Variations of Support Vector Machine classification Technique: A survey

Hetal Bhavsar[1], Amit Ganatra[2]
Assistant Professor, Department of Information Technology, SVIT, Vasad,
Gujarat Technological University, Gujarat, India and Research Scholar at CHARUSAT, Changa[1]
Dean, Faculty of Technology and Engineering, CHARUSAT, Changa, Gujarat, India[2]

Abstract

The Support Vector Machine (SVM) technique is emerged as a machine learning method used for classification, highly efficient and effective in the field of various applications like pattern recognition, image processing, fraud detection, text categorization etc. Its accuracy, robustness and providing best classification function to distinguish between members of the two classes in the training data are the main advantages, but the disadvantages can't be ignored even. The memory requirement and computation complexity are the main disadvantage of it. Many techniques are developed to overcome these limitations which are broadly classified into decomposition based and variant based algorithms. Also, SVMs were originally developed to perform binary classification. However, applications of binary classification are very limited. Most of the classification problems involve more than two classes. A number of methods to generate multiclass SVMs from binary SVMs have been proposed and is still a continuing research topic. In this paper, we present the survey of such techniques and falls them into three groups. The decomposition based method: overcome memory limitation, variant based techniques: reduce the computational complexity, and multiclass based methods handle the multi class classification.

Keywords

Classification, SVM, computational complexity, multiclass classification

1. Introduction

SVM have attracted a great deal of attention in the last decade and actively applied to various domains applications. SVMs are typically used for learning classification, regression or ranking function. SVM are based on statistical learning theory and structural risk minimization principal and have the aim of determining the location of decision boundaries also known as hyperplane that produce the optimal separation of classes [1][2][3]. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error [8].

SVM based classification is attractive, because its efficiency does not directly depend on the dimension of classified entities. Though SVM is the most robust and accurate classification technique, there are several problems. The data analysis in SVM is based on convex quadratic programming, and it is computationally expensive, as solving quadratic programming methods require large matrix operations as well as time consuming numerical computations [4]. Training time for SVM scales quadratically in the number of examples, so researches strive all the time for more efficient training algorithm[5], resulting in several variant based algorithm.

SVM can also be extended to learn non-linear decision functions by first projecting the input data onto a high-dimensional feature space using kernel functions and formulating a linear classification problem in that feature space [4]. The resulting feature space is much larger than the size of dataset which are not possible to store in popular computers. Investigation on this issues leads to several decomposition based algorithms. The basic idea of decomposition method is to split the variables into two parts: set of free variables called as working set, which can be updated in each iteration and set of fixed variables, which are fixed at a particular value temporarily. This procedure is repeated until the termination conditions are met [5].

Originally, the SVM was developed for binary classification, and it is not simple to extend it for multi-class classification problem. The basic idea to apply multi classification to SVM is to decompose the multi class problems into several two class problems that can be addressed directly using several SVMs [6]. Investigation on this issues leads to several multiclass based algorithms.

2. Decomposition Based Algorithms

The memory requirement of SVM is grows with the squares of number of training examples. So, the issue is can we scale up the algorithm for large dataset containing thousands and millions of instances [1].

Decomposition based methods break a large optimization problem into a series of smaller problems, where each problem only involves a couple of carefully chosen variables so that the optimization can be done efficiently. The following table 1 provides a brief description of some well known methods implemented to solve scaling problem [5].

Table 1: Comparative of Decomposition based SVM algorithms

Sr. No.	Algorithm	Key Idea	Advantages	Disadvantages
1	SVM^{light} [9]	many SVM learning problems have much less support vectors than training examples as well as many support vectors which have α_i at the upper bound C.	1. lower the training time with computational improvement like caching 2. efficient for large scale problem, especially for those with small support vector and most of their αi at the upper bound C.	
2	SMO (Sequential Minimal Optimization) [10]	Considers working set of size 2 in each iteration	1. Each sub problem can be solved analytically without invoking other solvers, thereby convergence is accelerated 2. SMO and its improved versions are effective for large scale SVM training	1. Inefficient due to the use of single threshold value. 2. Very slow for linear SVM
3	Alpha Seeding[11]	Speeding up SVM training by adapting alphas from previous training into appropriate seeds for the next training	1. Training cost is linear in the size of dataset.	• suited particularly for determining penalty coefficient and parameters in kernel using Leave-one-out-cross-validation estimations.
4	LIBSVM [12]	Add $b^2/2$ to the objective function	1. Due to no equality constraint, easier to deal with its dual bound constrained problem 2. comparable with SVM^{light} in terms of number of support vectors, the error rate and optimal value of objective function	Inefficient in performance
5	LASVM [13]	SMO sequential direction search is reorganised	1. Uses less memory 2. Significantly faster than state-of-the-art SVM solver 3. Gracefully handles noisy data 4. Converges to known SVM solution	Make equal number of process and reprocess iterations which does not guarantees optimal proportion

3. Variant Based Algorithms

Decomposition methods tackle only memory issue by splitting problem into a series of smaller ones, but they are time consuming for large scale problems. A number of methods to reduce the training time have been proposed at the price of accuracy, summarized in table 2 [5].

Table 2: Comparative of Variant Based SVM algorithms

Sr. No.	Algorithm	Key Idea	Advantages	Disadvantages
1	LS-SVM (least squared svm) [14]	based on Conjugate gradient scheme	Due to the equality constraints in the formulation, a set of linear equations has to be solved instead of a quadratic programming problem.	suitable for small dataset
2	LSVM (Lagrangian svm) [15]	Used for linear classification	1. Objective function is strongly convex and equality constrain disappear in its dual. 2. Capable of classifying data sets with millions of data in several minutes much faster than SMO and SVM[light] if the dimension of input space is small. (less than 100). 3. Better generalization capability compared with above methods.	Not able to scale up for very large problems
3	PSVM (Proximal SVM) [16]	Classifies points by assigning them to the closer of two parallel planes that are pushed apart as far as possible.	1. Allows to handle very large datasets. 2. Comparable with standard SVM in performance but fast by several orders of magnitude.	Suited for linear kernel SVM
4	RSVM (Reduced SVM) [17]	Randomly preselect a subset of m examples as support vector candidates	useful for larger problems as well as problems with many support vectors	Remark: Designed for large scale nonlinear kernel SVM.
5	LP-SVM [18]	Changing the metric of margin from 2-norm to 1-norm	Reduce number of dimensions	Convergence rate is similar to simple SVM

4. Multiclass Based Algorithms

Originally, SVMs were developed to perform binary classification. However, applications of binary classification are very limited especially in remote sensing land cover classification where most of the classification problems involve more than two classes. A number of methods to generate multiclass SVMs from binary SVMs have been proposed. The following table 3 provides a brief description of some methods implemented to solve multi-class classification problem with SVM [6][7].

Table 3: Comparative of Multiclass Based Algorithms

Sr. No.	Algorithm	Key Idea	Advantages	Disadvantages
1	OVA (one-Against-all) [8]	With the k classes, k binary problems are classified, where each problem discriminates a given class from the other k-1 classes.	Simple, provide comparable performance with other complicated approach when binary classifier is tuned well.	1. Training complexity is high, as the number of training samples are large 2. Memory requirement is very high during training phase
2	OVO (one-against-one) [19]	Binary classifier requires discriminating between each pair of classes, requiring k (k-1)/2 binary classifiers.	1. Memory required for kernel matrix is smaller 2. Better than OVA approach Shorter training time	Slower in testing especially when number of classes is big as every test sample has to be presented to large number of classifier

3	DAGSVM [20]	Same idea as OVO, and in recognition phase, the algorithm depends on a rooted Binary DAG to make a decision	Faster Testing and achieving similar recognition rate as OVO	Memory requirement and accuracy is similar to OVA and OVO
4	Error Correcting Codes [21]	Based on idea of error correcting code for neural network	Improve generalization ability	
5	MSVM (Multistage SVM) [22]	Used the support vector Clustering to divide the training data into two parts	Better Generalization capability	Controlling support vector clustering to divide the training dataset into exactly two classes is painful and unfeasible for large datasets
6	HSVM (Hierarchical svm) [23]	Based on clustering classes into a binary tree	Improve performance	1. Knowledge transfer problem 2. how to evaluate stopping criteria for mixed class samples
7	BTS (Binary tree of SVM) [24]	multiple SVMs arranged in a binary tree structure	Testing time is better than OVO and OVA	Require testing of each trained SVM with all the training samples in order to determine the next step, which significantly increasing the total training time.
8	SVM-BDT [6]	Multiple SVMs arranged in binary structure and based on efficient computation of tree architecture and the high classification accuracy of svm. K-1 svms needed for k class problem	1. During the recognition phase due to its logrithmatic complexity, it is much faster that widely used OVA and OVO methods. 2. more favourable as the no. of classes increases	

5. Conclusion

We compared the different Support Vector Machine techniques. Some of them are decomposition base, some are variant base and some are multi classification base. All the three kinds of techniques are improvements over basic SVM techniques. Improvements are proposed by researchers to gain speed efficiency, space efficiency and ability to handle multiple classes. Every technique hold good in particular field under particular circumstances. The future work will concentrate on analysing the performance of different kernel function on different application.

References

[1] J. Han and M. Kamber, "Data Mining Concepts and Techniques", Elevier, 2011.

[2] V. Vapnik and C. Cortes , "Support Vector Network," Machine Learning, vol. 20, pp. 273-297, 1995.

[3] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, 1998.

[4] H. Bhavsar, A. Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.

[5] G. Wang, "A Survey on Training Algorithms for Support Vector Machine Classifiers", Fourth International Conference on Networked Computing and Advanced Information Management, 2008, IEEE.

[6] G Madzarov, D. Gjorgievikj and I. Chorbev, " A Multi-class SVM Classifier Utilizing Binary Decision Tree", Informatica, pp. 233-241, 2009.

[7] M. Aly, "Survey on Multiclass Classification Methods", November 2005.

[8] V. Vapnik, "Statistical Learning Theory", Wiley, New York, 1998.
[9] T.Joachims, "Making large-scale support vector machine learning practical", In Advances in Kernel Methods: Support Vector Machines, 1999.
[10] J.Platt, "Fast training of SVMs using sequential minimal optimization", In B. Sch¨olkopf, C.Burges and A.Smola (ed.), Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 1999, 185-208.
[11] D. DeCoste and K. Wagstaff, "Alpha seeding for support vector machines", In Proceedings of International Conference on Knowledge Discovery and Data Mining (KDD-2000), 2000.
[12] C.C.Chang and C.J.Lin, LIBSVM: Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[13] A.Bordes, S. Ertekin, J. Weston, L. Bottou, "Fast kernel classifiers with online and active learning", Journal of Machine Learning Research (JMLR), 6, 2005, 1579- 1619.
[14] J.Suykens and J.Vandewalle, "Least squares support vector machine classifiers", Neural Processing Letters, 1999, 293-300.
[15] O.L.Mangasarian and D.R.Musicant, "Lagrangian support vector machines", Journal of Machine Learning Research, 1, 2001, 161-177.
[16] G. Fung and L.Mangasarian, "Proximal support vector machine classifiers", Proceedings of the 7th ACM conference on knowledge discovery and data mining, ACM, 2001, 77-86.
[17] Y.J.Lee and O.L.Mangasarian, "RSVM: Reduced support vector machines", In Proceedings of the First SIAM International Conference on Data Mining, 2001.
[18] O.L.Mangasarian, "A finite Newton method for classification, Optimization Methods and Software", 17(5), 2002, 913-929.
[19] J. H. Friedman. Another approach to polychotomous classification. Technical report. Department of Statistics, Stanford University, 1997.
[20] Platt, N. Cristianini, J. Shawe-Taylor. Large margin DAGSVM's for multiclass classification. Advances in Neural Information Processing System. Vol. 12, pp. 547–553, 2000.
[21] T. G. Dietterich and G. Bakiri , "Solving Multiclass learning problems via error correcting output code", Journal of Artificial Intelligence Research, 39:1-39, 1995.
[22] X. Liu, H. Xing, X. Wang, "A multistage support vector machine", 2nd International Conference on
Machine Learning and Cybernetics, pages 1305- 1308, 2003.
[23] Y.Chen, M. Crawford, and J. Ghosh, "Integrating support vector machines in a hierarchical output space decomposition framework", In Proceedings of Geosciences and Remote Sensing Symposium, volume 2, pages 949–952, 2004.
[24] B. Fei, J. Liu, "Binary Tree of SVM: A New Fast Multiclass Training and Classification Algorithm."IEEE Transaction on neural networks, Vol. 17, No. 3, May 2006.

Ms. Hetal B. Bhavsar (B.E.-'99, M.E. – '08, Ph.D.* '12) has completed her B.E from M.S. University, M.E. in Computer Science with second rank from D. D. University, Gujarat, India. She is pursuing her Ph.D in Classification Techniques in Data Mining under the guidance of Dr. Amit P. Ganatra from CHARUSAT, Changa. She is presently working as an Assistant Professor at SVIT, Gujarat Technolgical University. She has published a number of papers in the proceedings of National and International level conferences and journals. She is a life member of Computer Society of India (CSI) and ISTE.

Amit P. Ganatra (B.E.-'00-M.E. '04-Ph.D. '12) has received his B.Tech. and M.Tech. degrees in 2000 and 2004 respectively from Dept. of Computer Engineering, DDIT-Nadiad from Gujarat University and Dharmsinh Desai University, Gujarat and Ph.D. in Information Fusion Techniques in Data Mining from KSV University, Gandhinagar, Gujarat, India He is a member of IEEE and CSI. His areas of interest include Database and Data Mining, Artificial Intelligence, System software, soft computing and software engineering. He has 11 years of teaching experience at UG level and concurrently 7 years of teaching and research experience at PG level. In addition he has been involved in various consultancy projects for various industries. His general research includes Data Warehousing, Data Mining and Business Intelligence, Artificial Intelligence and Soft Computing. He had published and contributed over 70 papers (Author and Co-author) published in referred journals and presented in various international conferences.