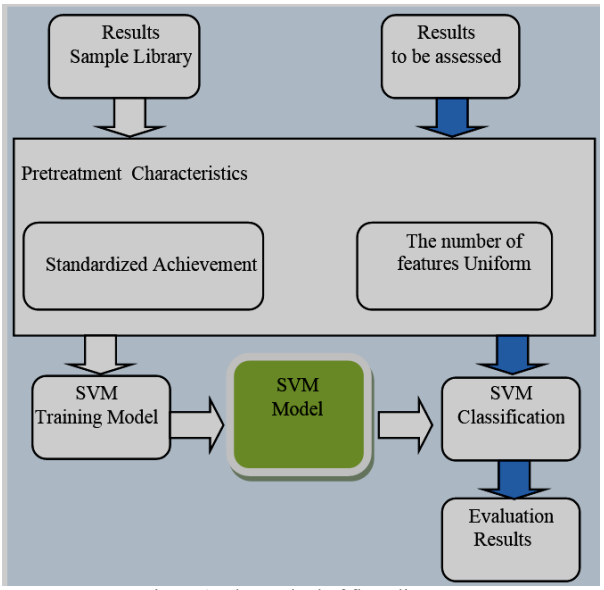

Figure 1. The method of flow diagram

In learning process, treating the results of the existing student scores and expert assessment as the training samples; no comment on the students' achievements as negative samples, the commentary scholarship students' achievements as positive samples; each achievement is managed by feature processing to standard feature vector input; then trained SVM classification model. During the experiment, after entered the student score by feature preprocessing and input it to the SVM classifier prediction, it will be determined whether he wins scholarship or not. As a scholarship for the assessment of the identification process, the key is to select and process characteristics, and select the machine learning algorithms. The method is presented in details below on student achievement pretreatment process.

A. Standardized academic scores

Due to the different scores from students in different classes are susceptible to different teachers, different papers and other factors, the scores differ a lot from classes in same course. This method of classification is based on a unified evaluation criteria, thus requiring different classes, different periods of standardized scores. There has been some fraction of the current standardized methods. Suppose a student's score is x_i, the minimum and maximum values of the classes of the course grade are

respectively min and max. One available method is to use the maximum quantization method, i.e. standardized score=100* x_i /max, and minimum and maximum extension methods, i.e., the standardized score=100*(x_i - min)/(max-min). The highest and lowest scores in both two methods are vulnerable to occasional factors, and do not reflect the overall learning situation in whole class. For example, One student is keen on a course which leads to the high performance, the method will result in great impacts on scores of other students by the highest score method rather than the impacts of teacher and papers complexity. In addition, each student can be standardized by normal distribution of scores. The method indicates that the student's scores that follow a normal distribution can be quantified by the average and standard deviation. However, many course points do not always match normal curriculum, for instance, teacher ratings, the award-winning scores, as well as some tricks of examinations.

Through above analysis, the method utilizes mean quantization method to standardize scores for one class in certain course, which is calculated as follows:

$$\overline{x_i} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where $\overline{x_i}$ is the average scores in the course. This standardized approach can maintain the difference between the scores of students, at the same time; eliminate the complexity of different teachers and tests.

B. The unification of quantity features

Machine learning method is used to regular assessment mining by historical data and expert evaluation results. Regardless of learning phase or stage classification, machine-learning methods are inputted feature vector. In the evaluation of scholarship applications, we need to represent all aspects of student achievement as a feature vector. A simple approach is to assess every scholarship as a characteristic feature, for example, teacher ratings, the results of a course and so on. However, different majors' courses are not the same, to set up automatic scholarship assessment algorithm for whole school, it requires consideration of inconsistencies between the different professional courses. In addition, the historical data of teaching program would be adjusted in different periods, different selective courses of the students are not the same, it should also be considered in assessing awards obtained by the learning methods applied at different times as well. Therefore, it requires great attention focus on how to learn the data that are not exactly the same and adapt to the changes of the various features based on the evaluation method of machine learning.

The method is treated as partition method aims at scholarship assessment features. The evaluation of the scholarship is divided into four basic categories: moral education division, intellectual stars, sports stars and skill points. Moral components consist of teacher ratings, apartment scores, class work scores and reward points. This method regards each feature as a moral score, a total of four features. According to the current teaching program, which consists of nine sub-intellectual freshman course grades, each course has corresponding credits, larger credits the means larger importance. This method will credit the same course, obtained by the weighted average credit score as a feature of the course is to reduce the impact caused by changes in curriculum; in addition also includes a sub-intellectual intellectual reward points. Therefore, the moral education credits classified section is divided into five grades and performance characteristics of reward intellectual characteristics. Sports stars include: physical fitness test, morning run score sports competition points, three results corresponding to three characteristics. Evaluation methods for skill points for passing a skill test plus one points, and expressed as a feature. From the analysis of specific examples above, the method presented through a 13-dimensional feature characterized in student achievement vector.

Approach above effectively reduces the impact of curriculum changes. When adding a course, the method can be fused according to its average weighted by credits to the corresponding features. When reducing a course, if its reduction in credit courses corresponded with features, other courses will not be affected.

C. Learning classification

This method uses a standard SVM learning approach. The score date of various majors and different learning stages should be prepared before training. Each student performance through the number of features and a unified standardized achievement obtained after 13-dimensional feature vector, and each student has evaluated the results of the assessment scholarships. The corresponding eigenvectors and evaluation result is treated as a sample. For the results in this method were standardized, scores should no longer be normalized during SVM training. Training support model parameter vector set are as follows: Error penalty coefficient C of 2000, gamma is 1.8445, radial basis function kernel function (RBF). Students are training samples is to get a scholarship; negative samples are students in the student sample library did not get a scholarship. The model after training can be effective to automatically determine whether the student can earn scholarships or not.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the effectiveness of this method in the scholarship automatic evaluation, we constructed a sample library. The sample library includes: freshman year all students of various professional assessment scholarships achievement data and assessment results, a total of 671 students. From these data, in which 71 students were randomly selected data as the test data, including scholarship selected 15; the other 56 students did not get the scholarship data. 600 student achievement as sample data to train the model parameters of SVM.

Learning from the evaluation target detection method, using two metrics: precision and the effectiveness of recall rate. The evaluation method will test the database of all students who receive scholarships set (T) as a set of real data; and students receive scholarships of the present method of automatic classification of the set to give (E), as estimated data collection. Automatic classification method of collection was a collection of award-winning student (E) with the actual data that matches students as a valid result set (C). Accuracy is defined as |C|/|E|. Recall is defined as |C|/|T|. The determination of recall rate reflects the accuracy of the award-winning list of students. Recall rate also reflects the ability whether the method can classify winning students from the student set out. An excellent automatic evaluation method should have a high accuracy and recall rate.

To test the impact of the different methods of standardization scoring method based on this experiment we tested the maximum quantization method, based on the minimum and maximum extension method, based on mean quantitative methods, and each method for independent learning and testing. Test results are shown in Table 1.

TABLE I. EXPERIMENTAL RESULTS

Methods	Quantitative methods based on the maximum value	Extension methods based on the minimum and maximum	Quantitative methods based on mean
Accuracy	0.87	0.81	0.93
Recall	0.89	0.87	0.95

In addition, the professional independence of the number of features aimed to test the effectiveness and consistency of the approach. We built an experimental test. Experimental data selections are training three electricity providers' professional class scholarship assessment results, and then use the computer application of professional two classes of students for evaluation. And the use of quantitative methods Mean standardized scores, the result is: accuracy rate of 0.92, 0.95 recalls. Where you can see the effect of the method to different professions quite small.

V. SUMMARY

This paper suggests scholarship automatic evaluation methods based on SVM. The classification model obtained by learning past scholarship evaluation result data which can be used to automatically assess student scholarships and eliminate the influence of human factors. Besides, to eliminate different teacher assessment results and the impact of the degree of difficulty of the exam, this automatic evaluation method will standardize student achievement; to eliminate the variation of professional training programs, the method proposed to classify scores, thereby performing the number of features of unification. Experimental results show that this method has a certain effect.

It should be emphasized that some problems still existing in applying this method. This method will assess scholarship as a classification problem, and that is no commentary or on the commentary. Therefore, the results are automatically evaluated on a neutral or no commentary on the results, did not like the existing scholarship evaluation method will give all students sort the results. In addition, the method is student achievement with previous data for comparison to determine whether the commentary on scholarships, the number of scholarships it cannot be determined in advance.

ACKNOWLEDGMENT

Zhejiang Province philosophy social science planning project "Vocational Training College Students' Ideal: Path online gaming community 'virtual character' formation, Model" (16NDJC245YB)

REFERENCES

[1] Chen Hailing. Scholarships based assessment model based on fuzzy mathematical methods, mathematical theory and Applications[J], Vol.29, No.3, PP.98-102 Sep 2009

[2] Ma Baoyun. Fellowships under the Credit Assessment Algorithm[D], Beijing Jiaotong University, 2011

[3] Liao Yiqiang. Scholarship Evaluation A quantitative evaluation model[J], Zhengzhou University of Light Industry, Vol. 23, No.2, PP.120-124, 2008.