http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 10 (11): 2140-2146, 2011
ISSN 1812-5638 / DOI: 10.3923/itj.2011.2140.2146

Teaching Assistant Evaluation Based on Support Vector Machines with Parameters Optimization

[1]Jing Yang, [2]Hua Jiang and [2]Honglei Zhang
[1]School of Kexin, Hebei University of Engineering, Handan 056038, China
[2]School of Economics and Management, Hebei University of Engineering, Handan 056038, China

Abstract: The quality of teaching assistants' work is important to students' education and inclusion, so it is of significance to evaluate and improve the performance of teaching assistants. Support vector machines with appropriate parameters may provide good tools for enhancing the recognition accuracy. Some basic knowledge on support vector machines was firstly introduced; then the paper applied the teaching assistant evaluation data set to examine the recognition effects of SVMs with default and chosen parameters, showing that different parameters may produce different evaluation results. Cross validation method and particle swarm optimization were respectively applied to optimize the parameters of support vector machines, both of which enhanced the recognition accuracy. Finally, conclusions and recommendations were given.

Key words: Teaching assistant evaluation, support vector machines, parameters optimization, cross validation, particle swarm optimization

INTRODUCTION

Teaching Assistants (TAs) are individuals who assist professors or teachers with instructional responsibilities. TAs include Graduate Teaching Assistants (GTAs), Undergraduate Teaching Assistants (UTAs), secondary school TAs and elementary school TAs. These TAs are typically responsible for most day-to-day, routine interaction with students, for ensuring students understand lectures and other course materials and for assessing student work (Osterlund and Robson, 2009). Teaching assistants also take an important role to create a good learning environment for students, especially for the first-year students (Samson and Millet, 2003). More and more studies have shown that the quality of their work is key to students' education and inclusion (Cobb, 2005; Gorsuch, 2006; Cremin *et al*., 2003). TAs' assignments are often used to pevaluate their performance. It may be effective to do the work by hand when the number of the assignments is small. However, when there are a large number of assignments, people have to turn to more effective and accurate methods.

Artificial Intelligence (AI) techniques provide good tools for enhancing the accuracy of performance evaluation. Support Vector Machines (SVMs), invented by Abibullaev *et al*. (2010), is a novel machine learning method based on Statistical Learning Theory (SLT) which is powerful for the problem with small sampling, nonlinear and high dimension. SVMs can be applied to analyze data and recognize patterns, used for classification and regression analysis. Owing to the advantages of SVMs, the technology has been used in various fields. However, an important problem encountered in setting up SVM models is how to select their parameters because inappropriate parameter settings may lead to poor classification results (Keerthi and Lin, 2003). Few literatures have focused on the parameters optimization of SVMs for teaching assistant evaluation. Present study tries to do some works for the issue.

TEACHING ASSISTANT EVALUATION BASED ON SVMS

Preview of support vector machines: The Support Vector Machine (SVM) is a new learning machine based on the statistical learning theory which embodies the Structural Risk Minimization (SRM) principle shown superior to the Empirical Risk Minimization (ERM) principle (Abibullaev *et al*., 2010). SRM minimizes an upper bound on the expected risk while ERM minimizes the error on the training dataset which equips SVM with better generalization ability. For the classification problems, the realization process of SVM is that input vectors are firstly mapped into a high-dimensional space through a nonlinear mapping and then a hyperplane is constructed and is moved until a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class. The details on SVM

Corresponding Author: Jing Yang, School of Kexin, Hebei University of Engineering, Handan 056038, China

wouldn't be introduced here. Kernel functions are important to the performance of SVM. Although, various kernel functions have been proposed by researchers, the most widely used ones are the following four kinds: Linear kernel function, Polynomial kernel function, Radial basis function and Sigmoid kernel function.

Description of the teaching assistant evaluation dataset: The Teaching Assistant Evaluation Data Set used in the paper comes from the UCI Machine Learning Repository (Frank and Asuncion, 2010). The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into 3 roughly equal-sized categories ("low", "medium" and "high") to form the class variable. The detailed information and the whole dataset can be accessed from [http://archive.ics.uci.edu /mL/ datasets/ Teaching+Assistant+Evaluation]. Each instance in the dataset owns 6 variables: (1) Whether the teaching assistant is a Native English Speaker (binary, 1 = English speaker, 2 = non-English speaker); (2) Course instructor (categorical, 25 categories); (3) Course (categorical, 26 categories); (4) Summer or regular Semester (binary, 1 = Summer, 2 = Regular); (5) Class size (numerical); (6) Class attribute (categorical, 1 = Low, 2 = Medium, 3 = High).

Teaching assistant evaluation modeling based on SVM: According to the SVM theory, we can conclude the basic process of teaching assistant evaluation modeling based on SVM. First, the training dataset and testing dataset are selected from the original data of teaching assistant evaluation; second, the selected training dataset and testing dataset are preprocessed in order to require the input and output standards of SVM; then, the training dataset is used to train SVM to get the teaching assistant evaluation model; after above processes, the trained SVM model can be applied to evaluate the performance of teaching assistants in the testing dataset.

Considering the limited instances, the paper selects all the instances in the Teaching Assistant Evaluation Data Set as the training samples and testing samples. The variables have different dimensions and magnitudes as well as different changing trends. In order to make these variables meet the input requirements of SVMs, the paper preprocesses them, making them same trending, dimensionless and quantitative. The paper mainly normalizes the dataset before inputting them into SVMs.

The normalization mapping is as follows:

$$f: x \rightarrow y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where, x is the value of some variable, $x_{min} = \min(x)$, $x_{max} = \max(x)$ and y is the normalized value which belongs to [0,1]. In MATLAB software, the mapminmax function can process matrices by normalizing the minimum and maximum values of each row to [0,1]. Mapminmax (X, YMIN, YMAX) takes X and optional parameters and returns Y and PS which are process settings that allow consistent processing of values.

After being preprocessed, the normalized values can be input into the SVM to train the classifier. The SVM toolbox that the paper uses is the Libsvm toolbox developed by [http://www.csie.ntu.edu.tw/ ~cjlin/ libsvm/ index.html]. The Libsvm toolbox is an integrated software tool for support vector classification (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM). It supports multi-class classification. Although, inappropriate parameter settings may lead to poor classification results, there are no effective methods to setup the parameters. However, the Libsvm provides many default parameters which may deal with some problems, so the paper first applies default parameters to train the SVM. The kernel function of SVM applied in the paper is the Radial Basis Function. Input the normalized values of the instances in the training dataset, we can train the SVM model.

By training the SVM, the paper gets the teaching assistant evaluation model which can be used to evaluate the performance of each testing instance. The output result is shown as Fig. 1, compared with actual classes. 81 of 151 instances are recognized right which means the accuracy is 53.6424%.

Results analysis: The results are shown in the Fig. 1 gotten with default parameters. Now we choose some different parameter values to see the recognition results. There are four kernel functions: Linear, Polynomial, Radial basis function and Sigmoid mentioned in cross validation method and the penalty parameter c and the kernel function parameter g also have important effects on the classifier (Note that: the kernel function parameter g represents the γ in polynomial, radial basis function and sigmoid kernel function and other related kernel parameters take default values). Table 1 gives the evaluation results with different parameters. As we can see, different kernel functions may result in different

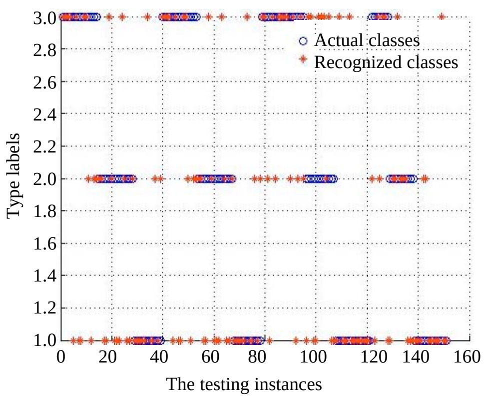

Fig. 1: Actual classes and recognized classes of SVM with default parameters

Table 1: Results with different parameters

The kernel function	Values of c and g	Accuracy
Linear	Default	54.3046% (82/151)
	'c = 0.5'	53.6424% (81/151)
	'c = 5'	58.9404% (89/151)
	'c = 10'	58.9404% (89/151)
Polynomial	Default	55.6291% (84/151)
	'c = 0.5, g = 1'	58.9404% (89/151)
	'c = 5, g = 6'	76.8212% (116/151)
	'c = 10, g = 8'	81.457% (123/151)
Radial basis function	Default	53.6424% (81/151)
	'c = 0.5, g = 1'	56.2914% (85/151)
	'c = 5, g = 6'	76.1589% (115/151)
	'c = 10, g = 8'	79.4702% (120/151)
Sigmoid	Default	51.6556% (78/151)
	'c = 0.5, g = 1'	31.1258% (47/151)
	'c = 5, g = 6'	32.4503% (49/151)
	'c = 10, g = 8'	31.7881% (48/151)

results with the same c and g and different cs and gs may also produce different classification results with the same kernel function.

PARAMETERS OPTIMIZATION OF TEACHING ASSISTANT EVALUATION SVM BASED ON CROSS VALIDATION

As Table 1 shows in Accuracy, SVMs' parameters are crucial for their classification accuracies. Sometimes the classification results SVMs produce with default parameters may be satisfactory but key parameters need to be adjusted to get more satisfactory accuracies. Then, how to choose the best parameters is the issue the following sections try to solve. Here, the study uses the Cross Validation (CV) method to optimize the parameters (mainly the penalty parameter c and the kernel function parameter g) of the RBF-SVM.

Cross validation methods: Cross validation is a statistical analysis method used to verify the performance of

```
Start
% parameters initialization
bestAccuracy = 0;
bestc = 0;
bestg = 0;
% meshing c and g for the grid search
for c = 2^(cmin):2^(cmax)
   for g = 2^(gmin):2^(gmax)
        Dividing equally train into K groups: train(1), train(2), …, train(K);
        Splitting the corresponding labels: train_label(1), train_label(2), …,
        train_label(K)
   for run = 1:K
        Taking train(run) as the validation dataset and others as training
        dataset
        Recording the verification accuracy into acc(run)
end
      cv = (acc(1)+ acc(2)+ …+ acc(K))/K;
      if (cv>bestAccuracy)
         bestAccuracy = cv; bestc = c; bestg = g;
end
      end
   end
   Over
where cmin, cmax, gmin, gmax and K should be given values
```

classifiers. The basic idea is that the original dataset is divided into training datasets which are used for training classifiers and validation datasets for testing the trained models to obtain the classification accuracy as the evaluation performance of classifiers. Several common CV methods include Hold-Out Method, K-fold Cross Validation (K-CV), Leave-One-Out Cross Validatlon (LOO-CV) and so on.

Compared with the K-CV method, LOO-CV has the advantage of using almost all the instances to train SVM models, so it may produce more reliable results. However, the main shortcoming of LOO-CV is the high calculation cost because the number of SVM models it needs to establish equals nearly to the number of instances in the original dataset. It is difficult or impossible to achieve when the original dataset includes considerable instances. Considering the advantages and disadvantages of above three methods, the paper mainly applies the K-CV to optimize the parameters of SVMs.

Algorithm design of optimizing parameters of SVMs based on CV: The whole process of teaching assistant evaluation based on SVMs is similar to parameters optimization but a CV algorithm for choosing the best parameters is embedded. The SVM toolbox used in the session is still the Libsvm.

At present, there are no unified methods recognized by researchers on the parameters optimization of SVMs. The method used commonly is: (1) c and g take values in a certain range; (2) For the set c and g, the training dataset is taken as the original dataset; (3) CV methods are

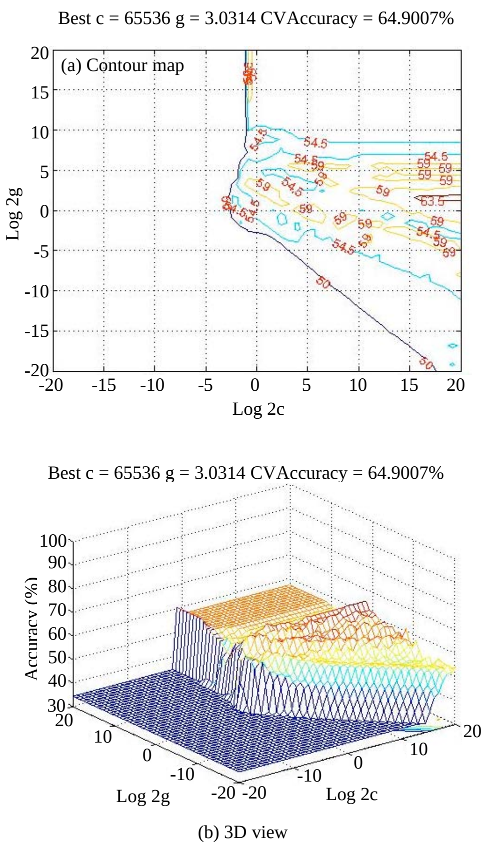

Fig. 2(a-b): Rough results of SVM parameters optimization based on CV

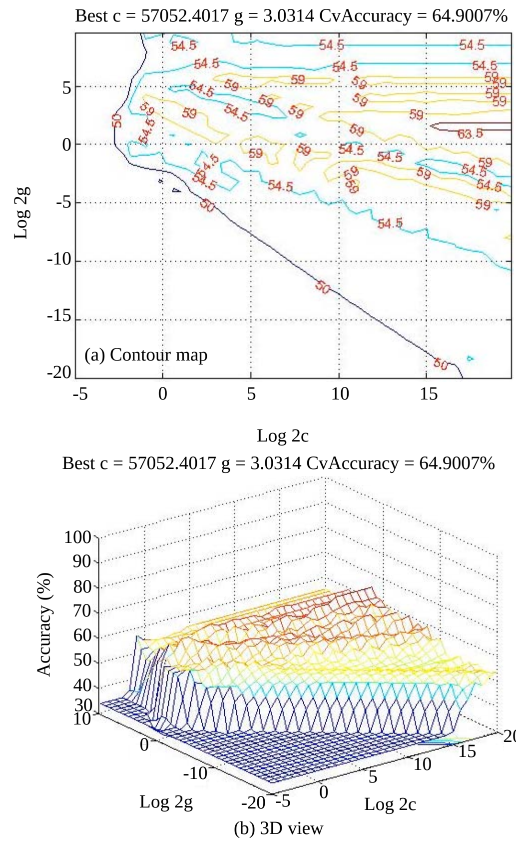

Fig. 3(a-b): Precise results of SVM parameters optimization based on CV

applied to obtain the best values of c and g; (4) Ultimately, the c and g which produce highest validation accuracy are regarded as the optimal parameters.

But there may be a problem that more than one set of c and g correspond to the highest validation classification accuracy. How to deal with this situation. The paper chooses the minimal c and g in the groups of cs and gs which achieve the highest accuracy as the optimal parameters; if more than one set correspond to the minimal c and g, the first group of c and g will be selected. The reason is that too high c may lead to the overlearning. The pseudocode of the above CV algorithm for optimizing c and g is as follows:

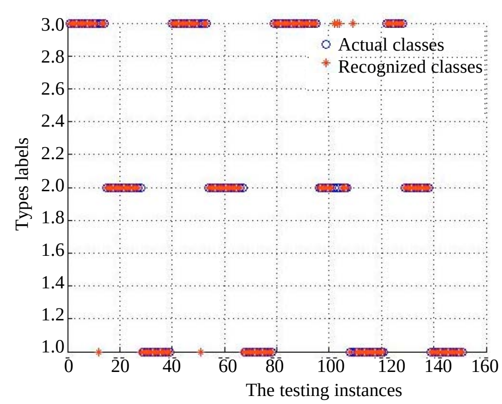

Fig. 4: Actual classes and recognized classes of SVM with optimized parameters by CV method

Results analysis: First, in order to cover the optimal parameters, a rough range is chosen. The ranges of c and g are from 2^(-20) to 2^(20). The selection results are shown as Fig. 2. The best CV accuracy is 64.9007% with the best c 65536 and g 3.03143. It took 417.685981 seconds in all.

According to the contour map and 3D view in Fig. 2, the paper changed the ranges of c into from 2^(-5) to ^(20) and g from 2^(-20) to 2^(10). Fine selection results are shown as Fig. 3 with the elapsed time 352.90543 seconds. Best cross validation accuracy equals 64.9007%, the best c 57052.4 and the best g 3.03143.

After getting the best c and g, the paper uses them to recognize the classes of the instances in the testing dataset. The recognized results are shown as Fig. 4. The teaching assistant evaluation accuracy is 94.702% (143/151) which is higher than any of those in the Table 1.

It can be seen that the CV method can attain the optimal parameters in some sense which may effectively avoid the overlearning and underlearning. A better accuracy can be achieved for the testing dataset. Example results show that the trained SVM models with the selected parameters by CV methods are more effective than the models trained with randomly selected parameters.

PARAMETERS OPTIMIZATION OF TEACHING ASSISTANT EVALUATION SVM BASED ON PARTICLE SWARM OPTIMIZATION

Although, CV methods can find the best c and g with the highest classification accuracy by the grid search function, sometimes it will be very time-consuming if you want to find the best parameters in a wider range. Heuristic algorithms need not traverse all the parameter values within the grid to find the global optimal solutions.

Particle swarm optimization: Particle Swarm Optimization (PSO) is a heuristic method that optimizes a problem by iteratively trying to improve a candidate solution. PSO does not use the gradient of the problem being optimized which means PSO does not require for the optimization problem to be differentiable as is required by classic optimization methods such as gradient descent and quasi-newton methods.

PSO algorithm first initializes a population of particles in the solution space. Each particle represents a candidate solution of the optimization problem and is characterized by the position, velocity and fitness value. The fitness values are the merits of the particles. These particles are moved around in the search-space. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position. When improved positions are being discovered these will then come to guide the movements of the swarm. The process is repeated and by doing so it is hoped but not guaranteed, that a satisfactory solution will eventually be discovered.

Given the search-space is D dimensions and a population $X = (X_1, X_2, \ldots, X_n)$ is composed by n particles. $X_i = (X_{i1}, X_{i2}, \ldots, X_{iD})^T$ is a D-dimensional vector, representing a candidate solution. According to the fitness function, the fitness value of X_i can be calculated. $V_i = (V_{i1}, V_{i2}, \ldots, V_{iD})^T$ is the velocity of the ith particle. $P_i = (P_{i1}, P_{i2}, \ldots, P_{iD})^T$ represents the particle's best known position and $P_g = (P_{g1}, P_{g2}, \ldots, P_{gD})^T$ is the swarm's best known position. In each iteration, each particle updates its velocity V_i and position X_i according to the following formula:

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 (P_{id}^{k} - X_{id}^{k}) + c_2 r_2 (P_{gd}^{k} - X_{id}^{k}) \quad (2)$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \quad (3)$$

where, ω is the inertia weight; d = 1,2,..., D; I = 1,2,...n; k is the current iteration; V_{id} is the particle velocity; non-negative constants c_1 and c_2 are the acceleration parameters; r_1 and r_2 are two random numbers between 0 and 1.

Algorithm design of optimizing parameters of SVMs based on PSO: According to the algorithm of PSO, the paper designed the whole process of teaching assistant evaluation based on SVM with parameters optimization by PSO, as Fig. 5 shows. The accuracy of the training instance by CV is taken as the fitness function of PSO.

Related initial parameters settings of PSO and SVM are shown as Table 2.

Results analysis: The parameters optimization process and results are shown as Fig. 6 and 7. Best validation accuracy equals 59.6026%, best c 84.2688 and best g 60.5749. The teaching assistant evaluation accuracy of the testing dataset is 95.3642% (144/151) which is equal to the optimal accuracy attained by CV method but with smaller c and g. And the elapsed time is reduced to 117.339507 seconds.

Table 2: Initial parameters settings of PSO and SVM

Parameters settings of PSO		Parameters settings of SVM	
Parameters	Initial values	Parameters	Initial values
c1	1.5	v	3
c2	1.7	cmax	100
maxgen	200	cmin	0.1
sizepop	20	gmax	1000
k	0.6	gmin	0.01
w	1		

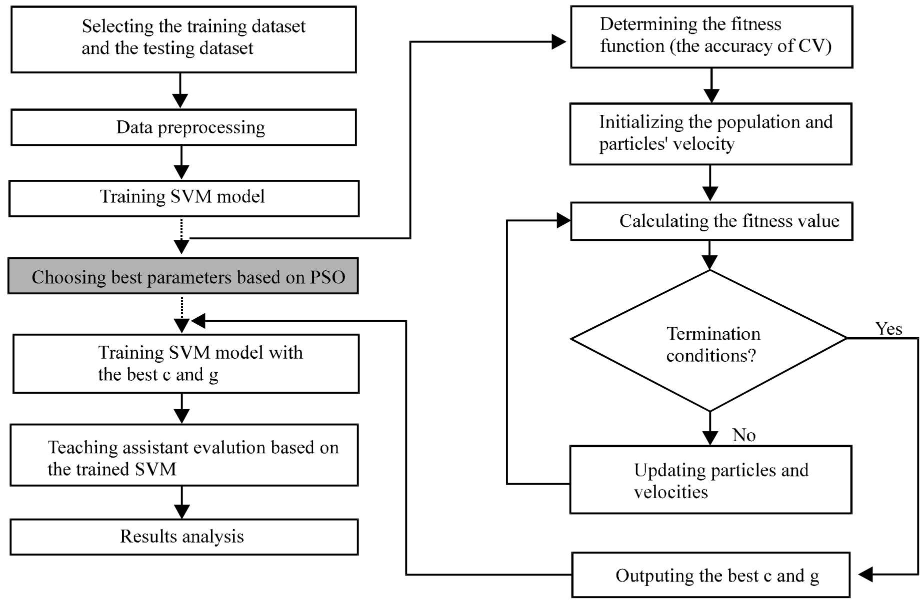

Fig. 5: The whole process of teaching assistant evaluation based on SVM with parameters optimization by PSO

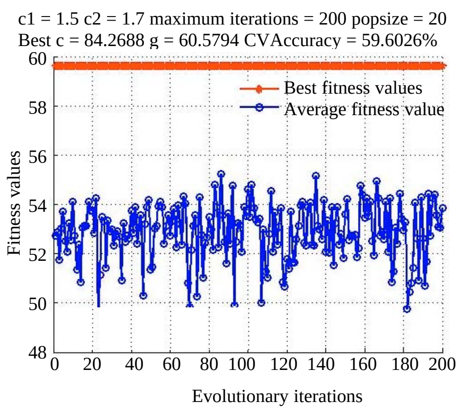

Fig. 6: The iterations of PSO to optimize SVMs' parameters

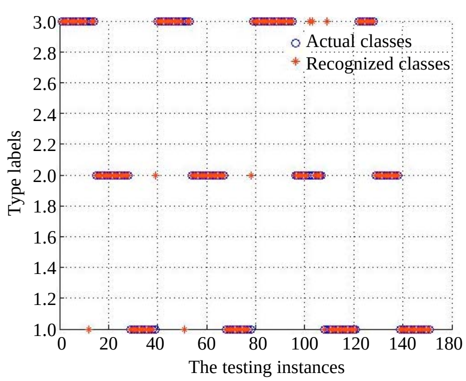

Fig. 7: Actual classes and recognized classes of SVM with optimized parameters by PSO

CONCLUSIONS

Support vector machine is a novel machine learning method which is powerful for the problem with small sampling, nonlinear and high dimension. However, an important problem encountered in setting up SVM models is how to select their kernel functions and parameters. Support vector machines with appropriate parameters can provide a good tool for enhancing the classification accuracy. Different kernel functions may result in different results with the same parameter and different parameters may also produce different recognition results with the same kernel function.

Cross Validation (CV) method can attain the optimal parameters in some sense which may effectively avoid the overlearning and underlearning. The trained SVM models with the selected parameters by CV method are more effective than the models trained with randomly selected parameters. Although, CV method can find best parameters with the highest classification accuracy by the grid search function, sometimes it is very time-consuming.

The best parameters of teaching assistant evaluation, SVM can be achieved by PSO which may produce the same accuracy with CV method. Although the paper produced satisfactory recognition results by applying PSO to select the best parameters of SVM, the choice of PSO parameters is not well considered which may have a large impact on optimization performance. How to select the appropriate parameters of PSO when it is used to optimize the SVMs' models needs further works.

REFERENCES

Abibullaev, B., W.S. Kang, S.H. Lee and J. An, 2010. Classification of cardiac arrhythmias using biorthogonal wavelets and support vector machines. Int. J. Advancements Comput. Technol., 2: 24-34.

Cobb, R., 2005. Training opportunities and career development for teaching assistants working with children with visual impairment in the United Kingdom. Proc. Int. Congress Ser., 1282: 811-815.

Cremin, H., G. Thomas and K. Vincett, 2003. Learning zones: An evaluation of three models for improving learning through teacher/teaching assistant teamwork. Support Learn., 18: 154-161.

Frank, A. and A. Asuncion, 2010. UCI machine learning repository Irvine. University of California, School of Information and Computer Science, USA. http://archive.ics.uci.edu/ml/

Gorsuch, G.J., 2006. Discipline-specific practica for international teaching assistants. English Specific Purposes, 25: 90-108.

Keerthi, S.S. and C.J. Lin, 2003. Asymptotic behaviors of support vector machines with gaussian kernel. Neural Comput., 15: 1667-1689.

Osterlund, K. and K. Robson, 2009. The impact of ICT on work-life experiences among university teaching assistants. Comput. Educ., 52: 432-437.

Samson, S. and M.S. Millet, 2003. The learning environment: First-year students, teaching assistants and information literacy. Res. Strategies, 19: 84-98.