International Journal of Database Theory and Application
Vol.7, No.6 (2014), pp.151-162
http://dx.doi.org/10.14257/ijdta.2014.7.6.14

The Nuclear Techniques and the Selection of Model Parameters in Big Data

Wang Chunhong

Computer Department, Inner Mongolia University of Finance and Economics, Hohhot, China
wangchunhongkeyan@163.com

Abstract

Now a large scale of data every day, the large-scale data is usually in the form of database storage. The law of the people wants to find useful or knowledge, thus was born the Data Mining technology. SVM (Support Vector Machine, SVM) is a very useful method in data mining, this paper mainly discusses the Support Vector Machine (SVM) play a key role in nuclear techniques and the selection of model parameters is analyzed and evaluated. This article some methods about how to construct the kernel function is introduced for the model to find suitable kernel function is to provide some reference strategies and proposed kernel function method for the simulation analysis.

Keywords:*big data; data mining; support vector machine (SVM); model parameters*

1. Introduction

As a result of information processing in the popularity of business, government, scientific research and so on all walks of life, produce the huge data every day. These mass data is usually stored in the form of a database. People want to find useful rules or knowledge, for business analysis and decision, scientific exploration, production testing, thus was born the Data Mining technology (Data Mining,) [1]. Data mining technology is different from traditional database technology and on-line Analytical Processing technology (On-Line Analytical Processing, OLAP), it can find the unknown, and potential information. Found in beer - just as the famous wal-mart diapers example, it is hard to predict in advance what relationship between beer and diapers. Data mining techniques can be used in noisy, sparse and redundant data sets, and has low computation complexity, easy to understand and less artificial parameter, *etc.*, with these advantages quickly has an important role in the analysis of the database.

In many data mining classification algorithm, more commonly used include naive Bayes, artificial neural network and Support Vector Machine (Support Vector Machine, SVM) [2], SVM has a solid theoretical foundation and beautiful geometric structure, make it more and more get the attention of the researchers. Since Nick and others from 60 s study of statistical learning theory, and the SVM algorithm is put forward in the ninety s, then the SVM attention and rapid development, for improved efficiency of the algorithm, the model hypothesis and explain, model parameter selection aspects put forward many new results.

SVM dealing with linear inseparable problem is the nature of the training set is mapped to high-dimensional feature space for maximum separation hyper plane, and map into the space to how to determine the SVM model, kernel function and model parameters are basically determines the mapped to high-dimensional feature space, the choice of

ISSN: 2005-4270 IJDTA

kernel function and the determination of model parameters is an important research direction of the SVM.

Machine learning algorithms [3] are often based on the distribution of training data, the decision function, estimated by empirical risk minimization principle, summing up the decision function of definition on all samples, namely from the training samples directly to the general rule, when decisions, in turn, with general rules to determine samples, and when the training sample of decision function expression is not enough, such as the characteristics of the training set is not obvious or number is less, can't induce complex decision function contains a large amount of information, the SVM by grasping the key of the sample, and on this basis to predict other unknown sample values, the decision function to retain some of the characteristics of the sample, is the training sample to the general law of excess. Because traditional pattern classification methods such as Bayes theorem, nearest neighbor, decision tree, Hsher discriminant, artificial neural network based on statistical theory, such as need to know the input sample prior distribution of the space, and normally the training sample is difficult to represent the whole input space distribution, often causes the generalization ability of classification algorithm. SVM based by the SLT, and said the geometry of a straightforward and simple mathematical model, makes it from immediately after the proposed important classification algorithm, and a lot of machine learning algorithm, the SVM decision function is sample and the combination of kernel function, similar to the nearest neighbor and case-based reasoning method, adopt the bypass key sample having the characteristics of the decision function of explicit expression, reduce the information loss during induction, simplifies the whole training process at the same time.

Standard SVM using existing quadratic programming algorithm, the training time to sample index scale growth, and put the whole Hesse matrix (Hessian) in memory, the size of its space occupied by the square of the number of samples increase, so the large-scale samples SVM is one of the important research direction.

Machine learning algorithm is a lot of unknown categories of samples with known samples to compare, to find the most similar with the category as a decision-making strategy, and the decision of the input should be similar to that of similar; Kernel function to a certain extent, reflects the input space to the similarity between the training sample, a polynomial kernel function with the distance in space, reflects the similarity of similar concept of judgment is one of the important basis for selection of kernel function, we can according to the research system of prior knowledge and the training set to choose kernel function.

Nuclear techniques [4] not only played a key role in the nonlinear SVM, the differential equation, harmonic Analysis, group theory, and other fields of machine learning is also doing well, since Aizerman scholars in 1964 will be the first nuclear function instead of the inner product technique to the research of data mining, a simplified algorithm to improve the operation efficiency, many scholars began to study by adding a Kernel function to many ordinary linear algorithm is extended to nonlinear, such as at the beginning of the end of last century to the world, such as Schkopf researchers combined Kernel function, and the extension of the traditional Principal Component Analysis can extract the nonlinear relationship between the characteristics of Kernel Principal Component Analysis (Kernel Principal Component Analysis, KPCA), and thought is the training samples mapped to high-dimensional space, to find a set of unrelated linear projection direction as characteristic direction, after the mapping of samples along the characteristic in the direction of projection is extracted to sample characteristics, because the whole process of nonlinear mapping and inner product appear

together, so you can use the Kernel function substitution and don't need to know the specific explicit nonlinear mapping.

Nuclear techniques rapid development, is the result of nuclear method has many advantages, such as (1) can be linearized nonlinear method of solving, the original training set the nonlinear mapped to high-dimensional feature space, in the feature space using the original structure of linear model, nonlinear model corresponding to the original space. (2) the introduction of kernel function to avoid explicitly shows the original space to high dimension feature space of nonlinear mapping, directly through the kernel function is nonlinear mapping is given and the result of the inner product, improve the speed of modeling. (3) the nuclear techniques the adjustment model, need to modify kernel functions and parameters, different areas of the practical problems facing the same framework based on prior knowledge and selects the best distribution of training samples and kernel function. (4) many kernel function can be added to the algorithm, the original linear algorithm suitable for nonlinear case.

2. Related Works

Research on machine learning, can be traced back to the 40-50 s of the 20th century, people at that time was conducted from the perspective of bionics research on learning mechanism of human brain and nervous system. In 1943, for example, McCulloch and Pitts (abbreviated to MP model) on neuron model research. In 1957, F.R osenblartt first learning machine model is put forward that sensors, this marked the beginning of the learning process were studied with mathematical method. He put the perceptron model characterized by a computer program, and illustrates its promotion ability through the experiment. 1962, Novikoff proved about perceptron's first important theorem, the theorem in creating learning theory has played a very important role, is the beginning of the learning theory.

After the perceptron is mentioned, people soon put forward other types of learning machine, such as adaptive learning machine, hidden markov model to solve practical problems. But these machines are only tools of solving practical problems, not the general model of learning problems. In 1986, completed the general learning machine structure after the research to the spread of technology. In the meantime, also produce the statistical learning theory[5] and has large development, produced the empirical risk minimization principle theory and algorithm complexity thoughts.

After the discovery to the spread of technology is a leap of perceptron, sensors, also known as neural network. The technology in modifying the model, the synthesis of new neurons is a continuous function. Using gradient of calculate the coefficient of neurons, can be applied to any based on gradient method to construct the expected function approximation. In many practical applications, the neural network has achieved good results.

In the 20th century, 60-70s, Vapnik established statistical learning theory. Compared with the traditional statistics, statistical learning theory is a specialized research theory of machine learning law under the condition of small sample. By the mid ninety s, with the continuous development of its theory and mature, also because of the neural network learning methods, such as lack of real progress in theory, statistical learning theory is starting to get more and more extensive attention. The theory is based on a more solid theoretical foundation, to solve the problem of limited sample study provides a unified framework. It can include many existing methods, is expected to help solve many difficult to solve the original problem (such as neural network structure selection problem

of local minimum point, *etc.*); At the same time, on the basis of the theory developed a new general learning method, support vector machine (SVM), it has shown a lot better than the performance of the existing methods. Some scholars believe that the SLT and SVM is becoming a new research hotspot after neural network research, and will effectively promote the development of machine learning theory and technology.

Because of the traditional standard SVM under large sample, the computing speed and storage size for hardware, there is a GouKe requirement, so need to improve to convenient for large-scale sample applications, domestic and foreign existing algorithm to solve such problems, mainly using the selected block algorithm (Chunking), Decomposition algorithm (Decomposition) [6], minimum sequence Optimization algorithms (Sequential Minimal Optimization, SMO), after the first cluster SVM, heuristic to find Support vector (the Support Vectors, v) and the smoothing method and the skill, reduce the running time and memory footprint.

Ms Cortes and proposed by Vapnik selected block algorithm (Chunking) in an iterative fashion constantly out of training set of china-africa support vector corresponding to a subset of the training sample, retain support vector, using Kuhn - Tucker violations (Karush - Kuhn - Tucker, KKT) conditions of outage criterion for judging the selection to add the basis of the training sample. Algorithm to calculate the size is decided by the number of spatially-integrated SVs, to a certain extent, reduced memory requirements, but when the number of spatially-integrated SVs many, algorithm efficiency does not increase a lot, and even falling in individual cases.

In 1997, Osuna spatially-integrated SVS is large Decomposition algorithm (Decomposition) was proposed, the training set is divided into work and non-work sets two subsets, each iteration only add or remove the same amount of work is sample, make the working set size remains the same. Joachiins continue to improve the update strategy, choose an algorithm and by contracting method to reduce the support vector and reduced support vector on the side of the nuclear matrix, get the SVMlight efficient algorithm. Platt with Minimal working set the two training samples, using the analytic method of quadratic programming solution directly referred to as the minimum sequence Optimization algorithms (Sequential Minimal Optimization, SMO), Lin, *etc.* The convergence of decomposition algorithm SVMlight, at the same time, Lin and Keerthi Platt demonstrates that the proposed SMO algorithm under certain conditions the linear convergence rate. Zanghiratitn et al. also parallel learning algorithm is proposed, to solve split training set are calculated respectively, after each results achieve the purpose of rapid training after the merger, but higher requirement to the computer hardware.

At present, the application of support vector machine (SVM) has gradually become the researchers focus, in pattern recognition, regression estimation, probability density function is estimated, and other fields have application results. In terms of pattern recognition, the most prominent application research is bell LABS to experiment, the usps handwritten digital library use three kinds of support vector machine (SVM) method to get the error rate of 4.0%, 4.1%, 4.2%, is better than that of the decision tree and neural network method. In the field of face detection and recognition, MIT face detection by support vector machine (SVM) experiments have achieved good results, you can better learn to identify possible location of a face in the image, also had more domestic research. Support vector machine (SVM) is widely used in image processing, image segmentation, image detection, image retrieval and so on all obtained the good application, in remote sensing image analysis, speech recognition, text classification, the 3 d object recognition and so on have greater application of research results.

Application research in the field of medical research and more. Support vector machine (SVM) is used to classify genes from the genetic data, human gene expression data analysis, gene function determination, categories of protein structure prediction, lung cancer diagnosis, fetal lung image detection, *etc.* In addition, the application of support vector machine (SVM) in industrial engineering research are gradually brought to the attention of the researchers, such as water wheel generator fault diagnosis, the internal combustion engine fault diagnosis, the industrial process of fault diagnosis, industrial object system identification, optimal control, predictive control, etc. In the field of signal processing, such as the financial sector is showing the powerful advantages of support vector machine (SVM).

Support vector although lagging behind in domestic research abroad, but developed rapidly in recent years, many domestic researchers have carried out effective research, made a lot of achievements, to promote the study of domestic support vector machine (SVM) has important significance.

3. Proposed Scheme

Although nuclear techniques with its many advantages, and its theory and application ofalgorithm has made considerable development, but also many direction is worth in-depth study.

(1) Construct a universal method, improved the original linear algorithm[7] to join the inner product, using the kernel function for nonlinear algorithm, increased the application areas of kernel function, and the kernel function in the field of multiple classification, also can explore the fast of the wholealgorithm.

(2) For the categories and the choice of parameters of kernel function, put forward the constructive theory, improve the classification model accuracy and enhance the learning ability and generalization ability of the model.

(3) For large training sets, create time for learning, memory footprint has the feasibility of the algorithm, at the same time reduce the decision error, in many applications, the data is often huge, and the algorithm of the timeliness and the demand of the hardware has certain limitations, for the promotion of large sample is helpful to broaden the scope of practical application, SVM is of great significance.

3.1 Commonly Used Kernel Function is Introduced

Commonly used kernel function is divided into rotation invariant, translation invariant kernel, *etc.* Can write to form such as $K(x,x^{'}) = \mathrm{f}((x \cdot x^{'}))$ kernel function is known as the rotation invariant core, one of them $f : R \to R$. Mainly has homogeneous polynomial kernel function, Sigmoid kernel function.

Polynomial Kernel function (Polynomial Kernel) : $K(x,x^{'}) = (\mathrm{a}(x \cdot x^{'}) + d)^{p}$, , ,parameter α is used to adjust the inner product to prevent a kernel function is too large, parameters d usually take 0 or 1, when take 0, we call it a homogeneous polynomial kernel function, using the kernel function is finally got an *p* order polynomial classifier, is equivalent to the maximum likelihood estimation of p order polynomial.,,

Sigmoid kernel function: $K(x,x^{'}) = \tanh(\sigma(x \cdot x^{'}) + c)$, Sigmoid kernel function is also known as the kernel function of neural network is derived from its produced a two layer

perceptron neural network, without the artificial given number of nodes in the hidden layer, as a result of the SVM to solve is convex programming problem, the solution must be global optimal solution, to a certain degree of photogenic results for direct use of neural network has better generalization ability.

3.2 The Structure of the Kernel Function

Commonly used kernel function is not conform to the study of the characteristics of the system accurately, can according to need to construct the kernel function, the use of a priori knowledge about two sample is similar to first determine the map, use this map to build kernel function; Can also according to the geometric characteristics of kernel function is used as the basis for choosing kernel function. The first case known as similarity criterion σ Such as distance and vector length or the angle,The name of their corresponding mapping, respectively $\phi(x) = x$, $\phi(x) = \|x\|$ or $\phi(x) = x / \|x\|$, kernel function is $K(x, x') = (x \cdot x')$ 、 $K(x, x') = \|x\| \|x'\|$ 、 $K(x, x') = (x, x') / (\|x\| \|x'\|)$.

Kernel function can also according to the focus point of drawing is divided into global kernel function (global kernel fonction) draw kernel function (local kernel function), when you need to distance in detail to distinguish the sample points, we use global kernel function; On the contrary, when we should be focusing on the analysis of distance is close, we use the local kernel function, generally speaking, the learning ability of local kernel function is higher than the global kernel function, and the generalization ability of local nuclear than global nuclear.

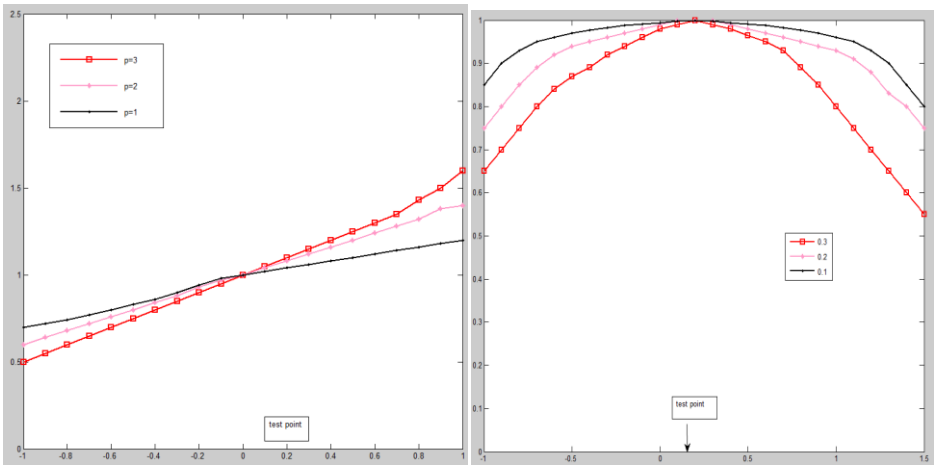

Figure 1. Global Kernel Function (left) and Local Kernel Function (right)

Polynomial kernel is a common global kernel function, for polynomial kernel parameter is larger, the higher the complexity, ability to learn stronger, in the left picture out of the variable dimension Figure 1 to 1,kernel function $K(x, 0.2) = ((0.2 \cdot x) + 1)^p$ for different parameters p ,The change of the function value, can be seen from the polynomial kernel function graph structure, it expanded the larger differences, originally function value tending to zero speed slower, global (quite a number of points) has an effect for kernel function.

Gaussian radial basis kernel function is a common local kernel function, the gaussian RBF [8]kernel parameter, the greater the significant influence of kernel function is the

sample range, the greater the ability to learn, the more strong, in figure 1 of the right to draw out the variable dimension is 1, kernel function $K(x,0.2)=\exp(-\frac{\|x-0.2\|^2}{2\sigma^2})$ for different parameters,function value change, from the graphical structure point of view, the shape is similar to the bell-shaped, it expanded the similar point of difference, and reduce the influence of the distance between the sample points far, front, for example, the condition of positive definite nucleus is another popular local.

3.3 The Selection of Model Parameters

Before selecting a model, we first establish a standard for accurate representation of a model is good, and have much good, for the usual two classification model, we use the most evaluation index is accuracy, speed, robustness to noise data, *etc.*

For a particular field of study, the use of prior knowledge selection model can improve the accuracy of model and parameters, reduce the complexity of training algorithm, are often encountered in the process of data mining is not system research field, we haven't come to the conclusion that effective prior knowledge, we can according to the distribution characteristics of the training sample selection of kernel function and parameters.

Parameter evaluation method can not only through learning ability evaluation of the training sample itself, and that the original empirical risk minimization principle, completely ignored the generalization ability of the calculation. We choose the training sample data, to verify the decision function of the time, so use test methods to choose parameters in training, the training samples are divided into two parts, one more part as new training set, the rest as a validation test parameters, advantages and disadvantages of the training set, the decision function in the original training set to produce as the choice of parameters according to the test set correctness.

We introduce the K fold cross-validation statistics (K fold Cross-validation) method, the training set K two mutually disjoint evenly divided into two subsets, K step iterative training and testing, each time a subset as the test set, as a training set, the rest of the sample after the decision function calculating the wrong the number of points, the final summary to calculate all error point number and the ratio of the total number of training set, this ratio is called K cross validation error, generally it is recommended to use 10 fold cross-validation, because of its overall low algorithm complexity, bias and variance.

One method (Leave-one-out, LOO) [9] is when K to take the training sample size of special K folding cross validation, each iteration only set aside a sample as a test set, K fold cross-validation error in one method and can be referred to as LOO error, error of LOO can approximate estimates expected risk.

If each fold in the proportion of sample classification is in accord with the original sample to K fold cross-validation is also known as layered Cross validation (Stratiffied Cross-validation), if only requires consistent proportion, without the need to each sample used, can use the back of sampling, this Method is called self-help Method (the Bootstrap Method), is Efron of statistical sampling Method is put forward in 1979; When extracting the sample set as well as the size of the original sample set, we called 0.632 self-help method, because in general, the sample size is large, with 63.2% of the original data by probability calculation is used as a method of self-help training sets, 36.8% of the remaining samples test set.

Trial and error method, according to the law of qualitative, the greater the value C, then for the training set error rate is less, but generalization ability is poor, so the accuracy

of the tests increase gradually, then because of the reduced number of support vector, precision down gradually. First give values of initial parameters between the model and the kernel function, and ongoing testing adjustment, until the test precision meet the requirements, usually, the bigger the scale of training data, parameter selection, the less influence on the accuracy of the model.

Grid search method [10], to choose all the parameters are given scope and step length, divided by step length range of node, after together make up a grid, the grid for the combination of the parameters on the optimal node, in search for groups of related parameters, can use the parallel algorithm increase the speed.

Optimization method is to find the optimal parameters in the space, but the direction is not fixed, but through the first training all each iteration solution of quadratic programming, and then put the solution into the objective function, the search parameters on the objective function of the negative gradient direction, according to the step length after updating parameters and model, continue to the next round of iteration, until the accuracy of results to meet certain parameters or no longer change over.

Cheng-hsuan li for a particular training set to determine the parameters σ of RBF kernel function, because RBF kernel function to map the input space to feature space of the unit sphere, he will nuclear function values as training samples in the feature space connected with the origin of the vector Angle cosine function value, when the two vectors is the same category, kernel function value incline to 1, the Angle of 0. When they are for different categories, the value of the kernel function tends to zero, the Angle is 90, W=all categories within the kernel function, the sum as a parameter to maximize the W.

And kernel parameter optimization heuristic intelligent algorithm, genetic algorithm is a simulated Darwinian evolution which are frequently used algorithm of natural selection in thought, the candidate solution set by coding for population, after repeated mutation and the fitness of the selection after a generation of population, until meet the termination conditions, choosing the best individual adaptability.

4. The Experimental Results and Analysis

4.1 Experimental Platform and Data

We in 2.2 GHz processor, 2 gb of memory, install Windows on a computer using matlab calls LIBSVM programming algorithm and the result analysis, by comparing the proposed algorithm and the comparison between results of the original algorithm,Validation choose $1/(n_{sv}+1)$ combination of nuclear parameter α ,training accuracy and the ROC is better than other values, but also to verify the combination kernel function is relative to a single kernel function has high robustness.

This experiment selects the database of sonar data sets, there are a large number of published articles based on the data set, is a common algorithm of machine learning test training data set, data set content for the metal cylinder and roughly cylindrical rock two types of sonar echo data, collected by the university of California Sejnowski, and space technology center of Gorman with signal joint company to improve. Frequency of signal is the sound of the rising frequency modulation pulse, by measuring and integrated in different periods of energy, get the size of the energy in a particular band, each observation sample recorded the sonar of received signals, by a group of 60 represents the frequency from low to high values of variables, each digital values range from 0 to 1, accurate to four decimal places, a total observed from different angles and under the condition of 111 metal cylindrical samples and 97 rock samples. Category attributes

contain not only the classification of real samples, also recorded the azimuth Angle of sonar signal size relations, but don't know the exact point of view.

Website of experts is given using the original sonar signal accuracy was 88% - 88% of human judgment, because the experts may be used to some of the original data information, this part of the information is likely to generate the data set out. Gorman and Sejncnvski using BP neural network to get the best accuracy from 84.7% to 84.7%, according to the nearest neighbor also receive a 82.7% probability of correct classification.

4.2 The Experiment Results Analysis

We according to the division of data provider Gorman and Sejnowski, they shall, in accordance with the azimuth evenly selected training set and test set, a total of 208 data of 49 metal cylindrical specimen and 55 rock samples as training set, and the remaining 62 metal cylindrical specimen and 42 rock observation samples as test set. Because sonar data set in addition to the class attribute for the "M" and "R", respectively "metal" and "stone", it is stated as some numerical value instead of 1 s and 0 s, data quality is good, so no other data preprocessing, we chose to use combination kernel function, $1/2\sigma^2$, $\alpha K_{global}(x,x^{'})+(1-\alpha)K_{local}(x,x^{'})$, $\alpha \geq 0$, $K_{global}(x,x^{'})=(\alpha(x\cdot x^{'})+1)^3$, $K_{global}(x,x^{'})=\exp(-\left\|x-x^{'}\right\|/2\sigma^2)$.

Table 1. A Different α Test Effectiveness and Efficiency

Theserial number	Combination coefficient α	The parameters of the nuclear $(\alpha, 1/2\sigma^2)$	The number of iterations	Number of SV	Number of BSV	accuracy
1	1	$\alpha=1$	1000	68	0	88.46%(92/104)
2	0.8	(1, 1)	1000	62	0	88.46%(92/104)
3	0.72	(1, 1)	1000	64	0	88.46%(92/104)
4	0.5	(1, 1)	1000	66	0	88.46%(92/104)
5	0.1	(1, 1)	900	64	1	89.42%(93/104)
6	1/61	(1, 1)	400	70	12	91.35%(95/104)
7	0.0000001	(1, 1)	150	98	40	80.77%(84/104)
8	0	(1, 1)	150	92	40	80.77%(84/104)

Can be found from Table 1, in the case of nuclear parameters unchanged, with the α decrease of the combination of nuclear factor, the number of iterations decline accelerated the speed of training,at the same time accuracy rising up to $\alpha=1/(n_{sv}+1)$, α continue to decrease, accuracy, kernel function degradation for local kernel function, thus, combination of nuclear factor $\alpha=1/(n_{sv}+1)$ good training speed and accuracy were obtained.

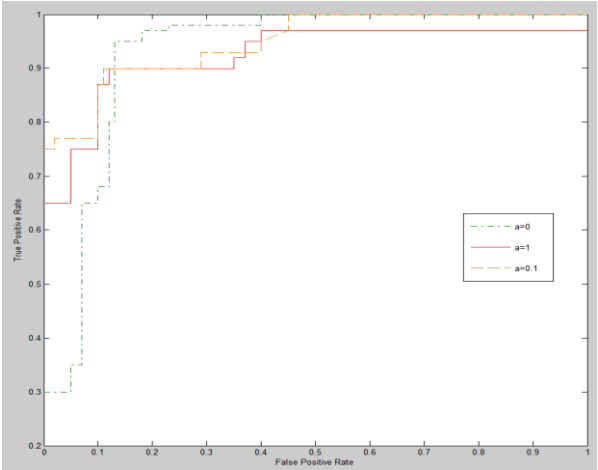

Figure 2. Different the Receiver α Operating Characteristic Curve

Figure 2 shows in the case $1/2\sigma^2=1$, according to the different α draw the receiver-operating characteristic curve, also calls the susceptibility curve, we can find that at the time $\alpha=0.1$, curve is the most close to the top left corner area, the best effect; At the same time, we calculated the area under the curve (AUC) in the size of the more accurate judgment α from the numerical value of the good or bad, α from 1 to 0 were 0.9174, 0.9251, 0.9370, 0.9174, can also be found at the time $\alpha=1/61$, the best training effect.

Table 2. Different Kernel Functions of Robustness

Theserial number	Combination coefficient α	The parameters of the nuclear $(\alpha,\ 1/2\sigma^2)$	The number of iterations	Number of SV	Number of BSV	accuracy
1	1	$\alpha=1$	1122	65	0	88.46%(92/104)
2	1	$\alpha=0.1$	184	70	65	82.69%(86/104)
3	1	$\alpha=0.01$	54	95	91	40.38%(42/104)
4	1/61	$(1,\ 1)$	464	70	15	91.35%(95/104)
5	1/61	$(0.1,\ 1)$	140	93	45	80.77%(84/104)
6	1/61	$(0.01,\ 1)$	132	94	48	80.77%(84/104)
7	0	$1/2\sigma^2=1$	139	94	46	80.77%(84/104
8	0	$1/2\sigma^2=0.1$	55	91	98	72.12%(75/104)
9	0	$1/2\sigma^3=0.01$	53	96	94	40.38%(42/104)

The three sets of data in Table 2 shows that, compared with a single kernel function when using combination kernel function especially $\alpha=1/(n_{sv}+1)$, accuracy is not sensitive to the choice of parameters, to reduce the effects of nuclear parameter selection, and according to LIBSVM software for first kernel function parameters α and $1/2\sigma^2$ the reciprocal of the samples for the default choice, in this case is the equivalent of 0.01 (1/104), the accuracy is poor.

5. Conclusion

Nuclear techniques using kernel function to replace the sample space to the high dimensional feature space transformation and the inner product of high-dimensional

feature space, can need not explicitly transformation is given, at the same time, simplify the operation steps. Introduces the commonly used kernel function and its characteristics, kernel function is divided into local and global nuclei and rotation invariant nucleus and translation invariant. In this paper, some methods about how to construct the kernel function is introduced as a model to find suitable kernel function is to provide some reference strategies. According to the nature of the kernel parameter evaluation standards of classification model and expounds the kernel parameter selection of several commonly used methods.

In the second part the article on the combination of global and local nuclear nuclear coefficient of determination, is put forward using kernel function is affected by the number of support vector as the judgment factors, namely the coefficient $1/(n_{sv}+1)$ for global nuclear, and through the data verify that the idea of combination selection of nuclear factor has a certain guiding role.

References

[1] C. Romero and S. Ventura, "Systems, Man, and Cybernetics, Part C: Applications and Reviews", EEE Transactions on, **(2010)**.
[2] E. Byvatov , "Support vector machine applications in bioinformatics", Schneider G. Applied Bioinformatics, **(2003)**.
[3] D. E. Goldberg, J. H. Holland, "Genetic Algorithms and Machine Learning", Machine Learning , vol. 3, no. 2-3, **(1988)**, pp 95-99.
[4] Y. Xi, C. Y. Gao and Z. Chen, "Advanced nuclear analytical techniques form et al loproteomic Journal of Analytical Atomic Spectrometry", **(2007)**.
[5] D. Corfield, B. Schölkopf and V. Vapnik, "Falsificationism and Statistical Learning Theory: Comparing the Popper and Vapnik-Chervonenkis Dimensions", Journal for General Philosophy of Science, vol. 40, no. 1, pp. 51-58,
[6] C. He, M. Liu, Z. X. Liao and B. Shi, "A learning-based target decomposition method using Kernel KSVD for polarimetric SAR image classification", EURASIP Journal on Advances in Signal Processing, **(2012)**, pp. 159.
[7] R. Zhu, Y. F. Zhao and Y. Z. Li, "Optimal linear precoding for opportunistic spectrum sharing under arbitrary input distributions assumption", EURASIP Journal on Advances in Signal Processing, **(2013)**, pp. 59.
[8] J. Lu, "Evolving Gaussian RBF network for nonlinear time series modelling and prediction Electronics Letters", vol. 34, no. 12, **(1998)**.
[9] Z. X. Yang and Y. J. Tian, "Leave-one-out bounds for support vector ordinal regression machine", Neural Computing and Applications, vol. 18, no. 7, **(2009)**.
[10] F. E. McNeil, "Grid search: an innovative method for the estimation of the rates of lead exchange between body compartments", Journal of Environmental Monitoring, no. 3**, (2005)**.

Author

Wang Chunhong received the B.Eng degree in Computer Science and Technology from Inner Mongolia Normal University(IMNU) and the M.Eng degree in Computer Application Technology from Inner Mongolia University(IMU)， CHINA in 2003 and 2010 respectively. She is currently director of the Department of Computer Science, Inner Mongolia University of Finance and Economics.She is currently researching on Computer Application Technology, Computer network and Data Mining.