Improvements in RBF Kernel using Evolutionary Algorithm for Support Vector Machine Classifier

[1]N.T.Renukadevi and [2]P.Thangaraj

[1]Assistant Professor, Dept.of CT-UG, Kongu Engineering College, Perundurai, Tamilnadu
[2]Professor and Head, Dept. of CSE, Bannariamman Institute of Technology, Sathy, Tamilnadu
renuka.kec@gmail.com

Abstract: The automatic medical image classification is useful in building a content-based medical image retrieval system. In this paper, a classification system for CT Medical Images is presented. Coiflet wavelets are used to extract feature from the CT images. The extracted features are then classified using Support Vector Machine (SVM) with Radial Basis Function (RBF). The accuracy of the SVM depends on the parameters C and gamma of the RBF kernel. The parameter selection is treated as an optimization problem wherein a search technique is used to the optimal parameters to maximize the SVM performance. In this work, Particle Swarm Optimization (PSO) is implemented to select the values of two SVM parameters for classification problems. The PSO is further modified using Genetic Algorithm to achieve optimal parameter values for the RBF kernel.
[N.T.Renukadevi and P.Thangaraj. **Improvements in RBF Kernel using Evolutionary Algorithm for Support Vector Machine Classifier**. *Life Sci J* 2013;10(7s): 454-459] (ISSN:1097-8135). http://www.lifesciencesite.com.

Keywords: Content Based Image Retrieval (CBIR), Computed Tomography (CT), Coiflet wavelets, Support Vector Machine (SVM), Radial Basis Function (RBF), Particle Swarm Optimization (PSO), Genetic Algorithm

1. Introduction

Computerized tomography uses X-rays, and a computer to create detailed body image. CT scan images are called tomograms having more detail than standard X-rays. A CT scan produces images of structures within the body like internal organs, blood vessels, bones and tumours [1]. The various types of CT scans that help investigate particular body areas include: Head scans, which can check for suspected brain tumours and arterial bleeding/swelling; head scans also investigate the brain after a stroke (when blood supply to a part of the brain is cut). Abdominal scans can detect tumours and diagnose conditions causing internal organs like liver, kidneys, pancreas, intestines or lungs, to become enlarged or inflamed. Vascular scans assess conditions affecting blood flow to various parts of the body. Bone scans assess bone injuries and disease, specially the spine.

The computer-aided diagnosis systems (CAD) are widely used for clinical diagnosis and treatment. Content-based image retrieval (CBIR) of medical images, according to its domain specific image features, is an valuable tool for physicians. A method for automatic classification of computed tomography (CT) images of different types is presented in this paper. The proposed method has three major steps: 1. Feature are extracted from the CT images using Coiflet wavelets; 2. The features extracted are classified using Support Vector Machine; 3. The parameters of the SVM are optimized using Particle Swarm Optimization (PSO), and modified PSO with a genetic algorithm.

Support Vector Machines (SVMs) achieve good empirical on different learning problems when compared to the other machine learning methods [2]. Though, the accomplishments of SVMs are governed by the adequate choice of parameters of the kernel and the regularization parameters. The parameter selection is treated as an optimization problem wherein a search technique is used to the optimal parameters to maximize the SVM performance [3-5]. Even though search techniques represent a systematic approach for parameter selection in SVM, it can also be expensive if the number of parameters to be evaluated during the search process is large [2].

A different method to SVM parameter selection is based on Meta-Learning wherein the SVM parameter selection is treated as supervised learning tasks [6]. In this, the characteristics of the training examples and the performance achieved for a set of parameters for the problem are stored. Meta-learners on the basis of the set of training examples received as input, predicts the best values for the parameters for a new problem. The Meta-Learning is also used as it is less expensive when compared to the search approach.

The SVM parameters are often selected by calculating different combinations of parameters and utilizing the combination which achieves the best performance for the particular dataset. To automatize the search process, various search and optimization techniques are used [7, 8]. The search space is made up of a set of possible combination of parameters and a fitness function corresponding to performance measure achieved by the SVM is considered. Different search techniques available in the literature are based on Evolutionary Algorithms [4], gradient based techniques [9], Tabu Search [10] and so on.

In the current work, PSO is implemented for parameter selection of the RBF kernel of SVM. "Swarm intelligence" is usually used for optimization, to maximize or minimize the cost function by searching for a set of variable is termed optimization. Swarm optimizations are based on the collective behaviour of the bees or ants, or social behaviour of bird flocking and fish schooling. The Particle Swarm Optimization algorithm is a population-based stochastic search algorithm and is efficient in solving complex non-linear optimization problem [11]. The PSO is popular as it is easily implemented, computationally inexpensive. To prevent premature convergence in PSO, the PSO is modified using genetic algorithm (GA). The rest of the paper is organized as follows: Section 2 reviews some of the related works available in the literature; section 3 details the materials and proposed methodology with section 4 discussing the results and section 5 concludes the paper.

2. Related Works

Kharrat et al [12] proposed a new approach for automatic classification of Magnetic Resonance (MR) human brain images as normal and abnormal. Wavelets Transform (WT) is used as input module to Genetic Algorithm (GA) for feature selection and Support Vector Machine (SVM) for classifying the MR images. The GA requires very less computation when compared with Sequential Floating Backward Selection (SFBS) and Sequential Floating Forward Selection (SFFS) methods. A reduction rate of 88.63% is realized and classification rate of 100% was obtained using the support vector machine.

Zhang et al [13] presented adaptive chaotic particle swarm optimization (ACPSO) for optimizing parameters. The methodology is used to classify MR brain image as normal or abnormal. Wavelet transforms are used to extract features and principle component analysis (PCA) is applied to reduce the dimensions of features. Feed forward neural network is used to classify the features. To enhance the generalization, K-fold stratified cross validation was applied. The proposed method was evaluated using 160 images (20 normal, 140 abnormal), and classification accuracy of 98.75% was achieved.

Gal et al [14] presented a multi-disciplinary approach to address the classification problem. The proposed methodology combined image features, meta-data, textual and referential information for classification of medical images. ImageCLEF 2011 medical modality classification data set was used to evaluate the system's accuracy. Multiple kernels were used for classification; significantly better classification accuracy was achieved as the kernels were selected for different features. Best classification accuracy of 88.47% obtained and outperforms the other methods available in the literature.

Umamaheswari et al [15] proposed PSO SVM for classification of DICOM images. The proposed method was used to recognise and classify brain images as normal and abnormal. Optimal recognition and detection of disease in DICOM images is crucial for the diagnosis process. The proposed method focused on recognition and classification based on combined approach of digital image processing incorporating PSO, GA and SVM. The combined approach by using PSO-SVM achieves high approximation capability and faster convergence.

3. Materials and Methods

Feature extraction using Coiflet wavelet

Coiflets are discrete wavelets designed to have scaling functions with vanishing moments [16]. The wavelet is near symmetric with N/3 vanishing moments and N/3-1 scaling functions. The function Ψ has 2N moments equal to 0 and the function φ has 2N-1 moments equal to 0. The support length of the two functions is 6N-1 [17]. The coifN Ψ and φ are considerably more symmetric than the dbNs. The coifN are compared to *db3N* or *sym3N* when considering the support length. When number of vanishing moments of Ψ is considered, coifN is compared to *db2N* or *sym2N*.

If s is a sufficiently regular continuous time signal, for large j the coefficient [18]

$\left\langle s, \phi_{j,k} \right\rangle$ is approximated by $2^{-j/2} s\left(2^{-j} k\right)$

If s is a polynomial of degree d, $d \leq N - 1$, then the approximation becomes equality.

Support Vector Machine (SVM)

Given a set of features that can be represented in space, SVM maps features non-linearly into n dimensional feature space when provided with features set that can be represented in space. When a kernel is introduced with high computation the algorithm uses inputs as scalar products with classification being solved by translating the issue into a convex quadratic optimization problem with a clear solution being obtained by convexity [19]. In SVM, an attribute is a predictor variable and a feature a transformed attribute. A set of features describing an example is a vector. Features define the hyperplane. SVM aims to locate an optimal hyperplane separating vector clusters with a class of attributes on one side of the plane with the on the other side. The margin is the distance between hyperplane and support vectors. SVM analysis orients the margin that space between it and support vectors is maximized. Figure 1 shows a simplified SVM process overview.

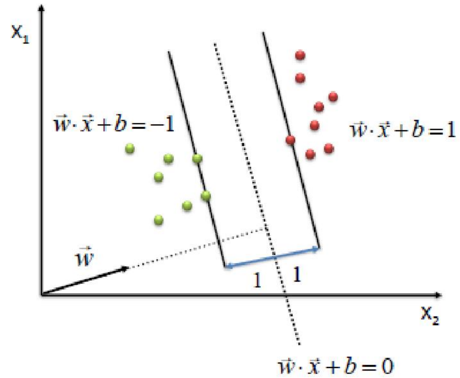

Figure 1: Support vector machine

Given a training set of $(x_i, y_i), i1,2,....l$ where $x_i \in R^n$ and $y \in \{1,-1\}^l$, SVM has to solve the optimization problem [20] of:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i$$

Subject to $y_i\left(w^T\phi(x_i)+b\right) \geq 1-\xi_i$ and $\xi_i \geq 0$.

The function ϕ maps the vectors x_i in higher dimensional space. C>0 is penalty parameter of the error term.

This optimization model is solved through the use of the Lagrangian method, equal to the method for solving optimization problems in a separable case. One maximizes the dual variables Lagrangian:

$$\underset{\alpha}{Max} \quad L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \left\langle x_i . x_j \right\rangle$$

subject to: $0 \leq \alpha_i \leq C \quad i = 1,..,m$ and $\sum_{i=1}^{m} \alpha_i y_i = 0$

To find the optimal hyperplane, a dual LagrangianLD(α) should be maximized as regards non-negative α_i under the constrains $\sum_{i=1}^{m} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$. The penalty parameter C, now the upper bound on α_i, is user determined.

A kernel function is defined as $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. The Radial Basis function is given as follows:

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \ , \gamma > 0$$

A proper parameter setting improves SVM classification accuracy. There are two parameters to be determined in the SVM model with the RBF kernel: C and gamma (γ). The value of C and γ influence the learning performance of the SVM [21]. Intuitively the γ parameter defines the distance a single training example can reach, with low values meaning 'far' and high values meaning 'close'. The C parameter trades off training examples misclassification against decision surface simplicity. A low C ensures a smooth decision surface while a high C attempts to classify training examples correctly. Experiments are undertaken to evaluate SVM performance through variations of the γ and C parameters.

To optimize the parameters C and γ, PSO [22] is adapted to execute the search for optimal combination (C, γ). The objective function for evaluating the quality of combination of parameters is based on Root Mean Squared Error (RMSE) achieved by the SVM in a 10-fold cross validation experiment. Thus, the PSO finds the combination of parameters with the lowest RMSE.

Each particle *i* represents a combination of parameters which also indicates the position of the particle in the search space. The velocity of the particle is indicative of the direction of the search of the particle. The PSO algorithm keeps updating the position and velocity of the particle in each iteration which leads to best regions in the search space. The velocity and the position of the particle are updated as follows:

$$v_i^d = wv_i^d + c_1 r_1 \left(p_i^d - x_i^d\right) + c_2 r_2 \left(p_g^d - x_i^d\right)$$

$$x_i^d = x_i^d + v_i^d$$

where w is the Inertia weight; d represents the number of dimensions; i is the size of the population; the two "best" values - pbest and gbest - of a particle where 'pbest' $\left(p_i^d\right)$ is the best solution achieved by the particle till then and 'gbest' $\left(p_g^d\right)$ is the best value obtained till then by any particle in the population; c_1, c_2 are positive constants; r1 and r2 are random values with value between [0, 1]. The flowchart of the achieving optimized parameter using PSO is shown in Figure 2.

The parameters w, c_1, c_2, r_1, r_2 in the PSO affect the performance of the algorithm significantly [23]. The inertia weight controls the exploration and exploitation; generally 0 < w < 1 for the particles to converge. Higher value of w (near 1) favours global search and lower values less than 0.5 favours local search. The random numbers r1 and r2 are with value between [0, 1]. The coefficients c_1 and c_2 are usually equal (i.e., c_1=c_2) and has a value in the range of [0, 4]. The value of c_1 and c_2 are significant as convergence is dependent on these values.

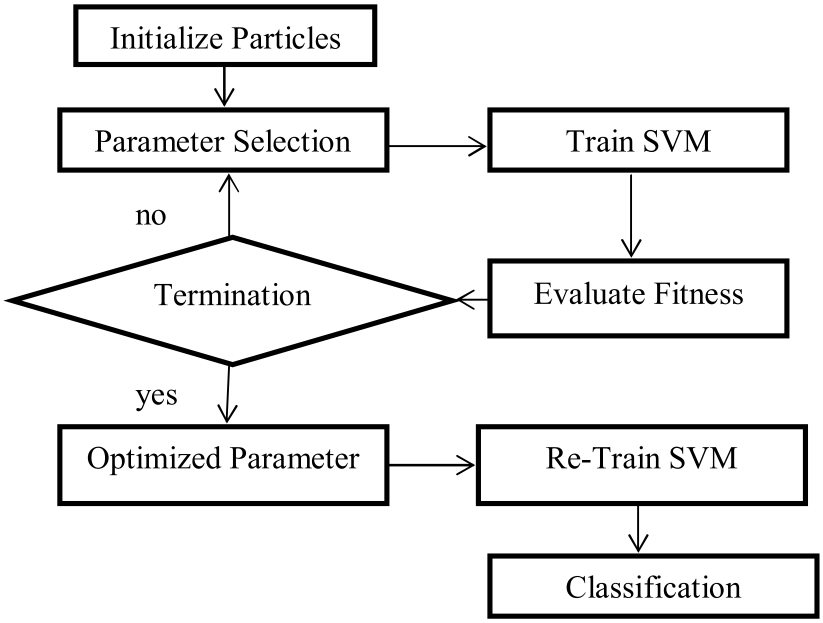

Figure 2: Flowchart – PSO optimizing SVM

Convergence is feasible when

$1 > w > \frac{1}{2}(\phi_1 + \phi_2) - 1$ where $\phi_1 = c_1 r_1$, $\phi_2 = c_2 r_2$

Also for stochastic ϕ_1 and ϕ_2, convergences results when

$\phi_{ratio} = \frac{\phi_{crit}}{c_1 + c_2}$ is close to 1 and

$$\phi_{crit} = \sup \phi \left[0.5\phi - 1 < w,\ \phi \in \left(0, c_1 + c_2\right) \right]$$

To avoid premature convergence and to combine the coordinates to achieve high convergence speed, the classical PSO is modified using GA. The GA incorporated, coordinates the relationship of the PSO parameters to maximize the performance. The GA generates a population by encoding the PSO parameters. The fitness value is calculated based on ϕ_{ratio} with the objective to maximize it. Genetic operators, selection, crossover and mutation, are applied to generate the next generation. On termination of the GA algorithm, the PSO parameters obtained are updated into PSO algorithm.

4. Results and Discussion

Experiments were conducted using 150 CT scans images of brain, chest and colon. Features were extracted using Coiflet wavelet. Experiments were conducted to evaluate the classification accuracy for SVM-RBF, with PSO and with modified PSO. All the experiments were conducted for 10-fold cross validation. The classification accuracy and the root mean square error (RMSE) achieved is tabulated in Table 1. Figure 3 shows the classification accuracy and Figure 4 show the RMSE.

Table 1: Classification Accuracy and RMSE

Classifier	Classification Accuracy %	RMSE
Naïve Bayes	90	0.2582
SVM-RBF	88.67	0.265
SVM, PSO	90.67	0.214
SVM, Modified PSO	92.67	0.196

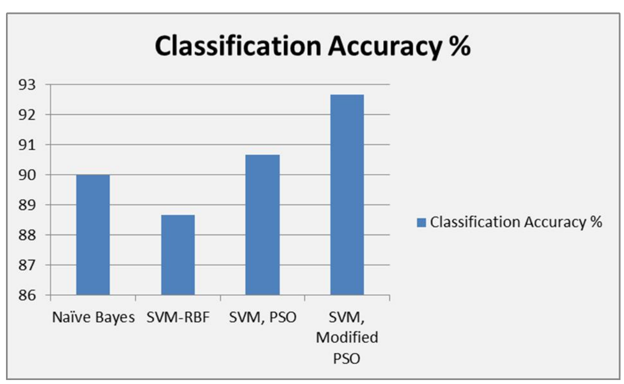

Figure 3: Classification Accuracy

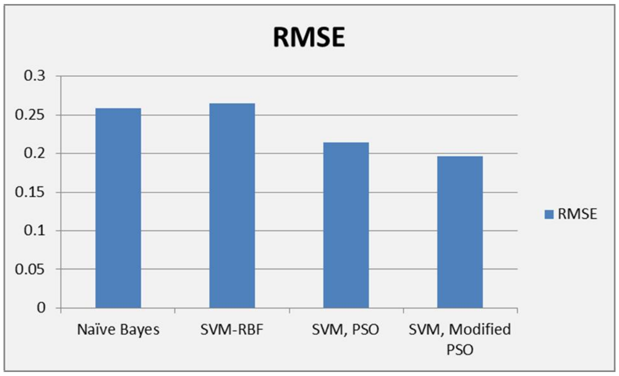

Figure 4: Root Mean Square Error

It is observed from the Table and Figures that the proposed modified PSO improves classification accuracy and reduces the RMSE significantly. Table 2 tabulates the precision and recall achieved.

Table 2: Precision and Recall

Classifier	Precision	Recall
Naïve Bayes	0.900	0.900
SVM-RBF	0.887	0.887
SVM, PSO	0.908	0.907
SVM, Modified PSO	0.928	0.927

Similar to the classification accuracy, precision and recall are high for the proposed modified PSO.

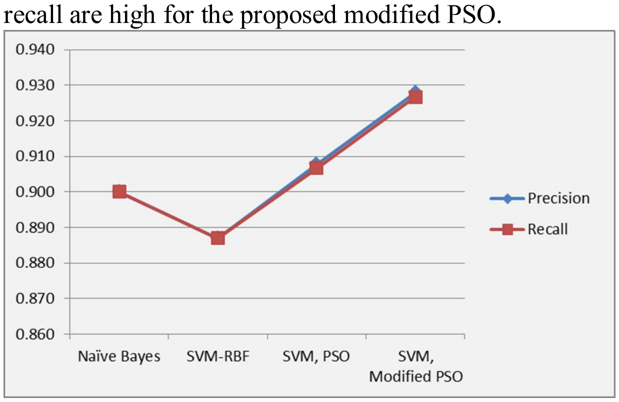

Figure 5: Precision and Recall

5. Conclusion

In this paper, to improve the performance of the SVM-RBF for classifying the CT images, the SVM parameters C and Gamma (γ) are optimized. Particle Swarm Optimization (PSO) is implemented to select the values of two SVM parameters for classification problems. To avoid premature convergence and to combine the coordinates to achieve high convergence speed, the classical PSO is modified using GA. The experiments were conducted for 10-fold cross validation. The classification accuracy and the root mean square error (RMSE) achieved for the proposed modified PSO is significantly better.

References

1. Kalender, W. A. (2011). Computed tomography. Wiley-VCH.
2. C. Soares, P. Brazdil, and P. Kuba, "A meta-learning approach to select the kernel width in support vector regression," *Machine Learning*, vol. 54, no. 3, pp. 195–209, 2004.
3. Lorena and A. de Carvalho, "Evolutionary tuning of svm parameter values in multiclass problems," *Neurocomputing*, vol. 71, pp. 16–18, 2008.
4. S. Lessmann, R. Stahlbock, and S. Crone, "Genetic algorithms for support vector machine model selection," in *International Joint Conference on Neural Networks*, 2006, pp. 3063–3069.
5. X. Guo, J. Yang, C. Wu, C. Wang, and Y. Liang, "A novel ls-svms hyper-parameter selection based on particle swarm optimization," *Neurocomputing*, vol. 71, pp. 3211–3215, 2008.
6. S. Ali and K. A. Smith-Miles, "A meta-learning approach to automatic kernel selection for support vector machines," Neurocomputing, vol. 70, no. 1-3, pp. 173–186, 2006.
7. F. Friedrichs and C. Igel, "Evolutionary tuning of multiple svm parameters," *Neurocomputing*, vol. 64, pp. 107–117, 2005.
8. B. de Souza, A. de Carvalho, and R. Ishii, "Multiclass svm model selection using particle swarm optimization," in Sixth International Conference on Hybrid Intelligent Systems, 2006, pp. 441–446.
9. O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," Machine Learning, vol. 46, no. 1, pp. 131–159, 2002.
10. G. Cawley, "Model selection for support vector machines via adaptive step-size tabu search," in Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms, 2001, pp. 434–437.
11. Satyobroto Talukder, Mathematical Modelling and Applications of Particle Swarm Optimization,
12. Kharrat, A., Gasmi, K., Ben Messaoud, M., Benamrane, N., & Abid, M. (2011). Medical Image Classification Using an Optimal Feature Extraction Algorithm and a Supervised Classifier Technique. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, *3*(2), 19-33.
13. Zhang, Y., Wang, S., & Wu, L. (2010). A novel method for magnetic resonance brain image

classification based on adaptive chaotic PSO. Progress in Electromagnetics Research, 109, 325-343.
14. Cheng, B., Wang, R., Antani, S., Stanley, R. J., & Thoma, G. R. (2012, January). Graphical image classification combining an evolutionary algorithm and binary particle swarm optimization. In IS&T/SPIE Electronic Imaging (pp. 829703-829703). International Society for Optics and Photonics.
15. Umamaheswari, J., & Radhamani, G. (2012). An Amalgam Approach for DICOM Image Classification and Recognition. World Academy of Science, Engineering and Technology, 62.
16. Ingrid Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, 1992.
17. G. Beylkin, R. Coifman, and V. Rokhlin (1991),Fast wavelet transforms and numerical algorithms, Comm. Pure Appl. Math., 44, pp. 141-183
18. Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. (1996). Wavelet toolbox. The MathWorks Inc., Natick, MA.
19. B. E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In Proceedings Of The Fifth Annual Workshop On Computational Learning Theory, pages 144-152. ACM Press, 1992.
20. Jrad, N., Congedo, M., Phlypo, R., Rousseau, S., Flamary, R., Yger, F., & Rakotomamonjy, A. (2011). sw-SVM: sensor weighting support vector machines for EEG-based brain–computer interfaces. Journal of neural engineering, 8(5), 056004.
21. P. Kuba, P. Brazdil, C. Soares, and A. Woznica. Exploiting sampling and meta-learning for parameter setting support vector machines. Proceedings of the IBERAMIA, pages 217–225, 2001.
22. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, 1995; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 1995; Vol.4, pp. 1942-1948.
23. Van Den Bergh, F. (2006). An analysis of particle swarm optimizers (Doctoral dissertation, University of Pretoria).

1/8/2013