Research on Kernel Function of Support Vector Machine

Lijuan Liu, Bo Shen and Xing Wang

Abstract Support Vector Machine is a kind of algorithm used for classifying linear and nonlinear data, which not only has a solid theoretical foundation, but is more accurate than other sorting algorithms in many areas of applications, especially in dealing with high-dimensional data. It is not necessary for us to get the specific mapping function in solving quadratic optimization problem of SVM, and the only thing we need to do is to use kernel function to replace the complicated calculation of the dot product of the data set, reducing the number of dimension calculation. This paper introduces the theoretical basis of support vector machine, summarizes the research status and analyses the research direction and development prospects of kernel function.

Keywords Support vector machine · High-dimension data · Kernel function · Quadratic optimization

Introduction

Support vector machine (SVM) was introduced into the field of machine learning and its related area in 1992 [1], having received widespread attention of researchers in later time and has made great progress in many fields. It uses a

L. Liu · B. Shen
School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
e-mail: 11120126@bjtu.edu.cn

L. Liu · B. Shen (✉)
Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
e-mail: bshen@bjtu.edu.cn

X. Wang
China Information Technology Security Evaluation Center, Bejing, China
e-mail: wangx@itsec.gov.cn

Y.-M. Huang et al. (eds.), *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*, Lecture Notes in Electrical Engineering 260,
DOI: 10.1007/978-94-007-7262-5_93, © Springer Science+Business Media Dordrecht 2014

nonlinear mapping to map original training data into high-dimensional data space in order to find the optimal classification hyper plane separating those data belonging to different categories. Support vector machine is based on SLT (statistical learning theory) [2, 3] VC dimension theory and structural risk minimization principle. Compared with traditional neural networks, support vector machine gains great enhancement in generalization ability and overcomes some problems existing in feed-forward neural networks, such as local minimum and the curse of dimensionality [4]. The introduction of kernel function greatly simplifies the complexity of dot product operation in support vector machine for nonlinear data classification, and it can be used to distinguish and enlarge the useful features, and support vector machine based on kernel function is playing a powerful role in the field of data mining.

Support Vector Machine

If the training data set is linear separable, the given data set is $\mathrm{D}:\ (X_1, y_1)$, $(X_2, y_2), \ldots, \left(X_{|D|}, y_{|D|}\right)$, among which X_i is training data with class label y_i. The scope of each y_i is $+1$ or -1, namely $y_i \in \{+1, -1\}$. In dealing with classification problem, the optimal classification hyper plane can be denoted as follows:

$$W \cdot X + b = 0. \tag{1}$$

W is a weight vector, that is to say, $W = \{w_1, w_2, \ldots, w_n\}$, where w_i is the weight of X_i; n is the number of attributes; parameter b is a scalar, and is often referred to as the bias. The formula $W \cdot X$ stands for the dot product of W and X. From geometric point of view, the entire input space can be divided by hyper plane into two parts: one part is positive value data set; another part is the negative one. Hyper plane is a line in two-dimensional space, a surface in three-dimensional space. The biggest edge distance between two types of training data set is $\frac{2}{\|W\|}$. Support vector machine discovers the optimum classification hyper plane by means of support vectors and the edges between them [5] and gets the maximum edge distance of two classes of data sets at the same time.

Research on Kernel Function

For linear separable data, support vector machine can directly classify the data set into two categories in the input space; for those nonlinear separable data, SVM has to map the original input data X with nonlinear mapping $(\Phi : X \rightarrow F)$ into another high-dimensional space where we can solve the maximum interval of classification, and this new high-dimensional space is the feature space. A dot product operation can be directly substituted by kernel function in feature space, and we needn't know the concrete eigenvector and mapping function, which is also known as kernel trick. Frequently-used kernel functions include the following ones:

$$\text{Linear kernel function: } K(X_i,\ X_j) = X_i \cdot X_j. \tag{2}$$

$$\text{Polynomial kernel function: } K(X_i,\ X_j) = (X_i \cdot X_j + 1)^h. \tag{3}$$

$$\text{Gaussian radial basis function kernel function: } K(X_i,\ X_j) = \ell^{-\left\|X_i - X_j\right\|^2/2\sigma^2}. \tag{4}$$

$$\text{Sigmoid kernel function : } K(X_i, X_j) = \tanh(\kappa X_i \cdot X_j - \delta). \tag{5}$$

The performance of support vector machine mainly depends on model selection, including the selection of the kernel function type and the kernel parameter selection [6]. In the study of kernel function selection, the kernel alignment is a good method [7]. Kernel alignment method is based on a hypothesis: the kernel matrix of a good kernel function should be as similar as possible to the calibration matrix. Under normal circumstances, the first consideration of choosing kernel function is the Gaussian radial basis function kernel function (RBF), which is because RBF has fewer parameters to select, and what's more, for some parameters, RBF has similar performance to the Sigmoid kernel function.

The development and application of support vector machine have been greatly promoted since the introduction of kernel function, and its application area has extended from hand-written numeral recognition, reference time series prediction test and other traditional application area to new areas such as information image processing [8], industrial process control, etc. The following content in this paper will center on the discussion of kernel function in support vector machine and put forward its future research directions.

Kernel Clustering

Clustering analysis divides data objects into different subsets and the data objects in the same subset are similar to each other, while those located in different subsets have different properties. Kernel clustering combines kernel function and clustering together, which is based on the characteristics of clustering [9]. In the first stepwise, kernel clustering clusters the training data and test data, and then constructs the kernel function on the basis of clustering results. Chapell et al. [10] proposed an overall framework of constructing kernel clustering, using different conversion functions to change the eigenvalue decomposed by kernel matrix. Jason Weston et al. came up with the bagged clustering kernel [10] to overcome the time complexity problem existing in Chapell's clustering kernel. Although the bagged clustering kernel shortened the kernel clustering time, there still is much room for improvement in terms of classification accuracy.

Fuzzy C-means (FCM) clustering algorithm [11] introduces fuzzy set theory to the process of clustering. Fuzzy kernel clustering algorithm firstly maps the data of input space into high-dimensional feature space to enlarge pattern differences between classes, and then carry through fuzzy clustering in the feature space [12, 13]. Fuzzy kernel clustering algorithm is able to highlight the differences of different sample characteristics, increasing clustering accuracy and speed [14]. It has always been an important research problem to expand SVM classifier to multi-class classification [15, 16], so Zhao et al. [17] applied fuzzy kernel clustering to multi-class classification method, solving the serious problem of fuzzy overlapping, but didn't have a large increase in classification speed.

The objective function of the fuzzy kernel clustering algorithm is as follows:

$$J_n(U,\ v) = \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij}^{n} \left\| \Phi(x_j) - \Phi(v_i) \right\|^2. \tag{6}$$

In the above formula, parameter c is the clustering number initially set; v_i is the initialized clustering center; u_{ij} is the membership function of sample j belonging to i category; $U = \{u_{ij}\}, v = \{v_1, v_2, \ldots, v_c\}$, parameter n is weighted index, and $n > 1$. The criterion of fuzzy clustering algorithm is for the minimum value of the above objective function until each membership value stabilized.

Kernel clustering support vector machine has been successfully utilized in biomedical, text classification [18] and many other application fields, undoubtedly, it will involve a wider range of application fields in later time.

Super-Kernel Function

Support vector machine for data classification often brings about two dilemmas using one single kernel function: one is unable to complete effective nonlinear mapping; the other is over-fitting or under-fitting [19]. The extrapolation ability of Gaussian radial basis function kernel function weakens along with the increase of σ parameter [20], so it has strong locality. Polynomial kernel function regulates different mapping dimensions through adjusting h parameter, and computation grows with h parameter, thus having strong global property and poor locality.

In order to adapt to the increase on data set and high efficiency requirements, many algorithms have been developed, such as large data set training [21] algorithm, super kernel learning [22] algorithm, fast convergence algorithm [23], etc. Combining several kernel functions of different categories with a polynomial composition can give full play to the excellent characteristics of different kernel functions in dealing with data classification, and overcomes the shortage of single kernel function while maintaining translational invariance and rotation invariance, which provides a fresh effective way to studying the construction of kernel function in support vector machine. A simple form of super-kernel function is like the following formula:

$$K(X_i, X_j) = \beta_1 \ell^{-\left\|X_i - X_j\right\|^2/2\sigma^2} + \beta_2 (X_i \cdot X_j + 1)^h. \qquad (7)$$

It is the linear combination of Gaussian radial basis function kernel function and h polynomial kernel function. Super-kernel function parameters were regulated in the form of parameter vector, that is to say, super-kernel function adjusts all parameters at the same time, and those several more parameters just add to the length of parameter vector but not affect the determination time of parameters.

Kernel Parameter Selection

Selection of kernel parameter has a direct impact on the performance of the SVM classifier. A commonly used parameter selection method is based on the Generalization Error estimates [24]. The Generalization Error estimates predicate and forecast the generalization capability of classification decision criteria by means of training data set [25]. Vapnik et al. [26] estimate generalization ability by span of support vectors on the basis of Generalization Error estimates, which has an advantage of higher accuracy but has more complex computation. A method of utilizing data set to evaluate the optimal kernel parameter method [27] is put forward on the basis of Vapnik's method, determining the optimal kernel parameter choice from a geometric point of view.

Qi et al. [28] proposed a kernel parameter selection method to solve LOO (leave-one-out) upper bound minimum point based on genetic algorithm, where we can choose the reproduction operator and combine genetic algorithm with steepest descent method, improving the accuracy of forecast but without leading to local optimal solution. Chen et al. [29, 30] adopted different generalization ability estimates as the fitness function of genetic algorithm, and it not only reduced the computation time to choose parameters but also the dependence on the initial value.

Parameter selection method based on kernel matrix similarity measure starts from research on kernel matrix in order to search for the optimal kernel parameter and learning model, and this method improved the calculation speed of SVM. Liu Xiangdong et al. [31] eventually found the optimal kernel parameters and the kernel matrix by means of experimenting on UCI standard data set and FERET standard faces library. Parameter selection method based on kernel matrix similarity measure can serve as a feasible method to choose the optimal SVM model, and it also has certain reference value for choosing other kernel parameters.

Conclusion

The study of kernel function of SVM is an important data mining research, so choosing the appropriate kernel function and its parameters can give full play to the performance of SVM and even has remarkable significance in promoting the popularization and application of data mining. This paper does research on the kernel function of SVM and does some summary comments on the kernel clustering, super kernel function and the selection of kernel parameters. Judging from the current study, the author believes that the study of kernel function in the following areas is to be further developed:

1. Finishing data mapping efficiently and reliably in the environment of big data. "Big data" has features of giant, high growth rate and diversification, and it needs new processing mode to excavate useful information from the massive data and get an insight in them. In this case, the conventional transformation of kernel function will face with new bottleneck in processing speed and processing quality. On the basis of existing research, it remains further study to extend the scale of the expansion of kernel function processing data and select the appropriate kernel parameters and further improve the quality and speed of processing data.
2. Giving full play to the advantages of different kernel functions in super-kernel functions. On the perspective of present research on the selection of kernel parameters, Gaussian radial basis function (RBF) and its parameter selection have been more detailed studied in the field by virtue of its favorable advantages in computer vision. How to expand the application area of polynomial kernel function and Sigmoid kernel function, especially how to give full play to the advantage of each kernel function in the super-kernel function still needs deeper exploration. It is worth studying that selecting and optimizing super-kernel function parameters and applying the concept of constructing super-kernel function to support vector machines.
3. Selecting appropriate kernel function of support vector machines for specific applications. The scope of data mining processing data is developing from structured data to the direction of semi-structured and unstructured data. As one part of the data mining classification algorithms, the application fields of SVM continue to expand, thus it appears particularly important to select the appropriate kernel functions of specific domain. The choice of kernel function is closely related to the data field [32], in the meantime, the performance of kernel function depends largely upon the selection of parameters. Future studies are required to determine the kernel function and its parameters according to different application areas of the support vector machine for the sake of reducing the consuming of storage space and computing time of computers.

Big data processing is the future research tendency, and with the arrival of the cloud era, big data is attracting more and more attention. In future work, the author will mainly focus on the study of achieving efficient and accurate classification with support vector machine in the environment of big data.

Acknowledgments This work has been supported by the National Natural Science Foundation of China under Grant 61172072, 61271308, and Beijing Natural Science Foundation under Grant 4112045, and the Research Fund for the Doctoral Program of Higher Education of China under Grant W11C100030, the Beijing Science and Technology Program under Grant Z121100000312024.

References

1. Boser B, Guyon I, Vapnik V (1992) A training algorithm for optimal margin classifiers [C]. In: Proceedings of the 5th annual ACM conference on computational learning theory, Pittsburgh, pp 144–152
2. Vapnik VN (2000) The nature of statistical learning theory [M]. Translated by Zhang Xuegong (trans: Zhang X). Tsinghua University Press, Beijing
3. Vapnik VN (2004) Statistical learning theory [M]. Translated by Xu Jianhua, Zhang Xuegong (trans: Xu J, Zhang X). Publishing House of Electronics Industry, Beijing
4. Vapnik V (1995) The nature of statistical learning theory [M]. Springer, New York
5. Ju C, Guo F (2010) A distributed data mining model based on support vector machines DSVM [J]. 30(10):1855–1863
6. Zhu S, Zhang R (2008) Research for selection of kernel function used in support vector machine [J]. Sci Technol Eng 8(16):4513–4517
7. Cristianini, N, Taylor Shawe J, Kandola J et al. (2002) On kernel target alignment. In: Proceedings neural information processing systems. MIT Press, Cambridge, pp 367–373
8. Yang Z (2008) Kernel-based support vector machines [J]. Comput Eng Appl 44(33):1–6
9. Li T, Wang X (2013) A semi-supervised support vector machine classification method based on cluster kernel [J]. Appl Res Comput 30(1):42–45
10. Tison C, Nicolas JM, Tupin F et al. (2004) A new statistical model for Markovian classification of urban areas in high-resolution SAR images [J]. IEEE Trans Geosci Remote Sens 42(10):2046–2057
11. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York
12. Wu Z, Gao X, Xie W (2004) A study of a new fuzzy clustering algorithm based on the kernel method. Journal of Xi 'an university of electronic science and technology magazine, 31(4):533–537
13. Zhang N, Zhang Y (2010) Support vector machine ensemble model based on KFCM and its application [J]. J Comput Appl 30(1):175–177
14. Cao W, Zhao Y, Gao S (2010) Multi-class support vector machine based on fuzzy kernel clustering [J]. CIESC Journal 61(2):420–424
15. Angulo C, Parra X (2003) K-SVCR Andreu Catala. A support vector machine for multi-class classification [J]. Neurocomputing 55(9):55–77
16. Platt JC, Cristianini N, Shawe-Taylor J (2000) Large margin DAGs for multiclass classification [J]. Adv Neural Inf Proc Syst 12(3):547–553
17. Zhao H, Rong L (2006) SVM multi-class classification based on fuzzy kernel clustering [J]. Syst Eng Electron 28(5):770–774
18. Yang Z (2008) Research progress of the kernel function support vector machine [J]. Sci Technol Inf 19:209–210
19. Jia L, Liao S (2008) Support vector machines with hyper-kernel functions [J]. Comput Sci 35(12):148–150
20. Guo L, Sun S, Duan X (2008) Research for support vector machine and kernel function [J]. Sci Technol Eng 8(2):487–489
21. Collobert R, Bengio S (2001) SVM torch: support vector machines for large-scale regression problems. J Mach Learn Res 1:143–160
22. Cheng SO, Smola AJ, Williamson RC (2005) Learning the kernel with hyper-kernel. J Mach Learn Res 6:1043–1071
23. Platt J, Burges CJC (1998) Fast training of support vector machines sequential minimal optimization. In: Sholkpof B, Smola AJ (eds) MIT Press, Cambridge
24. Tao W (2003) Kernels' properties, tricks and its application on obstacle detection [J]. National University of Defense Technology, Changsha
25. Y Fu, D Ren (2010) Kernel function and its parameters selection of support vector machines [J]. Sci Technol Innov Herald 9:6–7

26. Chapelle O, Vapnik V, Bousquet O et al (2002) Choosing multiple parameters for support vector machines [J]. Mach Learn 46(1):131–159
27. Men C, Wang W (2006) Kernel parameter selection method based on estimation of convex [J]. Comput Eng Des 27(11):1961–1963
28. Qi Z, Tian Y, Xu Z (2005) Kernel-parameter selection problem in support vector machine [J]. Control Eng China 12(44):379–381
29. Chen PW, Wang JY, Lee HM (2004) Mode selection of SVNs using GA approach [C]. Proceedings of 2004 IEEE international joint conference on neural networks. IEEE Press, Piscataway, pp 2035–2040
30. Zheng CH et al. (2004) Automatic parameters selection for SVM based on GA [C]. Proceedings of the 5th World congress on intelligent control and automation, IEEE Press, Piscataway, pp 1869–1872
31. Liu X, Luo B, Qian Z (2005) Optimal model selection for support vector machines [J]. J Comput Res Dev 42(4):576–581
32. Wang T, Chen J (2012) Survey of research on kernel selection [J]. Comput Eng Des 33(3):1181–1186