

On Evaluating the Balanced Development of English Education in Compulsory Schools Based on Proximal Support Vector Machine

Suping Sun^{1, a}, Yanwei Zhu^{2, b}

¹Foreign Languages Department, Tangshan Teachers' College, Tangshan, Hebei, China 063000

²Department of Mathematics and Informatics, Tangshan Teachers' College, Tangshan, Hebei, China 063000

^a2948109568@qq.com, ^bzhyw79@163.com

Keywords: compulsory schools; English education; balanced development; proximal support vector machine

Abstract. For a long time, researchers have achieved much in balanced development of national and regional compulsory schools, yet little has been done in the study of balanced development in a certain subject area. This paper aims at improving Proximal Support Vector Machine and constructing Weighted Proximal Support Vector Machine in order to scientifically evaluate the balanced development of English Education in compulsory schools. This article goes further to analyze the evaluating factors of the balanced development of English education in compulsory schools and the feasibility to apply PSVM to do so by determining the main evaluating indicators and working out a new evaluation method based on PSVM in the hope of boosting the study of balanced development of other courses therefore to achieve the ultimate goal of balanced development among compulsory schools.

1. Introduction

In the document, On the Promotion of Principal Teacher Exchange Rotation Among Compulsory Schools Within the Same County (District), Ministry of Education, Ministry of Finance, Ministry of Human Resources and Social Security point out: teacher allocation is the key element in furthering the balanced development of compulsory education. As far as allocation of teacher resources is concerned, if the school broadly considered as a whole, it is bound to have interdisciplinary faculty imbalance, therefore will in turn hinder the balanced development among schools. Thus, to achieve the balanced faculty allocation between rural and urban schools and between schools in the same district, we must consider the balance in a certain subject area because only when we have balanced faculty allocation for every discipline, shall we reach the goal of balanced development among all schools.

How to scientifically and effectively evaluate the balanced development of a discipline in compulsory schools, e.g. English education in a certain area is a key issue from both microscopic and macroscopic view therefore is only way to achieve ultimate balanced development among compulsory education. However for a long time, researchers have achieved much in balanced development of national and regional compulsory schools, yet little has been done in the study of balanced development about a certain subject. Especially English education has been much criticized for its "time-consuming and low-effectiveness". Therefore we take efforts to employ data mining method---support vector machine to evaluate English education in a certain area: to specify the decisive factors, to select evaluated elements and to determine the optimal support vector machine model. On the one hand, we will provide quantitative descriptions about the qualified and unqualified factors. On the other hand, we will supply a scientific method to adjust the unqualified elements. We aim to build up both a static and dynamic decision-making system to continuously boost the balanced development of English education in compulsory schools in the same district, therefore to achieve the final goal of balancing the development of every compulsory school as a whole.

2. Basic knowledge and algorithm about Support Vector Machine (SVM)

Support Vector Machine, first proposed by Cortes and Vapnik in 1995, based on Statistical Learning Theory, is a new tool to solve machine learning problems by means of optimization method. In recent years, it has been considered as one of the major achievements in the field of machine learning research for overcoming "dimensionality disaster" to some extent and traditional difficulties namely "over-learning". It is a research on data mining method, in another word, a research on how to carry on machine learning through empirical data. It is categorized as a cross-discipline and combination of computer science, pattern recognition, optimization and application of statistical methods. Currently in the United States and other developed countries, it has been successfully applied in military, economic and other fields, mainly to solve the pattern classification, regression analysis, function estimation problems. Researchers find the law from the observed data (sample), and then employ these laws to make decisions for future data or unobserved data. Specifically, SVM is known as an algorithm to build and solve optimization model from the training set (sample set), to construct decision-making functions, the purpose of which is to classify the actual problems or regression decision with the decision-making functions.

3. Feasibility Analysis of Applying SVM to the evaluation of the balanced development of English education in compulsory schools

We chose English teaching of the same leveled compulsory schools (e.g. rural junior high school) in a certain area for the evaluation object. After analyzing English teaching quality indicators, taking into consideration evaluating English teaching quality indicators from Educational Advisory Committees for different government levels, we determine the major English teaching quality indicators including English teachers' educational degrees, denoted by a_1 , English teachers' title structure denoted a_2 , English teachers' average age denoted a_3 , students evaluation on English teaching denoted a_4 , reform project per teacher denoted a_5 , number of English books per student denoted by a_6 , English data audio copies per student denoted by a_7 , student-teacher ratio denoted a_8 , students learning interest denoted by a_9 . Thus each school corresponds to a 9-dimensional vector ($a_1, a_2, a_3 \dots a_9$). Based on the size of the evaluation range, randomly selecting a certain number of schools, e.g. 100, we can obtain 100 9-dimensional vectors $x_1, x_2 \dots x_{100}$. Again according to the experts' opinions from the government supervision, the qualified ones are recorded 1 and the unqualified as -1, we can work out a training set (or sample set) $T = \{(x_1, y_1), \dots, (x_{100}, y_{100})\}$, where $x_i (i=1, 2, \dots, 100)$ is a 9-dimensional vector, $y_i (i=1, 2, \dots, 100)$ is ± 1 . With targeted training set T we can build an optimization model and by solving an optimization model we can construct decision function $f(x)$. With this decision function $f(x)$, we can evaluate whether English teaching in any schools of a certain region is qualified, that is, the 9 calculated values of English teaching in any school to make 9-dimensional vector x , substituting $f(x)$, if $f(x) = 1$, the English teaching in the school is qualified; If $f(x) = -1$, unqualified. The SVM algorithm will be compiled as a computer system software, for $f(x) = -1$ of x , x component can be adjusted on the computer, and by recording the adjustment component values, when $f(x) = 1$, the adjustment is completed. Thus it can clearly inform us about the would-be adjusted items for those unqualified schools and at the same time figure out the minimum adjustment values for the unqualified items. Thus we can acquire a quantitative target and methods of rectification for compulsory schools to avoid the blindness in terms of constructing and developing English Education. Meanwhile we can employ SVM decision function $f(x)$ to evaluate English education and even other subject education in all the schools within the region to save comprehensive assessments organized by large-scale expert supervisions. What is more, with the development of technology, economy and educational level, when supervision experts redefine the criteria, that is, to raise the standard, we can then re-build SVM algorithm to conduct a comprehensive evaluation, to provide data for schools re-rectification of English education. This will allows us to constantly improve the level of balanced development of English education and other course teaching.

4. Weighted Center Support Vector Machine

Based on the above analysis, the qualified and unqualified English education of different schools correspond with positive and negative points from the SVM training set, yet this relations are sometimes uneven because the components of the training set (sample set) carry differ from one another in the degree of importance. To solve these problems, we construct a weighted center support vector machine modeling on building weighted SVM method.

By introducing parameter C_+ to positive class points, parameter C_- to negative class points, we can obtain the original optimization problem:

$$\min_{w, \eta, b} \quad \frac{1}{2} (\|w\|^2 + b^2) + \frac{C_+}{2} \sum_{y_i=1} \eta_i^2 + \frac{C_-}{2} \sum_{y_i=-1} \eta_i^2 \quad (1)$$

$$s.t. \quad y_i ((w \cdot x_i) + b) = 1 - \eta_i, i = 1, 2, \dots, l \quad (2)$$

Or, more popularly, if we know in advance the importance of each point, we can introduce different penalty parameter C_i for each training point resulting in the original optimization problem:

$$\min_{w, \eta, b} \quad \frac{1}{2} (\|w\|^2 + b^2) + \frac{1}{2} \sum_{i=1}^l C_i \eta_i^2 \quad (3)$$

$$s.t. \quad y_i ((w \cdot x_i) + b) = 1 - \eta_i, i = 1, 2, \dots, l \quad (4)$$

Theorem 4.1, the dual problem of (3) --- (4) equals

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j ((x_i \cdot x_j) + 1) + \frac{1}{2} \sum_{i=1}^l \frac{\alpha_i^2}{C_i} - \sum_{i=1}^l \alpha_i \quad (5)$$

Proof the Lagrange function of question (3) --- (4) is

$$L(w, b, \eta, \alpha) = \frac{1}{2} (\|w\|^2 + b^2) + \frac{1}{2} \sum_{i=1}^l C_i \eta_i^2 - \sum_{i=1}^l \alpha_i (y_i ((w \cdot x_i) + b) + \eta_i - 1) \quad (6)$$

where $\alpha \in R^l$ is Lagrange multiplier vector, seeking the minimum of Lagrange function with respect to w, b, η , to gain the following conditions:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (7)$$

$$b = \sum_{i=1}^l y_i \alpha_i \quad (8)$$

$$\eta_i = \frac{\alpha_i}{C_i}, i = 1, \dots, l \quad (9)$$

$$y_i ((w \cdot x_i) + b) + \eta_i - 1 = 0, i = 1, \dots, l \quad (10)$$

Substituting (7) --- (9) into Lagrange function and seeking maximum α to achieve unconstrained dual problem (5), which is strictly convex quadratic programming. Its optimal solution is

$$\alpha = (I_C + Y(XX^T + ee^T)Y)^{-1}e = (I_C + HH^T)^{-1}e \quad (11)$$

Where $X \in R^{l \times n}$ is a composed matrix by inputting $x_i \in R^n, i = 1, \dots, l$; $Y \in R^{l \times l}$ is a matrix composed respectively by y_1, \dots, y_l and the remaining elements zero, $e = (1, \dots, 1)^l$, the diagonal against

$$I_C = \text{diag}\left\{\frac{1}{C_1}, \dots, \frac{1}{C_l}\right\} \quad (12)$$

$$H = Y[X, e] \quad (13)$$

By introducing the kernel function $K(x_i, x_j)$ to substitute the inner product $(x_i \cdot x_j)$ in problem (5), to obtain weighted proximal support vector machine. Specific algorithm is as follows:

Algorithm 4.1 Weighted PSVMC

1) Suppose the training set is known as $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$, where $x_i \in X = R^n, y_i \in Y = \{-1, 1\}, i = 1, \dots, l$;

2) Select the appropriate parameters $C_i, i=1, \dots, l$; select the appropriate kernel function $K(x, x')$;

3) Construct and solve optimization problems

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (K(x_i, x_j) + 1) + \frac{1}{2} \sum_{i=1}^l \frac{\alpha_i^2}{C_i} - \sum_{i=1}^l \alpha_i \quad (14)$$

Get the optimal solution α^* ;

4) According to equation (8) to determine b, constitute classification decision

$$\text{function } f(x) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i k(x, x_i) + b \right) .$$

5. Application of Weighted Proximal Support Vector Machine to bBalanced Development of English Education in Compulsory Schools

We randomly chose 100 rural junior high schools as samples in a municipal range and calculated each of 9 indicators of English education in each school. Giving a value to each indicator, we get a training set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_{100}, y_{100})\}$, where $x_i = (a_{i1}, a_{i2}, \dots, a_{i13})$, $i=1, 2, \dots, 100$. $y = \pm 1, i=1, 2, \dots, 100$. Thereby we make a set of practical training decisions, develop a decision-making systems based on weighted PSVM. Specifically we do the following:

A. Data preprocessing

Observing the existing data we can find that some data values are small, some large, and therefore we need to standardize data, which the minimum - maximum standardized method. The formula is as follows:

$$[x_j]_i' = ([x_j]_i - \min_{j=1, \dots, 100} ([x_j]_i)) / (\max_{j=1, \dots, 100} ([x_j]_i) - \min_{j=1, \dots, 100} ([x_j]_i)). \quad (15)$$

In this way the data can be normalized as D' .

Then we randomly divide the data set D' into two parts by the ratio of 7: 3: one part as a training set T , the number of training points referring to l (here $l=70$), while the other part a test set S , the number of training points referring to m (here $m=30$). Suppose the positive class points in the training set as $T+$, the negative number for the class point as $T-$; the positive class points in the test set as $S+$, the negative class points in the test set as $S-$. By comparing data, we find that the negative class points(not meeting the requirements) are 26, and positive class points(meeting the requirements) are 44, which displays much unevenness. Therefore we need to give different types of penalty parameters $C+$ and $C-$ in the application of the above model and $C+$, $C-$ are defined as the following formula:

$$C_+ = C \times \frac{T_-}{l}, \quad C_- = C \times \frac{T_+}{l} \quad (16)$$

where $C > 0$ is previously given parameter.

B. Model Selection

In response to these classification problems, we need to select the appropriate algorithm model and three SVM models were selected with the first one being weighted PSVM model given earlier, the original problem to be solved

$$\min_{\omega, \eta, b} \frac{1}{2} (\|\omega\|^2 + b^2) + \frac{C_+}{2} \sum_{y_i=1} \eta_i^2 + \frac{C_-}{2} \sum_{y_i=-1} \eta_i^2, \quad (17)$$

$$\text{s.t.} \quad y_i ((\omega \cdot x_i) + b) = 1 - \eta_i, i=1, 2, \dots, l. \quad (18)$$

Its dual problem

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (K(x_i, x_j) + 1) + \frac{1}{2C_+} \sum_{y_i=1} \alpha_i^2 + \frac{1}{2C_-} \sum_{y_i=-1} \alpha_i^2 - \sum_{i=1}^l \alpha_i \quad (19)$$

The second is weighted reasoning support vector machine model, and the third is weighted standard support vector machine model. After determining the three models, we need to select one

of the parameters, including the kernel function $K(x, x')$ and C, C^* , and kernel function parameters. Here we select the kernel function as radial basis function

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right). \quad (20)$$

Such parameters needed to be selected are C, C^* and σ .

For each model, according to the grid method to select the optimal parameters, namely, first the range for C and $C^* \{0.1, 1, 10, 100, 1000, 10000\}$ and σ ranging $\{0.1, 0.2, 0.5, 1, 2, 5\}$. This parameter group consisting of (C, C^*, σ) . For each set of parameter values, we chose minimum LOO error parameter set as the optimal parameters $(\bar{C}, \bar{C}^*, \bar{\sigma})$. The results are as follows: the weighted proximal SVM corresponds to parameter set $(\bar{C} = 10, \bar{\sigma} = 1.5)$; weighted reasoning SVM corresponds to parameter set $(\bar{C} = 100, \bar{C}^* = 100, \bar{\sigma} = 2)$ and weighted standard SVM to $(\bar{C} = 10, \bar{\sigma} = 5)$.

C. Results

We substitute the above-obtained optimal parameters into three optimal models to obtain a final decision function which in turn detect the points in the test set S . The results are shown in the following table:

Table 1 comparing the results

Test results	C-SVM	TSVM	PSVM
Detection accuracy	85%	88%	93%
False rate	3%	0%	1%
Detection rate	68.5%	68.7%	86%

Here in the table, detection accuracy is the ratio between the number of samples correctly detected and total number of samples; false rate is the ratio between false positives of normal samples as abnormal and the number of normal samples; detection rate is the ratio between the number of abnormal sample correctly detected and the number of actual abnormal samples.

Considering the result, we can see that the above-discussed weighted PSVM displays the highest detection accuracy to judge whether English education in a certain school is qualified. Therefore when we evaluate English education in a certain area, we just need to update the information into the decision system. If 1 is the result, then the English education in the school is qualified, if -1 is the result, we then need to adjust some data in the system until we reach the qualified goal of 1. We can gain the optimal resource allocation decision.

6. In conclusion

Based on these research and demonstration, it is definitely feasible to apply support vector machine to making decisions on the balanced development of English education in compulsory schools. A variety of support vector machines can improve conciseness of decision-making. It is only an issue of choosing a support vector machine model to work out the highest accuracy. As seen from the above examples, as long as the problem involving classification is concerned, support vector machine can be built to aid decision-making schemes. Conclusion can be made then that decision optimal system built based on the employment of support vector machine model will be of great help to evaluate English education in the hope of boosting the study of balanced development of other courses therefore to achieve the ultimate goal of balanced development among compulsory schools.

References

- [1] Fung G, Mangasarian O.L Proximal Support Vector Machine Classifiers·KDD2001 San Francisco CA USA, 2001.

- [2] Deng, Naiyang and Tian, Yingjie, Support Vector Machine Theory Improving Algorithm and its Expansion, Beijing: Science Publishing House. 2009.
- [3] Yu, Dong and Tang, Lin, “Two Evaluating Dimensions About Balancing Development of the Compulsory Schooling” Statistics and Decision, vol 13, pp122-173, 2011.
- [4] Wu, Jichun, “Foreign Index Managing High-quality Education: Advantageous Balance Behind Basic Education” Teaching Monthly, vol 5, pp34-36, 2007.
- [5] Wang, Ximai, “Analytic Frame and Evaluating Indicator of Educational Equality” Journal of Beijing Normal University, vol 3, pp21-23, 2008.
- [6] Shen, Youlu, “Educational Equality Dimensions” Educational Science Research, vol 7, pp34-38, 2009.
- [7] Zhai, Bo, “Educational Balancing Development: Theory and Target Methods and Measuring”. Educational research, Vol 3, pp17-28, 2006