The Classification Method Based on Hyper Surface

Qing He [1], Zhong-Zhi Shi [1], Li-An Ren [2]

1. The Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080
2. Graduate College of University of Science and Technology of China, Beijing 100039
e-mail: {heq,shizz}@ics.ict.ac.cn

Abstract * -The main idea of SVM, i.e. Support Vector Machine, is mapping nonlinear separable data into higher dimension linear space where the data can be separated by hyper plane. Based on Jordan Curve Theorem, a general classification method HSC, Classification based on Hyper Surface, is put forward in this paper. The separating hyper surface is directly made to classify large database. The data are classified according to whether the intersecting number is odd or even. It is a novel approach which has no need of either mapping from lower dimension space to higher dimension space or considering kernel function. It can directly solve the nonlinear classify problem. The experiments show that the new method can efficiently and accurately classify large data.

Key words: Support vector machine hyper surface Jordan Curve Theorem

I. INTRODUCTION

The aims of machine learning are to acquire new knowledge and reuse knowledge. Machine learning is the basic problem in artificial intelligence. Moreover, learning is regarded as intelligence action. Pattern recognition, function approximation and probability density evaluating are data learning. Classification is one of the important problems in machine learning.

Vapnik and his research group have studied machine learning based on finite samples since sixtieths in the last century. A complete theory, Statistical Learning Theory, has been established until ninths in the last century [1]-[2]. Moreover, a new universal learning algorithm SVM, Support Vector Machine, has been developed. It is an especially efficient classification algorithm for finite, nonlinear and high dimension data. SVM maps the nonlinear data to a higher dimension linear space where the data can be linear classified by hyper plane. The mapping is a nonlinear mapping defined by inner product function. So a lot of repeat inner product computations of m -matrix cannot be avoided, where m is the number of samples. It is almost impossible to classify large data more than 4000 samples by using PC [13].

In 1999, Ling Zhang and Bo Zhang proposed a geometry classification method[3]. In this method the original input space is transferred into a quadratic space by using global project function, and the well know point set covering method has been applied to perform partition of data in the transformed space. Moreover, a feasible covering-deletion design algorithm is present. The method solve covering problem in distance space instead of quadratic optimization problem in SVM. They have the same idea in the meanings that classify data by transforming samples into higher dimension space [2] [3].

Vapnik thinks that his works on machine learning during ninths in the last century is a regression to the times of apperceive machine. If his work is regarded as an effect to overcome the performance disadvantage of apperceive machine neural network then we can say that Vapnik make a great contribution to machine learning.

Because apperceive machine has linearity performance, it is impossible to solving nonlinear optimization problem, but the algorithm is pretty simple. How to solve the nonlinear classification problem by using the principle of apperceive machine is an important problem.

In the past, many means and special technology have been used to solve the problem. During the sixths in the last century Widrow and Hoff put forward a neural network Adaline constructed by adaptive linear implements and Madaline[11] constructed by many Adalines. These are attempt to solving the nonlinear optimization problem by using many hyper planes. The idea is very important, but how to find these hyper planes is a problem that has not been solved up to now. In this paper, the separating hyper surface is directly made to classify large database according to whether the intersecting number is odd or even. It is a novel approach that has no need of either mapping from lower dimension space to higher dimension space or considering kernel function. It can directly solve the nonlinear classification problem. The experimental reports show that the new method can efficiently and accurately classify large database.

II. THE CLASSIFICATION METHOD BASED ON SEPARATING HYPER SURFACE

In fact, SVM maps the data into higher dimension linear space where the data can be separated by hyper plane. On the contrary, the inverse mapping is a reducing dimension mapping. The separating hyper plane in higher dimension space is changed into separating hyper surface in the lower dimension space. It shows that SVM indirectly solve nonlinear

* This work is supported by the National Science Foundation of China (No.60173017,90104021)and the Nature Science Foundation of Beijing (No.4011003)

problems.

Ling Zhang and Bo Zhang put forward a universal classification method to solve the nonlinear problem based on the neighborhood principle. The main ideas of their work are the following. The original input space is transferred into a quadratic space and the well-know covering method of point set is applied to classifying the data in the quadratic space. At the same time, the simplicity of the M-P model and its corresponding network still maintain. There has no need of either increasing the complexity of network's structure or considering kernel function. A feasible covering-detection design algorithm is presented as well. Moreover, computer simulation results show that the neural network is quit efficient. But the method also use increasing dimension mapping. So it is can be regard as looking for separating hyper surface in the lower dimension space.

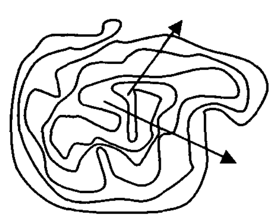
It is an important problem that how to find a classification method, which has no need of either increasing dimension mapping or computing a large of matrix products.

2.1 Jordan Curve Theorem

Jordan Curve Theorem is the theoretic base of HSC. Jordan Curve Theorem is stated as following:

Jordan Curve Theorem. Let X be a closed set in R^3. If X is homeomorphic to a sphere S^2, then its complement $R^3 \setminus X$ has two connected components, one bounded, the other unbounded. Any neighborhood of any point on X meets both of these components.

Jordan Curve Theorem show that each two-orient closed surface transformed from sphere by a series of conscious transforms separates three dimension space into two regions, one called inside, the other called outside. The surface called separating hyper surface can be used to classifying data. For any given point, how to determine whether it is inside or outside about the separating hyper surface is the first important problem.



2FIGURE 2-1 CLASSIFICATION THEOREM

Classification Theorem. Let X be a closed set in R^3. If X is homeomorphic to a sphere S^2, then its complement $R^3 \setminus X$ has two connected components, one called inside, the other called outside. For any $x \in R^3 \setminus X$, the point x is inside of $X \Leftrightarrow$ the intersecting number between any radial from x and X is odd; The point x is outside of $X \Leftrightarrow$ the intersecting number between any radial from x and X is even.

Theorem (Jordan Curve Theorem in High Dimension Space). Suppose that $X \subset S^n$ is homeomorphic to a sphere S^m, then $m \le n$, otherwise $X = S^n$. If $m < n$, then the homology group of $S^n \setminus X$ is

$$H_k(S^n \setminus X) \cong \begin{cases} Z \oplus Z, \text{if } m=n-1 \text{ and } k=0, \\ Z, \text{if } m<n-1 \text{ and } k=0, \\ 0, \text{otherwise.} \end{cases}$$

Specially, if $m = n - 1$, then $S^n \setminus X$ composed by two connected components. Moreover, if $m < n - 1$, then there exists only one connected component.

Based on the Jordan Curve Theorem, space can be separated by two-sided surface that is homeomorphic to sphere. For any given point, whether it is inside of the separating or outside depends on that the intersecting number between the separating hyper surface and the radial from the point is odd or even. This classification method is a direct and convenience method. But how to construct the separating hyper surface is an important and difficult problem. Separating hyper surface may be consisting of more than one local hyper surface. In the following, a new approach is given.

2.2 The Construction of Separating Hyper Surface

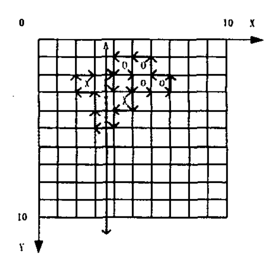


FIGURE3.1 BASIC ALGORITHEM

Based on Jordan Curve Theorem, we put forward the following classification method based on separating hyper surface.

Step1. Let the given samples distribute in the inside of a rectangle region.

Step2. Divide the region into some smaller regions, where there is only at most one sample.

If there is some smaller region in which there are more than one point, then repeat Step2 until in each smaller region there is only at most one sample.

Step3. Label each region according the class of inside sample then the frontier vectors of the smaller region and the class vector form a string.

Step4. Combine the connected region of the same class and obtain a separating hyper surface then saved it as a string.

Step5. Input a new sample and calculate the intersecting number of the sample about separating hyper surface. This can be done by drawing a radial from the sample, then according to whether the intersecting number between the radial and the separating hyper surface is even or odd decide the class of the sample.

III. TESTING RESULTS

Two-Spirals Discrimination Problem [3]: Two spirals K_1 and K_2 (in polar coordinates)

$$\begin{matrix} K1: \rho = \theta \\ K2: \rho = \theta + \pi \end{matrix} \qquad \frac{\pi}{2} \le \rho \le 8\pi . \qquad (3.1)$$

How can the two spirals are separated in two-dimension space. Based the above idea we design a new approach to solve the problem.

1)*Learning Process:* Let the training samples distribute in a region called sample region. The training samples are divided into two classes. Let X denote one class and O denote another class. Label the frontier for each unit. Remerge the frontiers of the same class regions and save it as a string.

2)*Local Elaborate Strategy:* If there are different training samples in the same unit lattice then divide the unit again and do the same things for all units. So a separating hyper surface formed and denoted by a string.

3)*Classifying Process:* If an unknown sample loads the sample region, then draw a radial from A. Calculate the intersecting number of the radial and the separating hyper surface. The class of the sample is decided according to whether the intersecting number is odd or even.

If the region has been divided, then standardize the sample's coordinate and continue classify in divided region.

3.2.1 Experience in Large Database

1) Training results list in table 3 –1

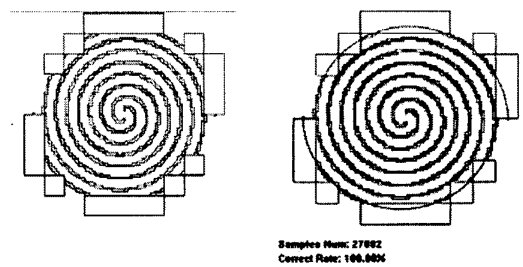


FIGURE 3-2 TRAINING RESULTS, SEPARATING HYPER SURFACE AND THE COVER MAP

TABLE3-1 THE TRAINING RESULT OF LARGE DATA

The Number of Training Samples	Training Time	Classification Time*	Recall Rate (%)
10, 800, 000	1h 34m 57s	2h 17m 35s	100.00
22, 500, 002	3h 16m 9s	4h 49m 55s	100.00
54, 000, 000	7h 42m 52s	11h 47m 7s	100.00

*The test set is the self of training samples set.

2) The result of classification SeeTable3 –2.

TABLE3–2 THE TESTING RESULT OF LARGE DATA

The Number of Training Samples	The Number of Testing Samples	Classification Time	Correct Rate (%)
10, 800, 000	22, 500, 002	4h 7m 4s	100.00
22, 500, 002	54, 000, 000	11h 25m 3s	100.00
54, 000, 000	67, 500, 002	14h 37m 6s	100.00

3.2.2 Small-scale Training and Large-scale Classification

TABLE3-3 SMALL-SCALE SAMPLES TRAINING AND LARGE-SCALE TESTING

The Number of Training Samples	The Number of Testing Samples ②	Classification Time	Correct Classification Rate (%)
5, 402	54, 002	41s	99.59
5, 402	540, 000	6m 45s	99.58
27, 002	540, 000	6m 44s	99.98
54, 002	540, 000	6m 47s	100.00
54, 002	5, 400, 000	1h 7m 7s	100.00

② Testing samples is another sample set which is made from the two spirals and ten times more than training samples.

Table3-3 shows that HSC has good generalization ability. HSC is a polynomial algorithm if the same class samples are distributed in finite connected components. In fact the string and the same class training samples need to be saved at the same time for the continuous of classifying hyper surface. But if the sample is very large, level must be controlled, and then the samples that can't influence the constructor of hyper surface will be deleted. So the computing speed is improved. For only the surface which is orthogonal to the radial from the sample is saved, the need of storage source computer is reduced.

Note: The above test results are given under the computing environment as following.

1) Main computer
   processor: Pentium III, 733MHz; memory: 256M;
2) operating system
   Microsoft Access 2000;
3) compile environment
   Visual C++ 6.0, Service Pack 4.

IV. THE COMPARISONS OF RESULTS

In order to show the performance of HSC two-spiral discrimination problem of two-dimension are studied above. The correct rates for all scale training samples are 100%. The classification correct rates for all scale samples are more than 99%.Specially; table3-3 and table 4-5 show that HSC has strong generalization ability. Compared to the results in [5] we can see that the classification of two spirals based on BP fails, and it takes 3000 iterations and only has 89.6% correct classification rate by using the generating-shrinking algorithm presented in [6]. Moreover, the data in this paper is much more than in paper [3].

V. CONCLUSION

In the paper, a new approach for classification based on hyper surface is put forward. This is a general classification method for large nonlinear data base. The new method has following advantages.

A. High Efficient and Accuracy

For large data (10^7), the speed of HSC is very fast. The reason is that the time for saving and extracting hyper surface is very short and the need for storage is very little, which is not the advantage of SVM. Another reason is that the decision process is very easy by using Jordan Curve Theorem.

B. Strong Ability of Generalization

The experiment of training on small scale samples and testing on large scale shows that HSC has strong generalization ability. The higher VC dimension is, the larger confidence domain is. It is a conclusion of statistic learning theory. So the difference between real risk and experimental risk possibly increase. This is the reason of excessive learning problem. So machine learning process is not only minimizing the experimental risk, but also reducing the dimension of VC. But the strategy is not useful in HSC. Because the hyper surface made by linear segmentation function. The function set has infinite VC dimension because the set can separate any more h samples that distribute in anyway. This is show that the bound of generalization given by Vapnik is loose when the VC dimension is too big. Moreover, we can see that the continuousness of the hyper surface is improved as the number of samples increases. This is show that the scale of unit should be larger than the margin of the samples. If the scale of unit is too small then the hyper space may be separated into two parts. Big unit is required in the region where samples scattered distribute. On the contrary, small unit is required in the region where samples densely distribute. But the samples are not always uniform distribution. So the local elaborate division is an important strategy. The local division strategy improves generalization ability and accuracy.

C.Robustness

Though the data noise can not be completely clear, but its effect can be controlled in a local region. If a noise sample locates inside of the hyper surface, then the hyper surface changes into complex hyper surface. In this case the classification theorem is still efficient, the noise may make mistake in classification, but the influence has been controlled in a local small unit.

D. The feature of samples in shape

In fact, HSC can solve the nonlinear classification problem that the samples distribute in any shape in a finite region. The method has nothing to do with the shape of the data, even though the shape is interlock or crisscross. The common condition as other classification methods is that the samples must reflect the feature of data distribution.

Reference

[1] Vapnik V. N. Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. Neural Information Processing Systems, Vol.9. MIT Press, Cambridge, MA

[2] Vapnik V. N. The Nature of Statistical Learning Theory. New York: Springer-Verlag press, 1995

[3] Ling Zhang, Bo Zhang, "A Geometrical Representation of McCulloch-Pitts Neural Model and Its Applications," *IEEE Transactions on Neural Networks,* Vol.10:4, pp.925-929, July 1999

[4] Cortes C, Vapnik V.N. "Support Vector Networks," *Machine Learning,* Vol.20, pp.273-297,1995

[5] S.E.Fahlman and C.Lebierere, "The cascade-correlation learning architecture," *Advances in Neural Information Process Systems,* D.S.Tourdzhy. Ed.San mateo, CA:Kaufmann, vol.2,pp.524-532, 1990.

[6] E.B.Baum and K.J.Lang,"Constructing hidden units using examples and queries," *Neural Information Processing Systems,* R.P.Lippman et al ,Eds. San Mateo , CA:Kaulfman, , Vol.3, pp. 904-910, 1991

[7] Vapnik V. N., Levin E, Le Cun Y. "Mearsuring the VC-Dimension of learning machine," *Neural Computation,* Vol.6, pp. 851-876, 1994

[8] Vapnik V. N. Statistical Learning Theory, J. Wiley press, New York, 1998

[9] Widrow B., Winter R. G. "Layered neural nets for pattern recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing,* Vol. 36: 3, pp. 1109-1118, 1988

[10] Burges C J C. "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery,* Vol. 2:2, 1998

[11] Widrow B., Hoff M. IRE Wescon Convension Record. Part 4, Institute of Radio Eng., New York, pp.96-104, 1960.

[12] William Fulton, Algebraic Topology A First Course, Springer-Verlag press, 1995

[13] Schokopf,B.,Burges,J.C.,Smola,A.J. Advances in Kernel Methods

Support Vector Learning. Cambridge, MA: MIT Press,1999

[14] Schouten T., Suykens J.A.K., De Moor B., " Multi resolution least squares SVM solver", Proc. of the 43rd IEEE Midwest Symposium on Circuits and Systems (MWSCAS2000), Lansing, Michigan, Aug. 2000

[15] cdrom p. Suykens J.A.K., ``Support Vector Machines : a nonlinear modelling and control perspective", *European Journal of Control*, Special Issue on fundamental issues in control, Vol. 7: 2-3, pp. 311-327 ,2001

[16] Lukas L., Suykens J.A.K., Vandewalle J., ``Least squares support vector machines classifiers : a multi two-spiral benchmark problem", Proc. of the Indonesian Student Scientific Meeting (ISSM 2001), Manchester, United Kingdom, pp. 289-292, Aug. 2001