International Core Journal of Engineering Volume 7 Issue 6, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

Review of Support Vector Machine Theory and Application Research

Yifan Zhang[1,a], Yong Liu and Xicheng Yang

[1]Henan University of Science and Technology University, Luoyang 471000, China.

[a]1837953025@qq.com

Abstract

Support vector machine (SVM) has a strong mathematical theory and theoretical foundation support, it is a machine learning method based on the VC dimension theory of statistical learning and the principle of structural risk minimization. The essence is a quadratic programming problem. Firstly, it introduces the theoretical basis of support vector machines, summarizes the application principles and current situation of support vector machines in the field of life, and finally looks forward to the research direction and development prospects of support vector machines.

Keywords

Statistical Learning Theory; Support Vector Machine; Classifier.

1. Introduction

As a good classification and regression method, support vector machines have achieved very significant research results, support vector machine is a machine learning method based on the VC dimension theory of statistical learning and the principle of structural risk minimization[1-3]. SVM has a strong mathematical theory and theoretical foundation support, it introduces a kernel function to map the linearly inseparable points in the low-dimensional space to the high-dimensional feature space through a nonlinearity, Find an optimal classification hyperplane in this space that can meet the classification requirements, in solving the problem of small sample, nonlinear and high-dimensional pattern recognition, the effect is considerable. At present, support vector machine classification technology is widely used in image recognition, text classification, fault diagnosis, and other fields.

2. Support Vector Machine Theory

2.1 VC dimensional theory and structural risk minimization principle

In order to study the speed and popularity of the consistent convergence of the learning process, the theory of statistical learning defines a series of indicators on the learning performance of function sets, one of the core concepts being VC dimensional. The intuitive definition of VC dimension [4] in the pattern recognition method is: For an indicator function set, if there are h samples, it can be separated by the functions in the function set in all possible forms of 2^h, then the function set is said to be able to break up the h samples; The VC dimension of a function set is the maximum number of samples it can break up h. If any number of samples have functions that can break them apart, the VC dimensionality of the function set is infinity. The VC dimension reflects the complexity of the problem, the larger the VC dimension, the more complicated the machine learning of the function set is. The concept of the VC dimension lays a theoretical foundation for the research of support vector machines, and the efficiency of the classifier can be measured well through the VC dimension.

For the principle of structural risk minimization, statistical learning theory introduces the concept of generalized error bounds. The theory clearly shows that the actual error of machine learning is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

composed of two parts, namely experience risk (training error) and confidence risk (VC confidence)[4-6].

$$R(w) \leq R_{emp}\ (w) + \Phi(h/n) \tag{1}$$

Where h is the VC dimension of the function set; n is the number of training samples; $R(w)$ is the actual risk, $R_{emp}\ (w)$ is the empirical risk; $\Phi(h/n)$ is the confidence risk. Among them, the confidence range Φ will change due to the size of the Φ value. When the number of samples represented by n is large, the confidence risk becomes smaller, and the optimal solution for minimizing the empirical risk is close to the actual optimal solution; The larger the VC dimension h of the classification function, the worse the generalization ability, and the greater the confidence risk. The goal of statistical learning is to change from seeking to minimize experience risk to seeking the smallest sum of experience risk and confidence risk, that is, to minimize structural risk, and support vector machines are just such an algorithm to minimize structural risk.

2.2 Support Vector Machine

The main purpose of supporting vector machine learning is to find an optimal classification superplane that meets the classification requirements, this allows the hyperplane to maximize the blank area on both sides of the hyperplane while ensuring the classification accuracy. First, take the analysis of two types of linear classification problems as an example. Given a set of training samples$(x_i, y_i), i = 1,2, \dots, l, x \in R^n, y \in \{\pm 1\}$, Can be separated by a hyperplane $\omega \cdot x + b = 0$, and the closest vector of the plane has the longest distance, then the realization of the optimal classification hyperplane becomes the solution to the following problem.

$$\begin{cases} min\, L(\omega) = \frac{1}{2} \parallel \omega \parallel^2 + C \sum_{i=1}^{n} \xi \\ s.t. y_i(\omega^T \Phi(x_i) + b) \geq 1 - \xi, \xi > 0 \end{cases} \tag{2}$$

in the equation, the $\Phi(x_i)$ is $R^n \to R^m$, where $m > n$ is a nonlinear mapping equation, Map sample points from low-dimensional space to high-dimensional space, Look for optimal classification hyperplanes in this space to meet classification requirements; ξ is a slack variable that represents an error measure of a data point. $C(C > 0)$ is the punishment factor, is the wrong sample of the degree of punishment control, the larger the penalty, but its generalization ability will also be reduced.

$L(\omega)$ it is a quadratic function, which has a unique minimum point. The Lagrange function is introduced for optimization, and the problem of constructing the optimal hyperplane is transformed into a more understandable dual form:

$$\begin{cases} maxW(\alpha) = \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ s.t. C \geq 0, i = 1 \cdots n \\ \sum_{i-1}^{n} x_i, y_i = 0 \end{cases} \tag{3}$$

in the formula, $\alpha_i(\alpha_i > 0)$ is the Lagrangian multiplier, and $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ is the kernel function.

2.3 Kernel function

In practice, most of the data that needs to be identified and classified are nonlinear and inseconcable, and for nonlinear problems, the researchers mapped the input space to high-dimensional space using appropriate nuclear functions, realize linear separability of high-dimensional space, transform nonlinear problems into linear problems to solve them. Choosing the appropriate kernel function determines the application performance of the support vector machine, so the kernel function selection problem has become a research hotspot[5].

Under normal circumstances, the appropriate kernel function for SVM can be selected according to the previous experience of the researchers, or it can be tried one by one using experimental methods. There are four main types of commonly used nuclear functions:

(1) Gauss Radial Base Function (RBF)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

$$K(x_i, x_j) = exp(-\| x_i - x_j \|^2 / \sigma^2) \tag{4}$$

(2) Sigmoid kernel function

$$K(x_i, x_j) = tanh\left((x_i \cdot x_j) + \theta\right) \tag{5}$$

(3) Polynomial kernel function

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \tag{6}$$

(4) Linear kenel function

$$K(x_i, x_j) = x_i \cdot x_j \tag{7}$$

3. Application

3.1 The field of image recognition

In the field of image recognition, face detection has a wide range of practical value in many applications such as intelligent human-machine interface, visual monitoring, image labeling, and retrieval, it's one of the research hotspots in the field of computer vision and pattern recognition. Early Osuna et al [8] used the SVM method in face detection technology to train a nonlinear SVM to classify faces and non-faces, and achieved good results. But this method still has shortcomings: One of them is that when using the SVM method for face detection, a large number of non-face samples need to be collected, this will affect the performance of the classifier, because SVM is relatively more suitable for small sample training sets. The other is that the nonlinear SVM classifier requires a large number of support vectors, and the calculation speed is slow. Regarding the problems of difficulty in classifier training and large amount of calculation in the method in literature [8], Yong Ma et al [9] proposed a frontal upright face detection method based on hierarchical support vector machine, the classifier of this structure consists of a combination of linear support vector machines and a nonlinear support vector machine. The former quickly eliminates most of the non-face areas in the image while ensuring the detection rate, and the latter further confirms the face candidate areas, experiments show that this method has a higher detection rate and a faster speed. Similarly, in the literature [10], the authors used a local ternary pattern (LTP) based on a genetic algorithm (GA) to identify faces by supporting vector machines, which used LTP and GA to extract feature vectors and separation features to further reduce computational time and improve recognition accuracy.

Based on the problems in the above research, some scholars have recently combined other machine learning algorithms with SVM and applied them to the field of image recognition. For example, Hongmin Yin et al[11] used the Principal components analysis (PCA) combined with SVM for face recognition, first using the fast PCA algorithm for feature degradation, and then using SVM classification has a better recognition rate. Xu and others [12]proposed a facial recognition method based on the fusion of Principal Components Analysis (PCA), Linear Discrimination Analysis (LDA) and SVM algorithm. Use the PCA to transform face images into new feature spaces,eliminate the correlation and noise between image features, extract the global features of the face, then use the LDA algorithm to further project and transform to reduce the data dimension, Finally, SVM classification is used. Combining the advantages of PCA, LDA and SVM three algorithms, the simulation experiment results show that compared with the joint method of PCA+SVM, the recognition rate of this method is increased by 5 percentage points.

3.2 Text classification field

Dumais [13] used the SVM method for text classification as early as 1998.The experimental results show that the application of the SVM method in text analysis is more effective than other methods such as Bayesian and decision tree, and it also has better generalization ability. This research result has attracted the attention of related scholars, who have conducted more in-depth research on the application of the SVM method in text classification. Hu Zhang and others [14,15] used the SVM classification method based on the integrated learning strategy to detect and recognize the deceptive

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

information in the Chinese text, The detection method is similar to the one-to-many method in multi-class classification SVM, the disadvantage is that the training sample is not abundant, and the reliability of experimental results needs to be further strengthened. In the study, Dian and others [16] proposed a sub-gradient-based fall method to optimize the support vector machine classifier, this can increase the speed of training data. The research shows that in data classification, this method can be used to optimize the training time of text data sets.

3.3 Fault diagnosis field

With the advent of industry 4.0 and the development of modern industry, fault diagnosis has become an indispensable and important link in the production process, this has undoubtedly promoted the widespread use of artificial intelligence in all walks of life. One of the research areas of intelligent fault diagnosis is the use of machine learning methods. Not only do we want machines to have human thinking, but we can also make faster and more accurate diagnoses than humans. Classification is an important method in the process of knowledge mining, and also an important method in fault diagnosis. We saw scholar L.B.Jack [17]successfully applied it to the fault detection of bearings and motors. Such methods have been formally introduced into the troubleshooting field. SVM is a commonly used fault diagnosis classifier, and the selection and optimization of its parameters are very important for the classification effect of the classifier. R. Hong et al [18] proposed a fault diagnosis method based on SVM and Particle Swarm Optimization (PSO) for electrical frequency identification, use the PSO algorithm to optimize the parameters of the SVM to avoid the blindness of manually determining the parameters, combining an example of an airport baggage sorting system based on radio frequency identification, it shows that this method is superior to traditional fault diagnosis methods. Later, the scholar Fang Yu [19] used the PCA method to extract feature indicators in his research,and used the gray wolf optimization algorithm to optimize the kernel parameters, use this method to diagnose faults in belt conveyors, the simulation experiment solves the problem that the SVM method cannot perform multiple fault classification.

3.4 Other application areas

The superiority of the support vector machine has led to a large number of applications. An improved time adaptive support vector machine for working with non-stable data sets is proposed in the literature [20], solve data loss and change, and reduce computational complexity, but are constrained by sample time and data set length. Bai Jing et al [21] aimed at the problems of the speech recognition system's recognition rate becoming worse in a noisy environment, a method to optimize SVM parameters based on parallel structure artificial fish group algorithm (PAFSA) based on habitat sharing mechanism is proposed. The artificial fish algorithm is improved, the diversity of sample individuals is maintained in the process of finding excellence, the accuracy of solution speed and reconciliation is improved, and its effectiveness is verified by experiments.

4. Conclusion

This paper focuses on the analysis of the basic principles of SVM and selects representative research literature on the application of image recognition, text classification, and fault diagnosis to explain the different roles of SVM in practical application, which has proved its advantages.

(1) The kernel function improvement and optimization of parameters The performance of SVM depends to a large extent on the selection of the kernel function, so further development is still needed in the selection of the kernel function; Secondly, the parameters of the kernel function are adjustable. Researchers can improve SVM performance by adjusting them based on certain parameter optimization methods. For example, the group optimization algorithm has achieved many important results in the optimization of SVM parameters. It has strong parallel processing capabilities, fast optimization speed, avoids falling into local optimality prematurely, and can perform global optimization, etc [22].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

(2) In combination with other models. Deep learning methods are becoming more mature and effective in pattern classification and target recognition applications[4]. SVM can be used in combination with commonly used algorithms in deep learning to maximize the detection and recognition rate while taking advantage of the advantages of both.

References

[1] VAPNIK V N. The essence of statistical learning theory (Beijing: Tsinghua University Press, China 2000.), Xuegong Zhang translated.

[2] CRISTIANINI N, TAYLOR Js. Introduction to Support Vector Machines(Beijing: Publishing House of Electronics Industry, China 2004), Guozheng Li, Meng Wang, Huajun Zeng, translated.

[3] Shifei Ding, Puyi Qi, Hongyan Tan. Summary of Support Vector Machine Theory and Algorithmic Research [J]. Journal of the University of Electronic Science and Technology. Vol. 40 (2011) No. 1, p. 1-10. (In Chinese)

[4] Fangyuan Liu, Shuihua Wang, Yudong Zhang. Overview of support vector machine models and applications [J]. Computer System Applications. Vol. 27 (2018) No. 4, p. 1-9. (In Chinese)

[5] Junjie Tao. Research on Rolling Bearing Fault Diagnosis Based on Wavelet Analysis and Support Vector Machine. Beijing University of Chemical Technology. 2019. (In Chinese)

[6] Haiyan Wang, Jianhui Li, Fenglei Yang. Summary of support vector machine theory and algorithm research [J]. omputer Application Research. Vol. 31 (2014) No. 5, p. 1281-1286. (In Chinese)

[7] Songlan Zhang. Overview of Support Vector Machine Algorithms and Applications [J]. Journal of Jiangsu Institute of Technology. Vol. 22 (2016) No. 2, p. 14-17,21. (In Chinese)

[8] E. Osuna, R. Freund and F. Girosit. Training support vector machines: an application to face detection [C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 1997, p. 130-136.

[9] Yong Ma, Xiaoqing Ding. Face detection based on hierarchical support vector machine[J]. Journal of Tsinghua University (Natural Science Edition). Vol. 43 (2003) No. 1, p. 35-38. (In Chinese)

[10] D.Tyagi, A.Verma, S.Sharma. An improved method for face recognition using local ternary pattern with GA and SVM classifier [C]. 2016 2nd International Conference on Contemporary Computing and Informatics(IC3I). Noida,2016, p. 421-426.

[11] Hongmin Yin, Guangying Ge. Research on face recognition based on fast PCA-SVM[J]. Journal of Science Education. Vol. 6 (2016), p. 147-147,157. (In Chinese)

[12] Zenze Xu, Zuhong Wu, Yan Xu, et al. Face Recognition with PCA, LDA, and SVM Algorithms [J], Computer Engineering and Applications. Vol. 55 (2019) No. 18, p. 34-37. (In Chinese)

[13] DUMAINSS S. Using SVMs for text categorization[J]. IEEE Intelligent System, Vol. 13 (1998),p.21-23.

[14] Lai Su, Xiong Zhongyang, Jiang Fan, et al. Using an improved polynomial kernel function support vector machine for text classification [J]. Journal of Chongqing University: Natural Science Edition. Vol. 35 (2012) No. S1, p. 45-51. (In Chinese)

[15] Lakshmi G., J. R. Panicker, Meera M. Named entity recognition in Malayalam using fuzzy support vector machine [C]. 2016 International Conference on Information Science (ICIS), Kochi, 2016, p. 201-206.

[16] Dian Puspita Hapsari,Imam Utoyo,Santi Wulan Purnami. Text Categorization with Fractional Gradient Descent Support Vector Machine [C]. Proceedings of the 1st International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019).2019.

[17] Hui Yi: Fault diagnosis and application research based on support vector machine [D], Nanjing University of Aeronautics and Astronautics, 2011. (In Chinese)

[18] R. Hong, C. Guorong, L. Yao and H. Hongli: Research on RFID Fault Diagnosis Method Based on SVM and PSO Algorithm [C], 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), Chongqing, China, 2018, p122-126.

[19] Yu Fang: Research on fault diagnosis and prediction of belt conveyor based on support vector machine [D], China University of Mining and Technology, 2019. (In Chinese)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0049

[20] Y.Z. Shi, F.L Chung, S.T. Wang. An improved TA-SVM method without matrix inversion and its fast implementation for nonstationary datasets [J]. IEEE Transactions on Neural Networks and Learning Systems. Vol. 26 (2015). No. 9, p. 2005-2018.

[21] Jing Bai, Lihong Yang, Xueying Zhang: An Anti-Noise SVM Parameter Optimization Method for Speech Recognition [J]. Journal of Central south University. Vol. 44 (2013). No. 2, p. 604-611. (In Chinese)

[22] Su Li, Zhigao Yuan, Cong Wang, et al. Group Intelligence Algorithm Optimization Support Vector Machine Parameter Overview. Journal of Intelligent Systems, Vol. 13 (2018) No. 1, p. 70-84. (In Chinese)