International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-4 Issue-7 December-2012

Improving RBF Kernel Function of Support Vector Machine using Particle Swarm Optimization

Smruti Rekha Das[1], Pradeepta Kumar Panigrahi[2], Kaberi Das[3],Debahuti Mishra[4]
[1]Mahavir Institute of Engineering and Technology, Bhubaneswar, Odisha
[2]Gandhi Institute of Engineering and Technology, Gunupur, Odisha, India
[3,4]Institute of Technical Education and Research, Siksha O Anusandhan University, Bhubaneswar, Odisha, India

Abstract

Support vector machine (SVM) has become an increasingly popular tool for machine learning tasks involving classification, regression or novelty detection. SVM is able to calculate the maximum margin (separating hyper-plane) between data with and without the outcome of interest if they are linearly separable. To improve the generalisation performance of SVM classifier optimization technique is used. Optimization refers to the selection of a best element from some set of available alternatives. Particle swarm optimization (PSO) is a population based stochastic optimization technique where the potential solutions, called particles, fly through the problem space by following the current optimum particles. In this paper, Principal Component Analysis (PCA) is used for reducing features of breast cancer, lung cancer and heart disease data sets and an empirical comparison of kernel selection using PSO for SVM is used to achieve better performance. This paper focused on SVM trained using linear, polynomial and radial basis function (RBF) kernels and applying PSO to each kernels for each data set to get better accuracy.

Keywords

Principal Component Analysis, Support Vector Machine, Linear Kernel Function Polynomial Kernel Function, Radial Basis Function, Particle swarm optimization.

1. Introduction

Classification is used to find a set of models that describe and distinguish data classes [1]. SVM has gained much attention; it presents a powerful new generation learning algorithm based on recent advances in statistical learning theory [2]. The SVM algorithm, which is based on mathematical programming, is originally established by Vapnik [3]. To work with gene expression data is not so easy due to high dimensionality, for this various feature reduction techniques are used. In this paper, PCA have been used to map the data set from higher dimension to lower dimension. SVM is designed to solve two-class problems. The kernel function is constructed by SVM algorithm to map the training data into a higher dimensional space when the linear separation is impossible in the original one, SVM can be generalized to compute nonlinear decision surfaces. In PSO [4], the search through the problem space can be thought of as the flight of a swarm of particles. The particles are initially distributed randomly through the problem space and given an initial velocity. Each particle keeps track of its location and fitness, as well as the best it has encountered so far in its flight [5]. The algorithm stops when some criterion is met perhaps after a certain number of iterations, or when many iterations pass without significant improvement. In this model, three different kernels using PSO are used to achieve the accuracy of SVM. This paper is organized as follows: section 2 describes related work; section 3 represents the preliminaries concepts of all the kernel functions with PSO used in SVM, section 4 deals with the proposed model, section 5 gives the experimental evaluation and result and finally section 6 deals with conclusion and future work.

2. Literature Review

Nasser H. Sweilam *et al*. [4] introduced Particle swarm optimization, Quantum-behave Particle Swarm for training SVM. Michael E. Matheny *et al*. [6] focused that radial-basis kernel SVM and polynomial kernel SVM mortality prediction models for percutaneous coronary interventions were optimized. Paulius Danenas *et al*. [7] showed that different SVM classifiers produced similar results, including Core Vector Machines based classifier, yet proper selection of classifier and its parameters remains an important problem. F Heppner *et al.* [8] proposed that PSO is an extremely simple algorithm that seems to be effecting for optimizing a wide range of functions. Chung-Jui Tu *et al*. [9] proposed that

PSO is used to implement a feature selection, and SVMs with the one-versus-rest method serve as a fitness function of PSO for the classification problem. Bo-Tsuen Chen *et al.* [10] found that the average classification accuracy rate of the approach is 100% in the training subset, and be 88.98% in the test subset, and it is evident that the PSO-SVM approach is as good as the grid search for SVM and original SVM. Dimitrios Bouzas *et al*. [11] showed that by applying PSO on the sub-class Linear Discriminant Error Correcting Output Codes framework get a significant improvement in the classification performance. Enrique Alba *et al*. [12] focused that PSOSVM is able to find interesting genes and to provide classification competitive performance. Smruti Rekha Das *et al.* [13] focused on SVM trained using linear, polynomial and RBF kernels.

3. Preliminaries

3.1 Dimensionality Reduction

The basic objective of dimensionality reduction is to remove the irrelevant or redundant attributes which can slow the mining process. There are so many feature reduction techniques like PCA, Discrete wavelet Transform (DWT), Fishers Linear discriminant analysis (LDA), Independent component analysis (ICA) and Factor analysis (FA). In PCA the original data are projected into a much smaller space, resulting in dimensionality reduction. [14]The dimensionality of the lung cancer data set, breast cancer dataset and heart disease data set has reduced by using PCA which is showed in fig.1, fig.2, fig.3 respectively.

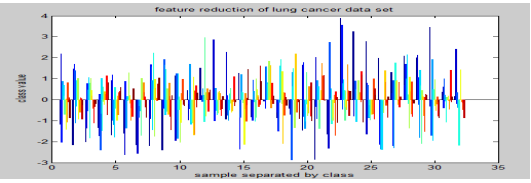

Fig .1 Feature reductions of lung cancer data set

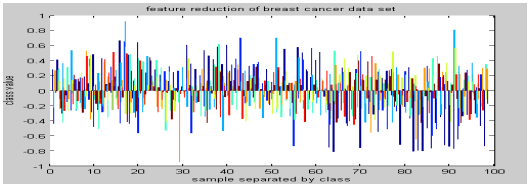

Fig. 2 Feature reduction of breast cancer data set

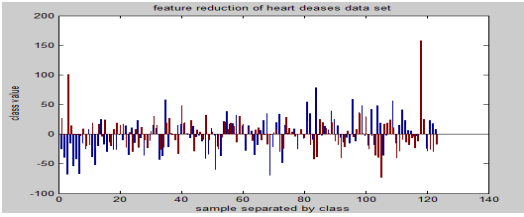

Fig.3. Feature reduction of heart disease data set

3.2 SVM

SVM is a promising new method for the classification of both linear and nonlinear data. SVM uses a nonlinear mapping to transform the original training data into a higher dimension and within this new dimension it searches for the linear optimal separating hyper plane. SVM performs the same function as an Artificial Neural Network (ANN). SVM can obtain the global optimum and the over fitting problem can be easily controlled [15]. SVM is capable of finding nonlinear decision boundaries in input space. Here, it transforms the original input data into a higher dimensional space using nonlinear mapping. For this mapping, different kernel functions are used. Once the data have been transferred to the new dimension it searches for the linear separating hyper plane in the new dimension [14].

Linear kernel function – $k(x, x_i)=x.x_i$ (1)

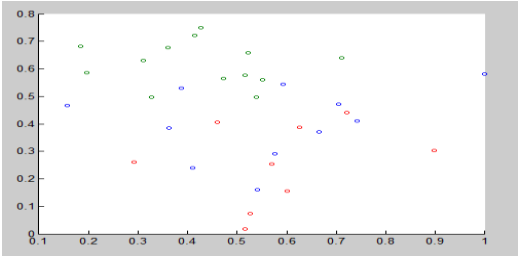

Fig. 4 Linear SVM kernel function separating class level of breast cancer data set

Here in the fig.4 it describes the SVM using linear kernel function separates the class level of breast cancer data set. Polynomial kernel function-polynomial kernel function is directional. This is due to the dot product in the kernel. The magnitude of the output depends on the output of the x_i.

$$k(x, x_i)=(x.x_i+1)^p \qquad (2)$$

Where, *p* is degree of polynomial

RBFs which have the property that each basis function depends only on the radial distance from a centre, so that $\theta_j(x) = \left(\left\|x - \mu_j\right\|\right)$. RBF was introduced for the purpose of exact function interpolation. Given a set of input vectors $\{x_1, x_2, x_3, \dots \dots, x_n\}$ along with corresponding target values, $\{t_1, t_2, t_3, \dots \dots, t_n\}$ the goal is to find a smooth function $f(x)$ that fits every target value exactly [16], so that $f(x_n) = t_n$ for *n=1,.....,N*. This is achieved by expressing $f(x)$ as linear combination of radial basis function, one centred on each data point.

$$f(x) = \sum_{n=1}^{N} w_n \, h(\|x - x_n\|), \qquad (3)$$

Where, w_n is the value of coefficients which is found by least-square. How the RBF kernel function is mapping is described in fig.5 and fig.6.

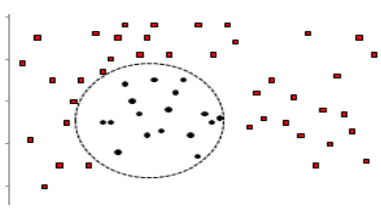

Fig. 5 Radial Basis Function

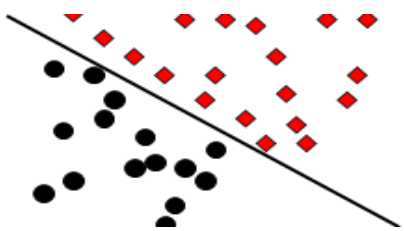

Fig. 6 RBF mapping

3.3 PSO

PSO is a relatively new computational learning algorithm. It bears some resemblance to evolutionary computation. The goal of PSO is to find the global optimum of some multidimensional (usually nonlinear) function. In PSO, each single candidate solution is *an individual bird of the flock*, that is, a particle in the search space. Each particle makes use of its individual memory and knowledge gained by the swarm as a whole to find the best solution. All of the particles have fitness values, which are evaluated by fitness function to be optimized, and have velocities which direct the movement of the particles. During movement, each particle adjusts its position according to its own experience, as well as according to the experience of a neigh boring particle, and makes use of the best position encountered by itself and its neighbour. The particles move through the problem space by following a current of optimum particles. The initial swarm is generally created in such a way that the population of the particles is distributed randomly over the search space. At every iteration, each particle is updated by following two *best* values, called *pbest* and *gbest*. Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) the particle has achieved so far. This fitness value is stored, and called *pbest*. When a particle takes the whole population as its topological neighbour, the best value is a global *best* value and is called *gbest* [9]. PSO is a versatile population-based optimization technique, in many respects similar to evolutionary algorithms (EAs). PSO has been shown to perform well for many static problems. However, many real-world problems are dynamic in the sense that the global optimum location and value may change with time. PSO is a population based heuristic search technique; each particle represents a potential solution within the search space. Each particle has a position vector x_i, a velocity vector v_i, the position at which the best position $pbest_i$ encountered by the particle so far, and the best [4] position of all particles gbest in current generation. The updating equations of PSO are as follows:

$$v_i(t+1) = w\, v_i(t+1) + c_1 r_1 \left(x_{pbest_i}(t) - x_i(t)\right) + c_2\, r_2 \left(x_{gbest}(t) - x_i(t)\right) \qquad (4)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (5)$$

Where the parameters c_1 and c_2 are set to constant value, which are normally taken as 2, r_1 and r_2 are two random values, uniformly distributed in [0,1], *w* is inertia weight which controls the influence of previous velocity on the new velocity.

Dynamic PSO

Optimization with particle swarms has two major ingredients, the particle dynamics and the particle information network. The particle dynamics are de-rived from swarm simulations in computer graphics, [17] and the information sharing component is

inspired by social networks. These ingredients combine to make PSO a robust and efficient optimizer of real-valued objective functions. PSO must be modified for optimal results on dynamic environments typified by the moving peaks benchmark (MPB).

Canonical PSO

In PSO, population members (particles) possess a memory of the best (with respect to an objective function) location that they have visited in the past, *pbest*, and of its fitness. In addition, particles have access to the best location of any other particle in their own network. These two locations (which will coincide for the best particle in any network) become attractors in the search space of the swarm. Each particle will be repeatedly drawn back to spatial neighbourhoods close to these two attractors, which themselves will be updated if the global best and/or particle best is bettered at each particle update. Several network topologies have been tried, with the star or fully connected network remaining a popular choice for unimodal functions. In this network, swarm is a single *gbest* global best attractor representing the best location found by the entire swarm. Particles possess a velocity which influences position updates according to a simple discretization of particle motion

$$v(t+1) = v(t) + a(t+1) \quad (6)$$
$$x(t+1) = x(t) + v(t+1) \quad (7)$$

Where *a*, *v*, *x* and *t* are acceleration, velocity, position and time (iteration counter) respectively. Though these two equations are similar to particle dynamics in swarm simulations, but PSO particles do not follow a smooth trajectory, instead moving in jumps, in a motion known as a flight [18].

4. Schematic Representation

Proposed model is verified with three different data sets i.e. breast cancer, lung cancer and heart disease. Three fourth of the total data set is used as the training data and one third of the total data is used as the testing data. With the aim of facing classification problem with gene expression data, an innovative version of PSO, based on the geometric framework presented in, has been developed in this work. In SVM the Lagrange multipliers α, constitute a vector $\alpha = [\alpha_1, \alpha_2, \dots\dots, \alpha_l]$ in the one-dimensional space, the optimization can be solved by PSO. Differing from the general PSO, all particles of the [4] PSO training SVM must satisfy both constraints

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \text{ and } c \geq \alpha_i \geq 0, \forall i \quad (8)$$

Thus the PSO algorithm must be improved according to constraint $c \geq \alpha_i \geq 0, \forall i$. The model uses PCA for feature reduction, and then SVM classifies using RBF, Polynomial and Linear kernel functions. Each and every data set is taken as the individual input for every kernel function of SVM which apply PSO technique to get better accuracy as shown in fig.7.

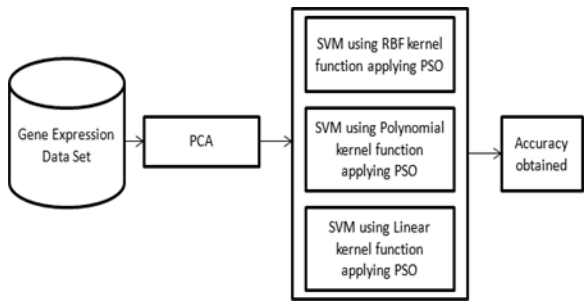

Fig. 7 Block diagram of proposed system

5. Experimental Evaluation

The model is tested on three different data sets and compared the performance in terms of accuracy and speed. The first one is Breast cancer data set [19], the second one is Lung cancer data set and the third one is Heart disease data set. The Breast cancer data is having the dimension 98*1213, after reduction it is 98*97. It has three class levels from which 1-11 belongs to class1, 12-61 belongs to class2 and rest is belongs to class3. The lung cancer data set contains total 32*56 numbers of data. From where 1-9 data is belongs to class1, 10-22 belongs to class2 and 23-32 belongs to class3. The Heart disease is having the dimension 123*14 from where 1-8 belongs to class0, 9-56 belongs to class1, 57-88 belongs to class2, 89-118 belongs to class 3 and 119-123 belongs to class5. The dataset we used in this study was obtained from the UCI Repository. SVM linear function divides the each individual data sets in two class level such as (1) and (-1). As the above data set is more than two class levels so again the data is further classified using linear function. At last the accuracy has obtained and given in table 1. SVM polynomial function maps a two dimensional input vectors into a multi-dimensional feature space. Applying the non-linear support vector classification to the linearly non

separable training data, it produces the classification *(c=∞)*. The margin is no longer of constant width due to the non-linear projection into the input space. The training data is now classified correctly. It is always possible to map the input space into a dimension which is greater than the number of training points and produce a classifier with no classification error on the training set. SVM RBF has [20] received significant attention, most commonly with a Gaussian of the form

$$k(x, x') = \exp\left(\|x - x'\|^2\right) \quad (9)$$

The kernel functions return the inner product between two points in a suitable Feature space, thus defining a notion of similarity, with little computational cost even in very high-dimensional spaces. PSO is used to serve as optimization technique [8] for classification problems. It helps to improve the performance owing to its smaller number of simple parameter settings. PSO is an evolution computing technology which simulates the social behaviour of fish in a school. At each iteration, a particle will, according to its fitness value and swarm fitness value, be optimized. The main task in this investigation is to find the best choice among different kernel functions. SVM using RBF gives 93% accuracy [13], but from table-1 it is found that Kernel based SVM using RBF applying PSO on gene expression data set gives 94% accuracy. In figure 8 it is shown SVM using RBF kernel function applying PSO technique separates the class level of lung cancer data set.

Table1. Classification results obtained using different kernels of SVM using PSO

Data sets	Actual Size	Reduce after PCA	Classification Using		
			Linear + PSO	Polynomial + PSO	RBF +PSO
Breast Cancer	98*1213	98*97	87	91.87	93.55
Lung Cancer	34*56	32*55	87.12	91.12	93.78
Heart disease	123*14	123*2	87.23	91.23	93.23

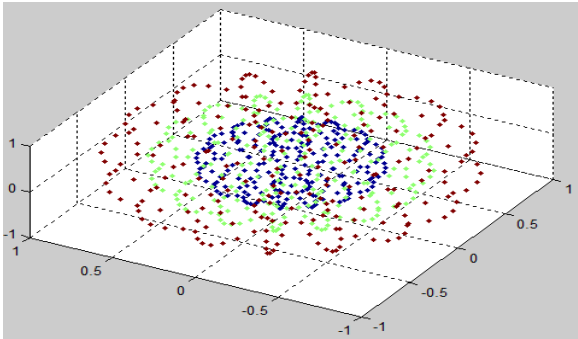

Fig. 8 SVM using RBF kernel function applying PSO technique for separating class level of lung cancer data set.

6. Conclusion and Future Work

To place the SVM's performance trained by PSO, records slightly higher overall accuracy than other techniques. When using the SVM, few obstacles are confronted: how to choose the kernel function and optimal input feature subset for SVM, and how to get the optimum result. These obstacles are crucial because the feature subset choice influences the appropriate kernel parameters and vice versa. Feature selection and appropriate kernel function is an important issue in building classification systems Building a model that can handle the three obstacles at the same time is a very important issue and needs further research and work in the future.

References

[1] Z. Zhanga, D. Zhang, Y. Tian, "Kernel-based multiple criteria linear programming classifier", Procedia Computer Science, Volume (1), pp. 2407– 2415, 2010.

[2] J. Suykens, A. K, j. Vandewalle, "Least Squares Support Vector Machine Classifiers", Neural Processing Letters, Volume (9), pp. 293–300,1999.

[3] V. N. Vapnik, "The Nature of Statistical Learning Theory", Springer, New York, 2000.

[4] N.H. Sweilam, A. Tharwat, A. Abdel, N. K. Moniem, "Support vector machine for diagnosis cancer disease", Egyptian Informatics Journal , Volume (11), pp. 81–92, 2010.

[5] M. Clerc, Kennedy, "The particle swarm–explosion, stability, and convergence in a multidimensional complex space", IEEE Transactions on Evolutionary Computation, Volume (6), Issue (1), pp. 58-73, February, 2002.

[6] M.E. Matheny, F.S. Resnic, N. Arora, L. Ohno-Machado, "Effects of SVM parameter

optimization on discrimination and calibration for post-procedural PCI mortality", Journal of Biomedical Informatics, Volume(40), pp. 688–697, 2007.

[7] P. Danenasa, G. Garsvab, S. Gudasc, "Credit Risk Evaluation Model Development Using Support Vector Based Classifiers", Procedia Computer Science, Volume (4), pp. 1699–1707, 2011.

[8] F. Heppner, U. Grenander, "A stochastic nonlinear model for coordinated bird flocks", AAAS publications, Washington, 1990.

[9] Tu, Chung-Jui, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang. "Feature selection using PSO-SVM." IAENG International journal of computer science 33, no. 1 (2007): 111-116.

[10] B.T. Chen, M.Y. Chen, "Applying particles swarm optimization for support vector machines on predicting company financial crisis", International Conference on Business and Economics Research, 2010.

[11] N. Arvanitopoulos, D. Bouzas, A. Tefas, "Subclass Error Correcting Output Codes Using Fisher's Linear Discriminant Ratio, Pattern Recognition", pp. 2953-2956, 2010.

[12] E.Alba, J. Garc´ıa-Nieto, L. Jourdan, E. Talbi, "Gene selection in Cancer Classification using PSO/SVM and GA/SVM Hybrid Algorithms", Congress on Evolutionary Computation, 2007.

[13] S.R. Das, K.Das, D.Mishra, K.Shaw, S. Mishra, "An Empirical Comparison Study on Kernel Based Support Vector Machine for Gene Expression DataSet", Procedia Engineering, Volume 38, pp.1340-1345, 2012.

[14] J. Han, M. Kamber, "Data Mining concepts and techniques", Second Edition, Elsevier.

[15] C. Burges, "A tutorial on support vector machines for pattern", Volume (2), pp. 1–47, 1989.

[16] M. Christopher Bishop, "Pattern recognition and machine learning", Springer, 2007.

[17] J. Branke, H. Schmeck, "Designing evolutionary algorithms for dynamic optimization problems", Theory and Application of Evolutionary Computation: Recent Trends, pp. 239-262, 2002.

[18] B. Mandelbrot, "The Fractal Geometry of Nature, W. H. Freeman and Company, 1983.

[19] www. uci repository.com.

[20] X. Y. Wang, E.N. Miao, H. Y. Yang, "A new SVM-based Image watermarking using Gaussian-Hermite moments", Applied Soft Computing, pp. 12: 887-903, 2012.

Smruti Rekha Das is working as Assistant Professor in the department of Computer Sc. & Engineering at Mahavir Institute of Engineering and Technology, Bhubaneswar, Odisha, India. She has completed her M.Tech from Siksha O Anusandhan University, Bhubaneswar in 2012. Her research area includes Data mining, Bioinformatics. She has many National and International papers to her credit.

Pradeepta Kumar Panigrahi is working as Assistant Professor in the department of Computer Applications at Gandhi Institute of Engineering and Technology, Gunupur, Odisha, India. He has completed his M.Tech from Biju Pattnaik University of Technology, Bhubaneswar. His research area includes Data mining, Bioinformatics.

Kaberi Das is working as Assistant Professor in the department of Computer Applications at Institute of Technical Education and Research (ITER), Siksha O Anusandhan University, Bhubaneswar, Odisha, India. She is perusing her Ph.D. (CSE) from Siksha O Anusandhan University, Bhubaneswar. Her research area includes Data mining, Bioinformatics. She has published fifteen papers in International conferences and Journals.

Debahuti Mishra is working as Associate Professor in the department of Computer Applications at Institute of Technical Education and Research (ITER), Siksha O Anusandhan University, Bhubaneswar, Odisha, India. She has awarded Ph.D. (CSE) degree from Siksha O Anusandhan University, Bhubaneswar in 2011. Her research area includes Data mining, Bioinformatics. She has published more than seventy papers in International conferences and Journals, four books in the field of Automata Theory and Computation, Data mining and Bioinformatics.