Research Survey on Support Vector Machine

Huibing Wang
Faculty of Software, Fujian
Normal University
Fuzhou, China 350108
wanghbfjnu@163.com

Jinbo Xiong
Fujian Engineering Research
Center of Public Service Big Data
Mining and Application
Fuzhou, China 350108
jinbo810@163.com

Zhiqiang Yao*
Fujian Engineering Research
Center of Public Service Big Data
Mining and Application
Fuzhou, China
zyy837603010g@163.com

Mingwei Lin
Fujian Engineering Research
Center of Public Service Big Data
Mining and Application
linmwcs@163.com

Jun Ren
Faculty of Software, Fujian
Normal University
jun_ren@outlook.com

ABSTRACT

Support vector machine (SVM) as a learning machine has shown a good learning ability and generalization ability in classification, regression and forecasting. Because of its excellent learning performance, SVM has always been a hotspot in machine learning. In hence, this paper will make a more systematic introduction of SVM, including the theory of SVM, the summarization and comparison of the quadratic programming optimizations and parameter optimizations for SVM and the introduction of some new SVMs, like FSVM, TSVM, MSVM, etc. Then, the applications of SVM in real life will be presented, especially, in the area of mobile multimedia. Finally, we conclude with a discussion of the direction of further SVM improvement.

KEYWORDS

SVM, mobile multimedia, optimization algorithms, SVM extension, application

*Corresponding author.

1 INTRODUCTION

SVM was first proposed by Vapnik in 1995 [1] and it was named as "support-vector networks" in the early time, which is a binary classification algorithm and implements the idea of mapping non-linearly vectors to a very high-dimension feature space to construct a linear decision surface (hyperplane) in this feature space. In hence, it can achieve good effect on solving separable and non-separable problems. This hyperplane is optimal in the sense of being a maximal margin classier with respect to the training data [2]. The Structural Risk Minimisation (SRM) principle, SVM follows, equips SVM with a greater ability to generalise. Corresponds to such significant advantages, SVM was applied to classification and regression problems quickly [3]. Even more, SVM embodies the characteristics of small samples, nonlinearly problem and "curse of dimensionality", for which SVM has always been the concern of many research scholars. As for application, it has been widely used in all areas in daily life, such as economic field, transportation field, medical field and so on. Except for these traditional areas, SVM also is used for new hot areas, like mobile multimedia.

Mobile multimedia refers to various types of content that are either accessed via portable devices or created using them. One of the areas that mobile multimedia has become ubiquitous is in smartphones that incorporate video and music playback capability [4], cameras, and wireless content streaming [5, 6]. It refers to many salient technology dimensions, like Media Storage, Retrieval and Synchronization [7]; Imaging, Display, Human-Machine Interface in Terminals; Multimedia Applications and Services. Successfully deploying multimedia services and applications in mobile environments requires adopting an interdisciplinary approach where

multimedia [8]. We also need to put a great effort in designing applications that take into account the way the user perceives the overall quality of the provided service. SVM has been applied to many areas of mobile multimedia. Thus, this article will focus on SVM applications in the field of mobile multimedia.

The remainder of the paper is organized as follows. Section 2 introduces the principle of SVM from three cases. In Section 3, some algorithm optimizations and parameter optimizations about SVM will be represented and compared. Then, Section 4 includes some popular extensions of SVM like FSVM, TWSVM and MSVM. The applications of SVM in many popular areas, especially, the area of mobile multimedia, will be presented in Section 5. Finally, in Section 6, there is a conclusion referring to the direction of the further SVM improvement.

2   THEORY OF SUPPORT VECTOR MACHINE

SVM shows its significant advantages on both separable problems and non-separable problems. Separable problems include linear separable problems and non-linear separable problems. Thus, this section will present the theory of SVM from three cases.

Case of Linearly Separable

Classification problems originated from the linearly separable situation, the method of SVM was applied to linearly separable problems at first [3]. Classification results often can bring some valuable information for further prediction or analysis. In hence, people are committed to using effective classification algorithms to get valuable results. For instance, say we are offered some training data with some people's weight, height and their corresponding sex, then we want to make use of them to predict the unknown gender data [2].

As Figure 1 shows, these two types of points represent male and female, respectively. It can be seen that there are a number of lines we can draw to divide the space into two regions in Figure 2. And it's easily to find that the black solid line would be the optimal line, which maximizes the margin between itself and the nearest points of each class.

SVM extends the two-dimensional linear separable problem to multidimensional, and aims to seed the optimal classification surface, also called as optimal hyperplane. The optimal hyperplane can be defined as:
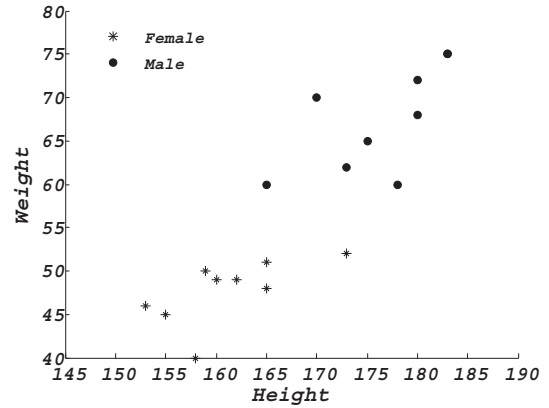
$$w^T x + b = 0, \tag{1}$$
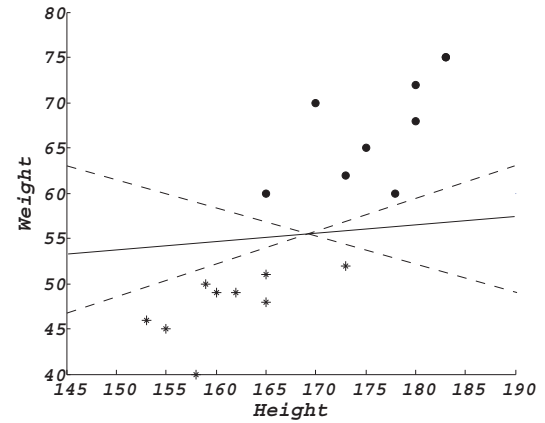

Figure 1: Points of Males and Females

Figure 2: Classification Lines.

in which w refers to the weight vector and b is the threshold. Thus, the relation between x_i and $f(x_i)$ can be defined as:

$$f(x) = w^T x + b. \tag{2}$$

Seeding the optimal hyperplane is equivalent to maximal the distance between the closest vectors to the hyperplane [1]. Define the Euclidean distance between the nearest points and the hyperplane $f(x)$ as:

$$r = |\frac{f(x)}{||w||}|, \tag{3}$$

where $f(x)$ refers to functional margin.

Assume the functional margin $f(x)$ between the nearest points and the hyperplane is 1 as the Figure 3 shows. The assumption was accompanied with a constraint condition $y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, n$, which implies that all training data are on the two hyperplane or behind them and the training data on the hyperplane
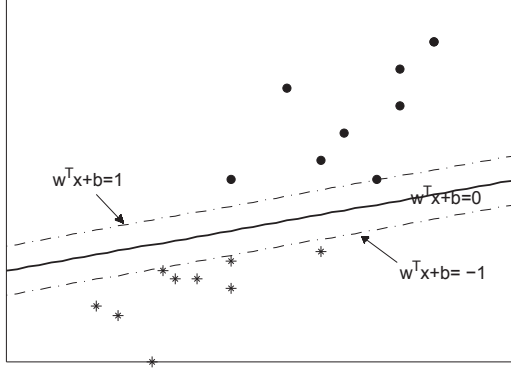
Figure 3: Optimal Classification Line.

Figure 4: Case of Non-linearly Separable.

are called support vectors (SVs). Thus, the margin between the parallel bounding planes d can be defined as: $d = 2r = \frac{2}{||w||}$.

Thus the objective function can be presented as:

$$max\frac{2}{||w||} \quad s.t. \ y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, n. \quad (4)$$

For ease of calculation, translate equation (4) into:

$$min\frac{1}{2}||w||^2 \quad s.t. \ y_i(w^T x_i + b) \geq 1, i = 1, 2, \ldots, n. \quad (5)$$

The solution to this optimization problem is given by the saddle point of the Lagrange functional:

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i[y_i(w^T x_i + b)] \quad (6)$$

where α_i are Lagrange multipliers, subject to $\alpha_i \geq 0$, and equation (6) has to be minimized with respect to w, b and maximized with respect to α_i. The objective function has been translated as:

$$p^* = min_{(w,b)}\theta(w) = min_{w,b}max_{(\alpha_i \geq 0)}L(w, b, \alpha), \quad (7)$$

The classical Lagrange duality enables the primal problem to be transformed into its dual problem, when satisfying the Karush-Kuhn-Tucker (KKT) condition. In hence, as for equation (7), when it fulfills the KKT condition of $\sum_{i=1}^{n} \alpha_i[y_i(w^T x_i + b) - 1] = 0$, equation (7) is equal to its dual function:

$$d^* = max_{(\alpha_i \geq 0)}min_{w,b}L(w, b, \alpha). \quad (8)$$

It is the KKT condition that determines only the SVs work on the extremum of the objective function. Then the minimum with respect to w and b of the Lagrange,
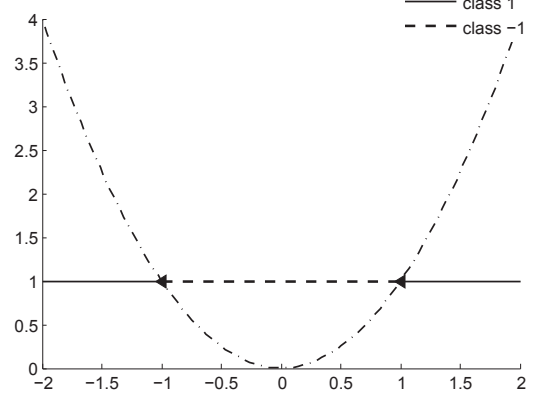
L, is given by,

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{n} \alpha_i y_i x_i,$$
$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0. \quad (9)$$

Hence from equations (6), (9) and the KKT condition, the dual problem d^* can be changed into:

$$\theta(\alpha) = L(w, b, \alpha) = \sum_{i=1}^{n} \alpha_i + \frac{1}{2}\sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i)^{\mathrm{T}} x_j, \quad (10)$$

which is a two convex optimization problems.

Supposing that $\overline{\alpha}_i$ is the result of equations (10), X_r is a support vector belonging to the positive class, while X_s belongs to the negative class. Then the optimal solution w and b can be expressed as follows:

$$\overline{w} = \sum_{i=1}^{n} \overline{\alpha}_i y_i x_i, \quad \overline{b} = -\frac{1}{2}\overline{w}[X_r + X_s]. \quad (11)$$

Hence from equations (2), (9) and (11), the final decision function is:

$$f(x) = sgn[\sum_{SVs} \overline{\alpha}_i y_i (x_i)^{\mathrm{T}} x + \overline{b}]. \quad (12)$$

Case of Non-linearly Separable

In addition to the linearly separable problems, there are much more nonlinearly cases in real life. As the example showed in Figure 4, there is no line can classify the two classes well, only curves.

SVM shows its superiorities for nonlinearly problems, which takes the way of mapping the input vectors in low dimension feature space to a high dimension feature space to hunt for a linear optimal separating hyperplane.

And it has been demonstrated that it did not account for large storage and computational expense [2].

The following is an example of mapping process. Define of the curve in two-dimensional space as:

$$g(x) = C_0 + C_1 x + C_2 x^2.$$

Map it into a four-dimensional space as follows:

$$y = \begin{bmatrix} y_3 \\ y_2 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}, \qquad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} C_0 \\ C_1 \\ C_2 \end{bmatrix}.$$

Then, the primal function is translated into:

$$g(x) \Rightarrow f(x) = <a, y> = ay.$$

It's a linear function, which means the primal nonlinearly problem has been converted to a linearly separable problem. SVM employs the empathy, mapping the w and x into high dimensional space, then gets new w' and x'.

$$f(x) = <w, x> +b \quad \Rightarrow \quad g(x') = <w', x'> +b.$$

In hence, new decision function can be defined as follows:

$$g(x') = \sum_{i=1}^{n} \alpha_i y_i <x_i', x_i> +b, \qquad (13)$$

where dot product is core operation. According to the existence theorem of kernel function: for any sample set K, there is a map of F, and in this map, $f(k)$ (in the high dimensional space) is a linear separable. Based on this theorem, SVM employs the kernel function during mapping process, and realizes the dot product in low dimensional space, which is supposed to be done in high dimensional space. Decision function with kernel function is shown as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K <x_i, x> +b. \qquad (14)$$

Different kernel functions can be choose to construct different types SVM, the following are a few commonly used kernel functions:
(1) Polynomial

$$K(x, x_i) = (x \cdot x_i)^q \quad q = 1, 2, \ldots; \qquad (15)$$

(2) Gaussian Radial Basis Function

$$K(x, x_i) = exp(-\frac{(x - x_i)^2}{2\sigma^2}); \qquad (16)$$

(3) Multi-Layer Perceptron

$$K(x, x_i) = tanh(b(x \cdot x_i) - c). \qquad (17)$$

Case of Non-separable

Last circumstance, some points belong to positive class may be classified into negative class. Another way to say, there is no hyperplane can separated the points of different categories accurately. Usually these error points are regarded as noise which can be ignored. However, machines can't deal with the error points as human do. Thus, Cortes [1] introduced slack variables $\xi_i \geq 0$ (a measure of the misclassification error) and punishment coefficient C (the emphasis on the outlier) to offset the effect of noise. And the constraint $y_i(w^T \cdot x_i + b) \geq 1, i = 1, 2, \ldots, n$ is modified to:

$$y_i(w^T \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, n.$$

C is a given value and subjects to above constraint. Too small ξ may lead to over learning, while too large may lead to owe learning. The equation (5) and equation (6) are changed into:

$$\Phi(x, \xi) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i, \qquad (18)$$

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i[y_i(w^T x_i + b) - (1 - \xi)]. \qquad (19)$$

While after derivation, the decision function for this case still is equation (12).

In summary, SVM is a powerful classifier, which is suitable for any case of classification with the same decision function, high classification accuracy and small computation.

3   THE OPTIMIZATION OF SVM ALGORITHM

SVM shows its superiorities in many ways, however, the high classification accuracy and small computation of SVM give the credit to well treatments of quadratic programming problem and parameters selection. Thus, this section will introduce some optimization algorithms for SVM.

Algorithms for Quadratic Programming Optimization

For the implementation point of view, training a SVM is equivalent to solving a linearly constrained Quadratic Programming (QP) problem [11]. QP problem is challenging when the size of the data points exceeds few thousands, which is often the case in practical applications. Thus, much researches for the difficult QP problem have been proposed.

A comparison for three common used QP optimization algorithms is shown as Table 1. From the comparison, the superiority of SMO is apparent and is most

Table 1: Comparison of QP Optimization Algorithms

Name	Introduction	Advantages	Disadvantages	Improvements
Chunking algorithm [9]	Break large QP problem into a series of smaller QP problems.	Reduces the requirement of storage capacity.	Overly relay on the numbers of SVs; has no effects when occurs to large data sets with high percentage of SVs.	Double chunking technique for solving large-scale problems [10].
Decomposition Algorithm [11]	Break large QP problem into a series of smaller QP problems (the breaking way is different from Chunking algorithm); its main work is selecting a suitable working set.	Reduces the training cost.	Its results are overly dependent on the choice of working set.	New methods for the working set selection [12]; two new decomposition algorithms for training bound-constrained SVM [13].
Sequential Minimal Optimization (SMO) [14]	A special case of decomposition algorithm; instead of using numerical QP as an inner loop as the formers, SMO uses an analytic QP step.	Mach faster than formers; reduces the requirement of storage capacity efficiently.	It is also time consuming for large-scale problems; training speed will slow when large margin values are used.	Methods for solving large size problems [15]; methods of speeding up training efficiency [16].

widely used in SVM nowdays. But for the large scale problems, it still needs further improvements.

Parameter Optimization for SVM

SVM refers to many parameters, slack variables, penalty parameter and Kernel parameters, and the selection of parameters takes many effects on the efficiency of SVM. Traditional optimization methods include cross-validation [17] and grid algorithm. The former improves the classification accuracy, while accompanying with high computation and low efficiency. The latter is a simple method and easy to use. However, it's sensitive to the initial value of the parameters and is time-consuming with low accuracy and intensive computation. Accounting for such defects, many new optimization methods have been proposed which are collectively referred to as swarm intelligence optimization.

The Table 2 is a comparison for parameter optimization methods. PSO and DE have shown their own superiorities. PSO has been combined with many other improvement algorithms to improve its performance. DE as a newer hot spot with significant advantages, still needs much more further research.

4 VARIANT SUPPORT VECTOR MACHINE

SVM is a strong classifier with a very good performance owing to its solid theoretical foundation and good classification results. Whereas SVM shows limitations in solving large-scale problems and multi-classification problems. Therefore, this section will represents some new SVM for improving traditional SVM.

Fuzzy Support Vector Machine (FSVM)

Many research results show that SVM is very sensitive to noises and outliners in the training sample due to overfitting [22]. In hence, FSVM [23, 24] was proposed to deal with the problem. Traditional SVM treated all training points uniformly and each training point belongs to either one or the other class. However, some input points may not be exactly assigned to one of these two classes. Some are more important to be fully assigned to one class so that SVM can seperate these points more correctly. Some data points corrupted by noises are less meaningful and should be discarded. The key point of FSVM is that it applies a fuzzy membership (importance weight) to each sample in the training datasets such that different samples can make different contributions in the construction of the SVM hyperplane. The fuzzy membership enhances the SVM in reducing the effect of outliers and noises in data points. Some newer methods assigned smaller weight or even zero weight

Table 2: Comparison of Swarm Intelligence Optimization

Name	Introduction	Advantages	Disadvantages
`Ant` colony optimization algorithm (ACO) [18]	An evolutionary algorithm based on population optimization; also called as a heuristic search swarm intelligence algorithm.	With strong robustness and strong global search ability; not sensitive to initial value of the parameters.	Convergence speed is slow; easy to fall into local optimum.
`Genetic` algorithms (GAs) [19]	Follows Darwin's evolutionary thought; catch the optimal value as follows through a series of operations: initializing population, calculating the fitness by fitness function, selecting, crossing, mutating and so on.	With high efficiency and global optimization; reduce the reliance on initial value selection.	Complex; fall into part optimum easily; not suit for peak problem.
Particle Swarm Optimization (PSO) [20]	An evolutionary algorithm newer than GA which originated in the behavior of birds foraging.	It has GA's advantages and is more simple than GA; has higher optimization efficiency and is easy to be implemented.	Premature convergence; not suitable for multi peak problems; easy to fall into part optimum.
Differential evolution (DE) [21]	It is arguably one of the most powerful stochastic real-parameter optimization algorithms in current use; has drawn the attention of many researchers all over the world and generates a lot of variants of the basic algorithm with improved performance.	More simple and fast; its algorithm control parameters are relatively less; with strong robustness and stronger global search ability; can solve multi peak problem.	Premature convergence; local search ability is weak.

to noisy and outlier data [25, 26] and a new type of FSVM [27] combined with some other algorithms was proposed to increase the difference of the meaningful training points and outliers. The future work for FSVM would focus on the selection of fuzzy membership function.

Twin Support Vector Machine (TWSVM)

TWSVM was proposed by Jayadeva et al. [28] for the binary classification problem and it has become the current researching hot spot in machine learning during the last few years [29]. TSVM generates two nonparallel hyperplanes by solving a pair of smaller-sized QPPs, such that each hyperplane is closer to one class and as far as possible from the other. The strategy of solving two smaller-sized QPPs rather than a single larger-sized QPP makes the learning speed of the TSVM faster than that of the standard SVM in theory. Due to TWSVM's lower computational complexity and better generalization ability, it has been studied extensively and developed rapidly in the last few years. Many variants of the TSVMs have been proposed, such as twin bounded support vector machine [30], nonparallel SVM [31], least-square TSVM [32], twin support vector regression [33],

etc. And recently, a novel twin support vector machine with the pinball loss (Pin-TSVM) was proposed to deal with the quantile distance to reduce the noise sensitivity and work fast [34].

Multi-class Support Vector Machine

SVM was originally designed for binary classification, while in real life, most applications involve multiple classification problems. Multi-class SVM [35] refers to two approaches: (1) consider all data in one optimization formulation directly; (2) construct and combine several binary classifiers. We typically construct a multiclass classifier by combining several binary classifiers because it made a great progress for algorithm research while the former way accounts for high computational complexity and achieves low efficiency on application [36]. Multi-class problems require that the number of variables should be proportional to the number of classes, which denotes that solve a multi-class problem consumes more calculations. Corresponding to such problems, some novel methods have been proposed. Lee et al. [37] proposed a multi-category SVM, which is a rightful extension from a binary SVM to a multi-category

SVM. Madzarov et al. [38] presented a novel architecture of SVM classifiers based on Binary Decision Tree architecture, takes advantage of both the efficient computation of the decision tree architecture and the high classification accuracy of SVMs. Recently, a novel multiple projection twin support vector machine (PTSVM) was proposed [39], which improves the generalization ability great.

Distribute Support Vector Machine (DSVM)

DSVM is based on connected networks. The network is divided into several sub-regions by some kind of certain strategy, and SVM is trained in each sub-region in serially or distributed. Finally, the sub-region adopts obtain the approximate solution or global optimal solution of the original SVM problem through some communications. Navia-Vazque et al. [40] proposed a true DSVM, which is relative to parallel SVM, includes two distributed schemes: a distributed packet method for propagating raw data, and a more detailed distributed semiparametric SVM. It reduces the total amount of information transferred between nodes and provides a privacy protection mechanism for information sharing. Then, Lu et al. [41] presented a strong connectivity network, Flouri et al. [42] proposed a DSVM over wireless sensor network and Kim et al. [43] proposed an improvement for this method later. Recently, distributed semi-supervised SVM [44], resilient DSVM [45] and some other new methods over DSVM has been proposed, which are different from existing parallel algorithms with good performance and experimental results. The various sub-problems only need to transfer hull information with each other, without the central processing unit. Accounting for such significant advantages, DSVM has become a popular direction for research.

It is worth mentioning that the new type of SVM achieves better and better performance, like above mentioned DSVM. A more newer method is called SVM classification trees [46], which is designed for multi-class problems. Giving the credit to these new methods' significant odds, they must be the future research hot spots.

5 APPLICATIONS OF SVM

SVM has a strong advantage on the basis of theory, it can guarantee that the solution of the extreme solution is the global optimal solution rather than the local minimum. In real application, SVM also has been successfully used for classification, regression and forecasting in all areas of life. Especially, it shows its own significant ability in pattern recognition, such as text recognition [47], handwriting recognition [48], face recognition [49], etc. In economic field, SVM is used for real estate forecasting [50], stock forecasting [51], etc. In transportation field, SVM is applied to identify license plate number[52], trained to distinguish driving performance and physiological measurements [53], etc. In medical field, SVM has been used in automatic medical decision support system for medical image [54], prediction of substrate specificities of membrane transport proteins [55], gene classification [56], etc.

Except for the applications in traditional areas, SVM also has been applied in the area of mobile multimedia. SVM is mainly used in the field of multimedia retrieval and multimedia event detection (MED). As for the multimedia retrieval, SVM has been applied to music and image retrieval in the early time. In 2001, Tong et al. [57] proposed the use of support vector active machine algorithm for conducting effective relevance feedback for image retrieval, which can quickly learn a boundary that separates the images and achieves high search accuracy. Mandel et al. (2006) [58] described a system for performing flexible music similarity queries using SVM active learning to achieve high search accuracy. As the demand for multimedia grows and with the evolution of digital technology, searching and organizing large scale datasets becomes a challenging task. Thus, some improved approaches have been presented in succession. For example, Yildizer et al. [59] proposed an extremely fast content-based image retrieval systems (CBIR) system which uses Multiple Support Vector Machines Ensemble, CBIR have become very popular for browsing, searching and retrieving images from a large database of digital images with minimum human intervention; Aryafar et al. [60] introduced a classifier fusion of multimodal audio and lyrics data to address single modality classification limitations. As for multimedia event detection, SVM has been applied to multimedia event detection (MED) based on audio or video in recent years. Wang et al. (2016) [61] proposed the method of classifying clips for events using "recurrent SVM" and Tzelepis et al. (2016) [62] using kernel SVM with isotropic gaussian sample uncertainty for video event detection, and both of them achieved good results. Except for those, SVM also can provide service of real-time multimedia quality assessment [63], mixed type audio classification [64], etc. In the future, accompanying with the development of SVM and its extension, SVM may be more widely used in the areas of mobile multimedia and achieve much more success.

6 CONCLUSION

SVM depends on statistical learning theory, which systematically studies the problem of machine learning, especially in the case of finite samples. It is based on the VC-dimensional theory and structural risk minimum principle and can effectively overcome the curse of dimensionality with the kernel function. As a result of these significant advantages, it has been successfully applied in many areas and achieves good efficiency. Accompanying with the development of technology, big data has become a mainstream. However, SVM is in a state of inferiority in dealing with large data problems. In hence, there has much further work to be done for further improvements for SVM, further improvement on parameters selection for SVM algorithm, integration of SVM and other disciplines, etc.

ACKNOWLEDGMENT

REFERENCES

[1] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[2] M. O. Stitson, J. A. E. Weston, A. Gammerman, V. Vovk, and V. Vapnik. Theory of support vector machines. *University of London*, 117(827):188 – 191, 1996.

[3] Richard G Brereton and Gavin R Lloyd. Support vector machines for classification and regression. *Analyst*, 135(2):230, 2010.

[4] Dapeng Wu, Junjie Yan, Honggang Wang, and et al. Socially incentive mechanisms for video distribution in device-to-device communications. *IEEE Trans. Multimedia*, 2017.

[5] Dapeng Wu, Hongpei Zhang, Honggang Wang, and et al. Quality-of-protection-driven data forwarding for intermittently connected wireless networks. *IEEE Wireless Commun.*, 22(4):66–73, 2015.

[6] Dapeng Wu, Yanyan Wang, Honggang Wang, and et al. Dynamic coding control in social intermittent connectivity wireless networks. *IEEE Trans. Vehicular Technology*, 65(9):7634–7646, 2016.

[7] Dapeng Wu, Boran Yang, Honggang Wang, and et al. An energy-efficient data forwarding strategy for heterogeneous wbans. *IEEE Access*, 4:7251–7261, 2016.

[8] Dapeng Wu, Boran Yang, Honggang Wang, and et al. Privacy-preserving multimedia big data aggregation in large-scale wireless sensor networks. *ACM Trans. Multimedia Computing, Communications and Applications*, 12(4s):60:1–60:19, 2016.

[9] Vladimir Naumovich Vapnik and Samuel Kotz. *Estimation of dependences based on empirical data*, volume 40. Springer-Verlag New York, 1982.

[10] Figueiras-Vidal et al. Pérez-Cruz, Fernando. Double chunking for solving svms for very large datasets. *Proceedings of Learning*, 2004.

[11] E Osuna, R Freund, and F Girosi. An improved training algorithm for support vector machines. In *Neural Networks for Signal Processing*, pages 276–285, 1997.

[12] Chih-Wei Hsu and Chih-Jen Lin. A simple decomposition method for support vector machines. *Machine Learning*, 46(1-3):291–314, 2002.

[13] Lingfeng Niu, Ruizhi Zhou, Xi Zhao, and et al. Two new decomposition algorithms for training bound-constrained support vector machines*. *Foundations of Computing and Decision Sciences*, 40(1):67–86, 2015.

[14] John C. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. In *Advances in Kernel Methods-support Vector Learning*, pages 212–223, 1998.

[15] Lijuan Cao, S. Sathiya Keerthi, Chong Jin Ong, and et al. Parallel sequential minimal optimization for the training of support vector machines. *IEEE Trans. Neural Networks*, 17(4):1039–1049, 2006.

[16] Shigeo Abe. Fusing sequential minimal optimization and newtons method for support vector training. *IJMLC*, 7(3):345–364, 2016.

[17] Grace Wahba, Yi Lin, and Hao Zhang. Margin-like quantities and generalized approximate cross validation for support vector machines. In *Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 12–20, 1999.

[18] C. B. Liu, X. F. Wang, and F. Pan. Parameters selection and stimulation of support vector machines based on ant colony optimization algorithm. *Journal of Central South University*, 39(6):1309–1313, 2008.

[19] Stefan Lessmann, Robert Stahlbock, and Sven F. Crone. Genetic algorithms for support vector machine model selection. In *IJCNN*, pages 3063–3069, 2006.

[20] J. Kennedy and R. Eberhart. Particle swarm optimization. In *IEEE Trans. Neural Networks*, pages 1942–1948, 1995.

[21] S Das and P. N Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evolutionary Computation*, 15(1):4–31, 2011.

[22] Yongqiao Wang, Shouyang Wang, and K. K Lai. A new fuzzy support vector machine to evaluate credit risk. *IEEE Trans. Fuzzy Systems*, 13(6):820–831, 2005.

[23] Han Pang Huang and Yi Hung Liu. Fuzzy support vector machines for pattern recognition and data mining. *Intl.j.of Fuzzy Systems*, 4(3):826–835, 2001.

[24] Chun-fu Lin and Sheng-De Wang. Fuzzy support vector machines. *IEEE Trans. Neural Networks*, 13(2):464–471, 2002.

[25] L. U. Yan-Ling, L. I. Lei, Meng Meng Zhou, and et al. A new fuzzy support vector machine based on mixed kernel function. In *ICMLC*, pages 526–531, 2009.

[26] Wan Mei Tang. Fuzzy svm with a new fuzzy membership function to solve the two-class problems. *Neural Processing Letters*, 34(3):209–219, 2011.

[27] O. N. Almasi and M. Rouhani. Fast and de-noise support vector machine training method based on fuzzy clustering

method for large real world datasets. *Turk J Elec Eng and Comp Sci*, 24(1):219–233, 2016.

[28] R Khemchandani, Suresh Chandra, et al. Twin support vector machines for pattern classification. *IEEE Trans. pattern analysis and machine intelligence*, 29(5), 2007.

[29] Divya Tomar and Sonali Agarwal. Twin support vector machine: A review from 2007 to 2014. *Egyptian Informatics Journal*, 16(1):55–69, 2015.

[30] Yuan Hai Shao, Chun Hua Zhang, Xiao Bo Wang, and Nai Yang Deng. Improvements on twin support vector machines. *IEEE Trans. Neural Netw*, 22(6):962–968, 2011.

[31] Y. Tian, Z. Qi, X. Ju, Y. Shi, and et al. Nonparallel support vector machines for pattern classification. *IEEE Trans. Cybernetics*, 44(7):1067, 2014.

[32] Yitian Xu, Xianli Pan, Zhijian Zhou, and et al. Structural least square twin support vector machine for classification. *Applied Intelligence*, 42(3):527–536, 2015.

[33] Sugen Chen, Xiaojun Wu, and Renfeng Zhang. A novel twin support vector machine for binary classification problems. *Neural Processing Letters*, 44:1–17, 2016.

[34] Yitian Xu, Zhiji Yang, and Xianli Pan. A novel twin support-vector machine with pinball loss. *IEEE Trans. neural networks and learning systems*, 28(2):359–370, 2017.

[35] Rameswar Debnath, N. Takahide, and Haruhisa Takahashi. A decision based one-against-one method for multi-class support vector machine. *Pattern Anal. Appl.*, 7(2):164–175, 2004.

[36] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Networks*, 13(2):415–425, 2002.

[37] Vojtech Franc and Václav Hlaváč. Multi-class support vector machine. In *IEEE Trans. Pattern Recognition*, volume 2, pages 236–239, 2002.

[38] Gjorgji Madzarov, Dejan Gjorgjevikj, and Ivan Chorbev. A multi-class SVM classifier utilizing binary decision tree. *Informatica (Slovenia)*, 33(2):225–233, 2009.

[39] Chun-Na Li, Yun-Feng Huang, He-Ji Wu, and et al. Multiple recursive projection twin support vector machine for multi-class classification. *IJMLC*, 7(5):729–740, 2016.

[40] A Navia-Vazquez and E Parrado-Hernandez. Distributed support vector machines. *IEEE Trans. Neural Networks*, 17(4):1091–1097, 2006.

[41] Yumao Lu, V Roychowdhury, and L Vandenberghe. Distributed parallel support vector machines in strongly connected networks. *IEEE Trans. Neural Networks*, 19(7):1167–1178, 2008.

[42] K Flouri, B Beferull-Lozano, and P Tsakalides. Distributed consensus algorithms for svm training in wireless sensor networks. In *EUSIPCO*, pages 1–5, 2008.

[43] W Kim, M. S Stankovic, K. H Johansson, and et al. A distributed support vector machine learning over wireless sensor networks. *IEEE Trans. Cybernetics*, 45(11):2599–2611, 2015.

[44] S Scardapane, R Fierimonte, Lorenzo P Di, and et al. Distributed semi-supervised support vector machines. *Neural Netw*, 80:43–52, 2016.

[45] Zhixiong Yang and Waheed U. Bajwa. Rd-svm: A resilient distributed support vector machine. In *ICASSP*, pages 2444–2448, 2016.

[46] Peter De Boves Harrington. Support vector machine classification trees based on fuzzy entropy of classification. *Analytica Chimica Acta*, 2016.

[47] Xiao-Liang Liu, SF Ding, Hong Zhu, and et al. Appropriateness in applying svms to text classification. *Comput Eng Sci*, 32(6):106–108, 2010.

[48] Mohamed Elleuch and Monji Kherallah. An improved arabic handwritten recognition system using deep support vector machines. *IJMDEM*, 7(2):1–20, 2016.

[49] W Ouarda, H Trichili, A. M Alimi, and B Solaiman. Face recognition based on geometric features using support vector machines. In *Soft Computing and Pattern Recognition*, pages 89–95, 2014.

[50] Xibin Wang, Junhao Wen, Yihao Zhang, and et al. Real estate price forecasting based on svm optimized by pso. *Optik - International Journal for Light and Electron Optics*, 125(125):1439–1443, 2014.

[51] Wei Shen, Yunyun Zhang, and Xiaoyong Ma. Stock return forecast with ls-svm and particle swarm optimization. In *BIFE*, pages 143–147, 2009.

[52] Michael Del Rose and Jack Reed. Support vector machine application on vehicles. In *International Symposium on Optical Science and Technology*, pages 144–149, 2001.

[53] Huiqin Chen and Lei Chen. Support vector machine classification of drunk driving behaviour. *International Journal of Environmental Research and Public Health*, 14(1), 2017.

[54] P LATHA et al. Svm based automatic medical decision support system for medical image. *JATIT*, 66(3), 2014.

[55] L Li, J Li, W Xiao, and Y Li. Prediction the substrate specificities of membrane transport proteins based on support vector machine and hybrid features. *IEEE/ACM Trans. Computational Biology and Bioinformatics*, pages 1–1, 2016.

[56] Bin Yu, Shan Li, and Haijun Liu. A hybrid gene selection method for tumor classification based on genetic algorithm and support vector machine. *J COMPUT THEOR NANOS*, 12(11):4730–4735, 2015.

[57] Simon Tong. Support vector machine active learning for image retrieval. In *ACM International Conference on Multimedia*, pages 107–118, 2001.

[58] Michael I Mandel, Graham E Poliner, and Daniel PW Ellis. Support vector machine active learning for music retrieval. *Multimedia systems*, 12(1):3–13, 2006.

[59] Ela Yildizer, Ali Metin Balci, Mohammad Hassan, and Reda Alhajj. Efficient content-based image retrieval using multiple support vector machines ensemble. *Expert Systems with Applications*, 39(3):2385–2396, 2012.

[60] K Aryafar and A Shokoufandeh. Multimodal sparsity-eager support vector machines for music classification. In *ICMLA*, pages 405–408, 2014.

[61] Yun Wang and Florian Metze. Recurrent support vector machines for audio-based multimedia event detection. In *ICMR*, pages 265–269, 2016.

[62] Christos Tzelepis, Vasileios Mezaris, and Ioannis Patras. Video event detection using kernel support vector machine with isotropic gaussian sample uncertainty (ksvm-igsu). In *MMM*, pages 3–15. Springer, 2016.

[63] Wenbi Rao and Yingshu Li. Real-time multimedia quality assessment based on media delay index and support vector machine. In *IEEE, CINC'09*, volume 1, pages 111–114, 2009.

[64] Lei Chen, Sule Gunduz, and M. Ozsu. Mixed type audio classification with support vector machine. In *IEEE, ICME*, pages 781–784, 2006.