Journal of Theoretical and Applied Information Technology
10th February 2013. Vol. 48 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

VIRTUAL LABORATORY TEACHING QUALITY EVALUATION MODEL BASED ON ROUGH SET AND SUPPORT VECTOR MACHINE

[1]ZUFENG ZHONG
[1] School of Business, ZhanJiang Normal University, Zhanjiang, 524048, China

ABSTRACT

Virtual laboratory teaching quality evaluation helps to realize scientific teaching management. Virtual laboratory teaching quality evaluation is multi-level and multi-objective system engineering. In this paper, a teaching quality evaluation model based on rough set (RS) and on improved Binary-Tree and multi-category support vector machine (SVM) was provided. Firstly, the attribute reduction of RS was applied as a preprocessor to delete redundant attributes and conflicting objects without losing efficient information. Then an improved multi-category SVM based on Binary-Tree classification model was built to make a forecast. Finally, the model was applied to validation. Results show that this model performs well both in data classification accuracy and predictive accuracy with wide applicability.

Keywords: *Virtual Laboratory; Quality Appraisal System; Rough Set; Support Vector Machine;*

1. INTRODUCTION

Practical ability training has been enhanced for college students. The key for such training is laboratory teaching. Virtual laboratory teaching is a critical stage in training of high-quality and innovative talents, and plays an important role in higher education. In virtual laboratory teaching quality evaluation, the evaluations from the department, guiders, experts and students were first collected, and then the values of each evaluation index were analyzed by traditional methods [1]. These methods are inadequate because the inputs and outputs of the evaluation system are not in simple linearity. Therefore, it is urgent to set up a reasonable and science-based model, and to scientifically, fairly and objectively evaluate teaching quality.

Statistical learning theory (SLT) is the optimal for small-sample statistical estimate and prediction learning. Support vector machine (SVM) based on this theory has a simple mathematical expression and good generalization ability, and can well solve small-sample, nonlinear, and high-dimensional problems with local minimum [2]. Based on the SLT of VC Dimension and structural risk minimization, it seeks in finite sample a compromise between model complexity and learning ability, and aims to achieve the best promotion performance. Rough set (RS) is a math tool that solves imperfect and uncertain knowledge, and conducts knowledge reduction without changing the classification ability of decision table. The combination of RS for preprocessing and SVM for building a classifier has been widely used, because it can provide excellent learning performance and unique advantage in small-sample identification.

In this paper, a model based on RS and SVM is proposed to improve accuracy in laboratory teaching quality evaluation. RS algorithm was used for discretization and reduction, and binary-tree SVM was used for building a classifier and training. The experiments with this model reveal higher accuracy in college laboratory teaching quality evaluation.

2. INTRODUCTION INTO THE THEORIES

2.1 Rough Set

As a mathematic tool that depicts imperfection and uncertainty, RS can effectively analyze and process all types of imperfect, inaccurate, inconsistent and incomplete information, and thereby identify the hidden knowledge and reveal the potential rules [3].

In RS, an information system SI can be expressed by a quadruple:

$$IS = (U, A, V, f)$$

where U is a non-void finite set of targets (or events), called discourse domain; A is a non-void

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

finite attribute set; V is the value range of attribute a; F is an information function.

Suppose [4][5] $X \subseteq U$, R is an equivalence relation of U, and $A = (U, R)$ is an approximation space, then [4]

$$R_*(X) = \cup\{Y \in U/R : Y \subseteq X\}$$
$$R^*(X) = \cup\{Y \in U/R : Y \cap X \neq \Phi\}$$

They are called the R lower and R upper approximations of X, respectively.

Not all attributes are needed in an information system, so the attributes should be reduced.

Suppose P and S are of equivalence relations in U, and the P positive domain of S, marked as $POS_p(S)$:

$$POS_p(S) = \cup P_*(X) \qquad X \in U/S$$

Suppose Q is the S independent subset of P, and $POS_Q(S) = POS_P(S)$, then the subset Q is called the S reduction of P. The attributes in P that cannot be reduced from S are called the S core of P, marked as $CORE(R) = \cap RED(R)$, where RED(R) is the reduction set of R.

2.2 Support Vector Machine

SVM is a machine learning tool based on SLT of VC Dimension and structural risk minimization. It seeks in finite sample a compromise between model complexity and learning ability, and is the achievement of SLT [6]. It is defined as:

Suppose the training sample set under linearly separable condition is $D = \{(x_i, y_i), i = 1,2\ldots,m\}$, $x_i \in R^m$, $y_i \in \{+1,-1\}$. In dimensional space, if there is a classification hyperplane [7][8] $g(x) = w \cdot x + b = 0$,

and it satisfies the following conditions: $\min \frac{1}{2}\|w^2\|,\ s.t.\ \ y_i[w \cdot x + b] \geq 1, (i = 1,2\cdots,m)$

This sample can be classified into two categories with the largest inter-category distance, and then this is the optimal hyper-plane.

To get the optimal solution, we should introduce a Lagrange function, whose saddle (minimum) is the optimal solution.

$$F(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} a_i(y_i((w \cdot x_i) + b) - 1)$$

Where a_i is Lagrange multiplier. Calculate the partial derivatives of w, b and let them be 0, and then solve the decision function:

$$f(x) = \operatorname{sgn}\left\{\sum_{i=1}^{m} a_i^* y_i (x_i \cdot x) + b^*\right\}$$

where is b^* value range of classification; the sample when a_i is not 0 is the support vector.

3. SET UP AN INDEX SYSTEM

3.1. Principles for Setting up Indices for Virtual Experimental Teaching Evaluation

In evaluation of the feasibility and scientificity of virtual laboratory education, the introductory, principled and abstract information should be concretized. Indices are the decomposed attributes related to the object being evaluated, and are concrete, measurable, behavior-based and operable. The index system should be:

1) Objective-based. The system should reflect the requirements for talent training raised by colleges and universities, and reveal the objectives and quality standards of virtual experimental teaching.

2) Comprehensive. Virtual laboratory teaching evaluation involves multiple areas and factors, and is related to such experiment resources as open-sharing rate of lab, number of misreported faults, and number of machines per person. Manifold, multilevel and multi-angle considerations are needed in virtual laboratory teaching evaluation for ensuring objectivity.

3) Dynamic. The system should meet the requirements by virtual laboratory teaching evaluation and the teaching objectives. A reasonable evaluation index should be dynamic. The teachers' teaching level, the students' comprehensive quality, and the resources and conditions in the laboratory are changing. Therefore, virtual laboratory teaching evaluation is staged, varying and developing. Teaching evaluation should be science-based and objective.

4) Feasible. The indices should be well arranged, with an appropriate number and a simple quantizing method. It is expected that during evaluation, a practical and quantifiable method can be used, and

Journal of Theoretical and Applied Information Technology
10th February 2013. Vol. 48 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the specified contents can be objectively collected, analyzed and summarized.

3.2. Design of Building Virtual Laboratory Teaching Evaluation Indices

Based on the above principles for building an index system, and on influence factor analysis, a virtual laboratory teaching quality evaluation model can be built (Table 1). The first-level indices include laboratory resources, students' experiment condition, and teachers' guidance. The second level and third level include 14 indices like open-sharing rate. This index system can be used for college virtual laboratory teaching quality evaluation.

Table 1. Virtual Laboratory Teaching Evaluation Indices

Lab resources	Students' experiment status	Teachers' guidance
Open-sharing rate	Completion rate of theory support courses	Titles
Number of misreported faults	Ratio of guide-time/practice-time	Length of service
Number of machines per person	Absence rate	Pre-class preparation
	Completion rate of experiment	Number of in-class answer
	Completion rate of lab reports	
	Pass rate	
	Excellence rate	

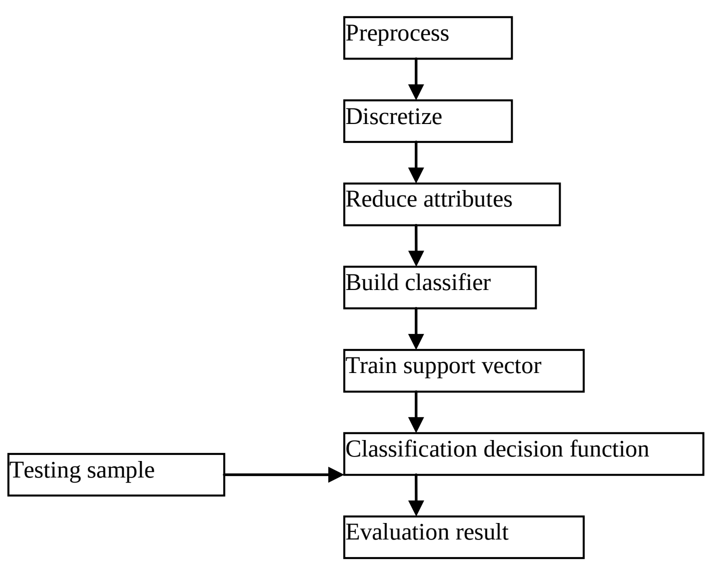

Figure 1 Virtual Laboratory Teaching Quality Evaluation Model

Journal of Theoretical and Applied Information Technology
10th February 2013. Vol. 48 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. EVALUATION MODEL

Figure 1 shows the modeling for the mixed classification algorithm based on RS and SVM. The modeling process includes three stages: 1, collect data for the indices; preprocess and discretize the data; reduce attributes and dimensions by RS. 2. Build a classifier with SVM; train SVM and build a classification decision function. 3. Test with the testing data, and get the results.

4.1 Preprocessing and Discretization

The statistical data from a college laboratory training center during the past few years were used in this paper. First, in preprocessing, the "dirty" data were cleaned, such as deleting the abnormal data, and supplementing the incomplete data. Then, since RS cannot be directly used in processing continuous data, the 14 scattered or continuous indices in this paper should be discretized prior to attribute reduction. For instance, the teachers' titles of professor, associate professor, lecturer and assistant can be discretized into 4, 3, 2 and 1, respectively. The teachers' service lengths of below 5, 5-10, 10-20, and above 20 years can be discretized into 1, 2, 3 and 4, respectively.

After preprocessing and discretization, the resulting normal data set of 14 indices are suitable for attribute reduction.

4.2 Attribute Reduction

In attribute reduction, the uncorrelated or unimportant knowledge is deleted without changing the classification or decision ability of knowledge base. Consequently, the attributes and decision rules can be reduced, and data utilization can be improved.

The steps for attribute reduction include: 1) Reduce the 14 indices (condition attributes), like delete the redundant attributes (dispensable attributes), while maintaining consistency with the original decision algorithm. 2) Reduce the attribute values, that is, reduce each decision rule; after the redundant attribute values are deleted, the consistency of the original decision is kept. 3) The minimum decision rule set is the required minimum algorithm.

After attribute reduction, 6 indices are obtained (number of machines per person, pass rate, completion rate of theory support courses, completion rate of experiment, absence rate, teachers' titles). The reduction attributes and the corresponding data set are used to build an SVM classifier.

4.3 Building a Classifier for Testing

The core of this laboratory evaluation model is the construction of classifier, which directly affects the evaluation results. In this paper, 800 data sets were acquired, with 400 training sets and 400 testing sets. Since SVM applies to two categories, a multi-category classifier should be built: an improved binary-tree method was used [9]. Virtual laboratory teaching quality was classified into Excellent, Good, Pass and Flunk. The steps are as follow: 1) based on the classification requirements, to set up the training sample set $\{x_i, y_i\}, y_i = \{1,2,\cdots,n\}, n$ is number of categories. The attributes are adjusted linearly to $[-1,+1]$. 2) Calculate the core $\bar{x}^j$ for each category sample set, $\bar{x}^j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i^j = \{1,2,\cdots K\}, n_j$ is the number of samples for category j. 3) Calculate the minimum hypersphere radius $\bar{r}^j{}_{\max}$ for each category. 4) Calculate the distance $r_i^j = \| x_i^j - \bar{x}^j \|$ of the first category from its core $\bar{x}^j$, x_i^j is sample I of category j. Sort the samples by distance. 5) Find the distance of 80% of samples as the hypersphere core radius of this category; based on $R_{dec} = R_{\ker} + cof \times (R_{\max} - R_{\ker})$, calculate the hypersphere decision radius, where $R_{\ker}$ is the hypersphere core radius, $R_{\max}$ is the minimum hypersphere radius, R_{dec} is the hypersphere decision radius, 0<cof<1 is weight coefficient. 6) Based on the hyperspace decision radius of each category, build the structure of the binary tree; larger decision radius is first given leaf node, and first separated from other categories. 7) Build a bi-category SVM. In training sample sets, define the categories to be separated as positive sample, and other categories as negative samples. Select suitable core functions and core parameters for training SVM. Build the optimal hyperplane and save the trained structural model. 8) When only two categories of training sets are left, training finishes; otherwise, delete the categories separated from the previous step, and return to Step 7 for bi-category training.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Build a classifier with the 400 training sets, and train the classifier with the 400 testing sets. The results are listed in Table 2. After classification and training, the accuracy is >85%; after testing, the accuracy is >80%. The error rate of each category is about 7%, which is satisfactory.

Table 2. Accuracies Of Training Sets And Testing Sets

	Accuracy	Error rate			
		Category 1	Category 2	Category 3	Category 4
Training set	86.20%	4.32%	5.06%	7.63%	6.34%
Testing set	84.32%	5.73%	6.20%	8.45%	6.89%

5. CONCLUSIONS

Rough set, support vector machine (SVM), and their combination are all promising issues. The upcoming work includes: First, rough set only processes discrete data, so in terms of discretization, the selection of breaking points will greatly influence the reduction results. Therefore, it should be further studied how to select more effective breaking points by using optimization algorithms. Second, much progress has been made in basic SVM training algorithms and the efficiency of the improved binary-tree SVM algorithms has been increased, but the SVM theories which emerged later should be still further perfected.

The virtual laboratory teaching quality evaluation model is a guarantee for higher teaching quality. Teaching evaluation is a multifactor, multivariate, blurred and nonlinear process. RS can be used to solve problems of large sample and uncertainty in data processing. Binary-tree SVM was used for building a classifier, which has good nonlinear classification ability. This classifier was used to evaluate virtual experimental teaching quality, and has prospects in improving learning speed and evaluation accuracy.

ACKNOWLEDGEMENTS

This work was supported by Natural Science Foundation of ZhanJiang Normal University (W1107).

REFERENCES

[1] Zhu Xufen and Jia Xiaoming, Mobilize all Positive Factors of Student, Deepen Teaching Reform of Microbiology [J], Microbiology, 2007, 34(1): 185~187 (in Chinese).

[2] Zhang Xuegong, Statistical Learning Theory and Support Vector Machine [J], Acta Automatica Sinica, 2000, 26(1): 32-42 (in Chinese)

[3] Yang Wu, Huang Zhizhen and Li Lixin, An Inference Quantized Analysis Scheme of MLS DBMS Using Rough Set [J], Computer Science, 2005, 32(8): 75-76 (in Chinese).

[4] Zeng Huanglin and Yao Yi, Rough Set Theory and Application [J], Computer Science, 2001, 28(5) (in Chinese).

[5] Zhang Kun and Chi Guotai, Establishment of ecological evaluation indicators system based on correlation analysis-rough theory. Journal of systems engineering [J], 2012, (2): 119-125. (in Chinese)

[6] T.Van Gestel, B.Baesens, J.Garcia, P.Van Dijcke. A support vector machines approach to credit scoring. Bank en Financiewezen, 2003, (2):73-82.

[7] Deng Naiyang and Tian Yingjie, New Method in Data Mining: Support Vector Machine [M], Science Press: Beijing, 2004 (in Chinese)

[8] Liu Jianwei and Li Shuangcheng, Classification Algorithm of Support Vector Machine via p-norm Regularization. ACTA Automatica Sinica[J], 2012, (1): 76-85. (in Chinese)

[9] Vapnik V N, translated by Zhang Xuegong, Improved Multi-Category Support Vector Machines Based on Binary Tree [J], 2010, 46(33): 117-120.