

Research on Confidence Machine

Fangchun JIANG^{1, 2}

¹School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

²ShenZhen Institute of Information Technology, ShenZhen, GuangDong 518172, China

06112061@bjtu.edu.cn

Keywords: machine learning, confidence machine, controlled confidence, reject option, SVM

Abstract. Manageable confidence machine learning is one of the important approaches to implement confidence machine application. This paper is based on two class confidence classifier, adopting two class classifier as tool to convert learning results of classifiers and achieve confidence management through setting threshold values. The research accomplished manageable general accuracy of the classification and manageable positive/negative classification accuracy. Such method is tested in 5 experimental data sets of cardiopathy and diabetes, achieved preferable research result.

Introduction

Confidence Machine (CM) is a sub-division of machine learning research which has not been studied for long time. The methods of implementing confidence machine learning are few [1]. The research of reference [2] transforms classification distance of machine learning results into posterior probability, then uses posterior probability to estimate confidence. The research of reference [3] adopts multi-layer feed forward neural network, and uses neural network output expectation as posterior probability, to perform confidence estimation. The research of reference [4] eliminates unreliable part using refuse option in output result in order to improve confidence of remaining output to achieve confidence classification. The research of reference [5] carries out algorithm research on manageable confidence machine, proposes to use slot method to achieve confidence machine classification algorithm. Confidence machine has practical significance in high risk fields such as medical diagnosis.

Support Vector Machine (SVM) [6] is a powerful tool in machine learning, which was proposed by Vapnik in 1992 and achieved full in-depth application and development. The SVM is reinforced by maximum margin hyper plane, kernel function, convex quadratic programming, slack variable, sparse solution and some other technologies, becoming a very successful method in machine learning field. It also becomes one of the standard tools of machine learning.

This paper is based on support vector machine as the tool of two-class classifier. The research adopts indirect convert confidence method and setting refuse option method to implement manageable confidence algorithm design. Such approach has been tested on cardiology and diabetic experimental data and met the research goal.

Algorithm design and theoretical principle

Such Algorithm involves technologies like SVM and refusing option, which will be discussed respectively.

(1) 、 SVM

Support vector machine [7], is a two class classifier model, its basic model is defined as the linear classifier of maximum interval on feature spaces. Its learning strategy is maximizing interval, eventually transforms to a solution of one convex quadratic programming problem.

Firstly, we need to define the classification problem, based on given training set $S = \{(x_1, y_1), \dots, (x_l, y_l)\} \subset (X \times Y)^l$, $(x_i, y_i) = (X, Y) \in (\mathfrak{R}^n, \pm 1)$, l is the number of sample points. Its purpose is to seek one real-valued function $f(x)$ on $X = \mathfrak{R}^n$, in order to apply decision function

$$g(x) = \text{sgn}(f(x)) = \text{sgn}((w \bullet x_i) + b) \quad (1)$$

to confirm the value y corresponding to x . The essence of solving classification problem is to find out one rule of dividing points on \mathfrak{R}^n to two groups.

If the training set can be divided without error, and if each class of data has maximum distance with hyper plane, then this hyper plane is called optimal classified hyper plane, written as function $(w \bullet x) + b = 0$, it's constraint is displayed as

$$y_i((w \bullet x_i) + b) \geq 1, \quad i = 1, \dots, l \quad (2)$$

In the condition of approximately dividable sample set, the research introduces slack variable $\xi = \{\xi_1, \dots, \xi_l\}$, allowing certain error in classification of training set. After introduction of slack variable, constraint (2) becomes:

$$y_i((w \bullet x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l \quad (3)$$

The geometrical interval of classification is $\Delta = \frac{1}{\|w\|}$, maximizing classification interval Δ is to minimize $\|w\|$, hence the optimal target function and constraint of learning problem are as follows:

$$\min_{\omega, b, \xi} \tau(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (4)$$

$$s.t. \quad y_i((w \bullet x_i) + b) \geq 1 - \xi_i \quad (5)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (6)$$

where $C > 0$ is the penalty factor, also known as the regularization factor. It controls the penalty degree to error samples, achieving balance between sample error and machine generalization ability. The target function juggles empiric risk and confidence limit, satisfying minimum structural risk principle.

The aforementioned optimal classified hyper plane is introduced in case of linear sample classifiable or near linear classifiable, which is referred to as linear support vector machine. When the samples can not linearly classified, constraint (2) is no longer tenable, more complex hyper-surface will replace classification hyper plane, which is non-linear support vector machine and is used to treat with linearly un-classifiable problems.

The working principle of non-linear support vector machine is through some projecting function Φ , projecting training samples from input space to one high dimension (or even infinite dimension) feature space Z . This approach will make the training sample to present linear (near) classifiable feature in feature space, then form the optimal classification hyper plane in feature space. The transition of space enables linear method to solve non-linear decision function. After the application of projecting function, the optimal question equations (4) to (6) are turned to

$$\min_{\omega, b, \xi} \tau(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (7)$$

$$s.t. \quad y_i((w \bullet \Phi(x_i)) + b) \geq 1 - \xi_i \quad (8)$$

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (9)$$

Normally, we do not directly solve (7) to (9), but use Wolfe dual form to solve them

$$\min_{\alpha} W(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \quad (10)$$

$$s.t. \quad 0 \leq \alpha \leq Ce \quad (11)$$

$$y^T \alpha = 0 \quad (12)$$

In which, the α is the Lagrange multiplier, e is a vector with elements that are all 1, Q is called Hessian matrix, $Q_{ij} = y_i y_j K(x_i, x_j) \in \mathfrak{R}^{l \times l}$. These are typical Quadric Programming (QP) problems. If

Q is (positive definite) semi-positive definite matrix, then above problem is (strict) convex quadratic programming.

Solving quadratic programming process corresponds to learning machine's parameter optimization process (training process), finally constructs the learning machine's discrimination function as

$$y = \text{sgn}\left(\sum_{i \in SV} \alpha_i^* y_i K(x_i, x) + b^*\right) \tag{13}$$

Optimal classification hyper plane only is determined by corresponding $\alpha_i^* > 0$ samples, these samples are support vectors, their geometrical meaning is that they are on the maximum interval hyper plane. The so called support means the classification hyper plane H and its edges are defined by these samples.

(2) 、 Refuse Option

Dividing the sampling space into two complementary fields: refuse field *R* and accept field (classification field) *A*, their definition are as follows:

$$R = \{x \mid 1 - \max_i p(\omega_i \mid x) > t\} \quad A = \{x \mid 1 - \max_i p(\omega_i \mid x) \leq t\} \tag{14}$$

In which, *t* is the threshold value, the smaller threshold value *t* gets the larger refuse field *R*. If sample *x* is in accept field *A*, then according to some learning method to classify *x*. If sample *x* is in refuse field *R*, then refuse classification to *x*.

Reference [8] performed research on classifier with refuse option, proposed optimized classifier and refuse rule. To two class classification problems, according to Bayes learning method, if $p(\omega_1 \mid x) \geq (1-t)$, then *x* belongs to ω_1 ; If $p(\omega_2 \mid x) \geq (1-t)$, then *x* belongs to ω_2 , else refuse *x*. The parameter *t* is the refuse threshold value and also a constant value, $0 \leq t \leq 0.5$, which is used to set posterior probability threshold value. If *t* is 0, the posterior probability is 1, all classifying sample are accepted; if *t* is 0.5, it's like Bayes learning without refuse option. The posterior probability of refusing any sample is less than $1-t$, ensuring high creditability to achieve confidence determination.

Algorithm implementation

Directly set threshold value to achieve manageable confidence classification.

Such method is relatively intuitive and simple, according to the distribution of outputs after classification, directly set the threshold value, then compute the accuracy, which is the confidence. The procedure is shown in Figure 1.

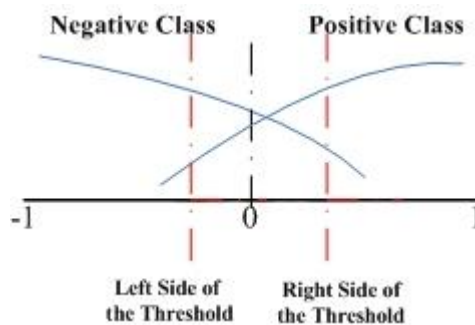


Fig.1 Algorithm 3

The average result of executing such algorithm for 10 times is shown in table 1.

Tab.1 Average result of Algorithm 3 executed 10 times

No.	Data set name	The threshold is not set		The threshold is set		
		Accuracy %	error rate %	The actual accuracy %	The left side of the threshold	The right side of the threshold
1	heart	82.29	17.71	88.17	-0.8	0.8
2	pima indians diabetes	79.56	20.44	89.81	-0.8	0.8
3	colon-cancer	87.50	12.50	94.67	-0.8	0.8
4	liver-disorders	74.22	25.78	79.73	-0.8	0.8
5	australian	85.56	14.44	91.17	-0.8	0.8

Conclusions

This paper based on manageable confidence machine theory and proposed algorithms: directly setting threshold value to achieve manageable confidence classification, and tested on 5 data sets including cardiology and diabetes data sets to verify the design requirement.

At the same time, further research will be carried out to improve the control accuracy on manageable confidence machine.

Acknowledgements

Special thanks to Natural Science Foundation of Guangdong (No. S2011010000824); Shen Zhen Collaborative Science and Technology Innovation Project (GJHS2012062809424003)

References

- [1] Jiang Fangchun, Tian shengfeng, Yin chuanhuan. A Survey on the confidence mechanism research in machine learning[J]. Journal of Beijing Jiaotong University, 2014, Vol. 38 (3): 111-117.
- [2] Liu C L, Nakagawa M. Precise Candidate Selection for Large Character Set Recognition by Confidence Evaluation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(6):636-642.
- [3] Richard M D, Lippmann R P. Neural Network Classifiers Estimate Bayesian a Posterior Probabilities[J]. Neural Computation, 1991, 3(4): 461-483.
- [4] Grandvalet Y. Support Vector Machines with a Reject Option[C]// Vancouver B.C. Canada: Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference, 2009: 537-544.
- [5] Li M, Sethi I K. Confidence-based Classifier Design[J]. Pattern Recognition, 2006, 39: 1230-1240.
- [6] Ethem Alpaydm. Introduction to machine learning[M]. Beijing: China Machine Press,2009: 142-149.
- [7] Nello Cristianini, John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods [M]. Beijing: Electronic Industry Press, 2005: 82-98.
- [8] Chow C K. On Optimum Recognition Error and Reject Tradeoff[J]. IEEE Trans. on Info. Theory, 1970, 16: 41-46.