Enclosing Machine Learning

Xunkai Wei*
Department of Aircraft and Power Engineering
University of Air Force Engineering
Xian, Shaanxi 710038
China
voice: +86 136-0108-7369
fax: +86 029-8478-7527
email: xunkai.wei@gmail.com

Yinghong Li
Department of Aircraft and Power Engineering
University of Air Force Engineering
Xian, Shaanxi 710038
China
voice: +86 137-0924-8339
fax: +86 029-8478-7527
email: yinghong_li@126.com

Yufei Li
Department of Aircraft and Power Engineering
University of Air Force Engineering
Xian, Shaanxi 710038
China
fax: +86 029-8478-7527
email: horizon_lyf@hotmail.com

(* Corresponding author)

Enclosing Machine Learning

Xunkai Wei, University of Air Force Engineering, China

Yinghong Li, University of Air Force Engineering, China

Yufei Li, University of Air Force Engineering, China

INTRODUCTION

As is known to us, Cognition process is the instinct learning ability of human being, and this process is perhaps the most complex but highly efficient and intelligized information processing process. For the cognition process of the natural world, human always transfers the feature information to the brain through perception, and then the brain will process the feature information and remember the feature information for the given objects. Since the invention of computer, scientists are always working toward improving its artificial intelligence, and hope one day the computer could have their own genuine intelligent "brain" like the human brain. However, according to the cognition science theory, the human brain can be imitated but cannot be completely reproduced. Thus, to let the computer truly "think" by themselves seems easy yet there is still a long way to accomplish this objective.

Currently, artificial intelligence is still an important and active direction of function imitation of the human brain. Yet traditionally, the Neural-computing and neural networks families are the majority part of the direction (Haykin, 1994). By imitating the working mechanism of human neuron of the brain, scientist found the neural networks computing theory according to experimental progresses such as Perception neurons and Spiking neurons (Gerstner

& Kistler, 2002) in understanding the working mechanism of neurons. For a long time, the related research works mainly emphasize on neuron model, neural network topology, learning algorithm, and thus there are quite flourish large families (Bishop, 1995) such as, Back Propagation neural networks (BPNN), Radical Basis Function neural networks (RBFNN), Self Organization Map (SOM), and various other variants.

Neural-computing and neural networks (NN) families have made great achievements in various aspects. Recently, statistical learning and support vector machines (SVM) (Vapnik, 1995; Scholkopf, & Smola, 2001) draw extensive attention, show attractive and excellent performances in various areas (Li, Wei & Liu, 2004) compared with NN, which imply that artificial intelligence can also be made via advanced statistical computing theory. Nowadays, these two methods tend to merge under statistical learning theory framework.

BACKGROUND

It should be noted as for NN and SVM, the function imitation is from the microcosmic view utilizing the mathematic model of neuron working mechanism. However the whole cognition process can also be summarized as two basic principles from the macroscopical view, i.e. the first is that **human always cognizes things of the same kind**, the second is that **human recognizes things of a new kind easily without affecting the existing knowledge**. These two common principles are easily concluded.

In order to make the idea more clearly, we firstly analyze the function imitation explanation of NN and SVM. The function imitation of human cognitive process for pattern classification via NN and SVM can be explained as follows (Li & Wei, 2005). Given the training pairs (sample features, class indicator), we can train a NN or a SVM learning machine. The training process of these learning machines actually imitates the learning ability of human being.

For clarity, we call this process “cognizing”. Then, the trained NN or SVM can be used for testing an unknown sample and determine the class indicator it belongs to. The testing process of an unknown sample actually imitates the recognizing process of human being. We call this process “recognizing”.

From the mathematic point of view, both these two learning machines are based on the hyperplane adjustment, and obtain the optimum or sub-optimum hyperplane combinations after the training process. As for NN, each neuron acts as a hyperplane in the feature space. The feature space is divided into many partitions according to the selected training principle. Each feature space partition is then linked with a corresponding class, which accomplishes the “cognizing” process. Given an unknown sample, it only detects the partition where the sample locates in and then assigns the indicator of this sample, which accomplishes the “recognizing” process. Like NN, SVM is based on the optimum hyperplane. Unlike NN, standard SVM determines the hyperplane via solving a QP convex optimization problem. They have the same “cognizing” and “recognizing” process except different solving strategies.

Now, suppose we have a complete sample database, and if a totally unknown and novel sample comes, both SVM and NN will not naturally recognize it correctly and conversely prefer to assign a most close indicator in the learned classes (Li & Wei, 2005).

However, this phenomenon is quite easy for human to handle with. If we have learned some things of the same kind before, given similar things we can easily recognize them. If we have never encountered with them, we can also easily tell that they are fresh things. Then under supervised learning of them, we can then remember their features in the brain without changing other learned things.

The root cause of this phenomenon is the learning principle of the NN or SVM "cognizing" algorithm, which is based on feature space partition. This kind of learning principle may amplify each class's distribution region especially when the samples of different kinds are small due to incompleteness. This makes it impossible to automatically detect the novel samples. Here comes the concern: how to make it automatically identify the novel samples like human.

MAIN FOCUS

Human being generally cognizes things of one kind and recognizes complete unknown things of a novel kind easily. So the answer is why not let the learning machine "cognize" or "recognize" like human being (Li, Wei & Liu, 2004). In other words, the learning machine should "cognize" the training samples of the same class regardless of the other classes, so that our intention is focused only on each single class. This point is important to assure that all the existing classes are precisely learned without amplification. To learn each class, we can just **let each class be cognized or described by a cognitive learner**. It uses some kind of model to describe each class instead of using feature space partition so as to imitate the "cognizing" process. Therefore, now there is no amplification occur different from NN or SVM. The bounding boundary of each cognitive learner scatters in the feature space. All the learners' boundaries consist of the whole knowledge of the learned classes. For an unknown sample, the cognitive class recognizer then detects **whether the unknown sample is located inside a cognitive learner's boundary** to imitate the "recognizing" process. If the sample is completely new (i.e., none of the trained cognitive learner contains the sample), it can be again described by a new cognitive learner and the new obtained learner can be added to the feature space without affecting others. This concludes the basic process of our proposed enclosing machine learning paradigm (Wei, Li & Li, 2007A).

Mathematic Modeling

In order to realize above ideas for practical usage, we have to link the ideas with concrete mathematic models (Wei, Li & Li, 2007A). Actually the first principle can be modeled as a minimum volume enclosing problem. The second principle can be ideally modeled as a point detection problem. In fact, the minimum volume enclosing problem is quite hard to solve for samples from arbitrary distribution， and the actual distribution shape might be rather complex for calculating directly. Therefore, a natural alternative is to use regular shapes such as sphere (Fischer, Gartner, & Kutz, 2003), ellipsoid and so on to enclose all samples of the same class with minimum volume to approximate the true minimum volume enclosing boundary (Wei, Löfberg, Feng, Li & Li, 2007). Moreover, the approximation method can also be easily formulated as a convex optimization problem. Thus it can be efficiently solved in polynomial time using state-of-the-art available open source solvers such as SDPT3 (Toh, Todd & Tutuncu, 1999), SEDUMI (Sturm, 1999), YALMIP (Löfberg, 2004) etc. Consequently, the point detection algorithm can be easily concluded via detecting its location inside it or not.

Enclosing Machine Learning Concepts

Using previous modeling methods, we can now introduce some important concepts. Note that the new learning methodology now actually has three aspects. The first aspect is to learn each class respectively, we call it cognitive learning. The second aspect is to detect unknown samples' location and determine its indicator, we call it cognitive classification. While the third aspect is to conduct a new cognitive learning process, we call it feedback self-learning, and the third aspect is for imitation of the character of learning samples of new kind without affecting the existing knowledge. The whole process can be depicted in Fig 1. We now can give following two definitions.

Class Learner. A cognitive class learner is defined as the bounding boundary of a minimum volume set which encloses all the given samples. The cognitive learner can be either a sphere or an ellipsoid or their combinations. Fig 2, Fig 3 and Fig 4 depict the examples of sphere learner, ellipsoid learner, and combinational ellipsoid learner in 2D.

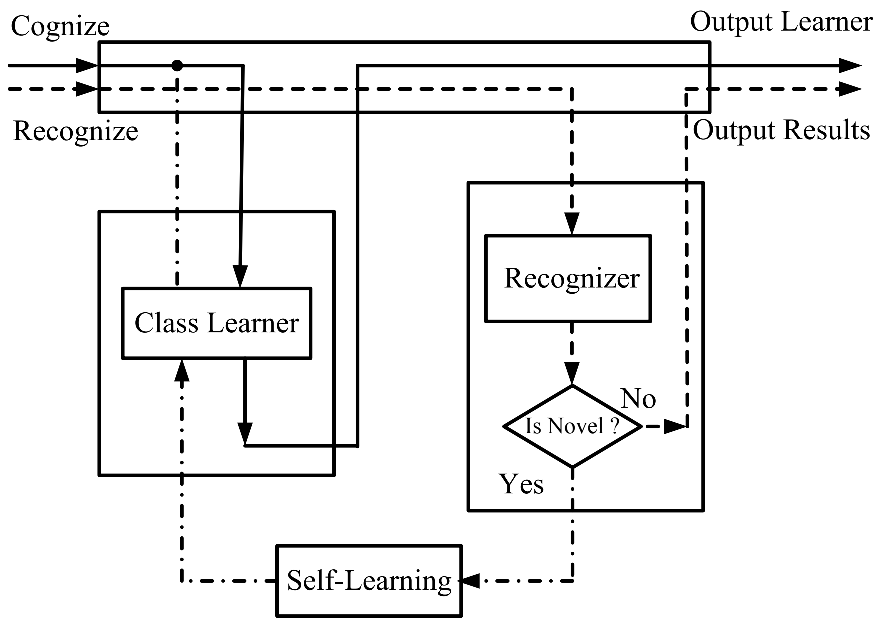


Fig.1 Enclosing Machine Learning Process. The real line denotes the cognizing process. The dotted line denotes the recognizing process. The dash-dotted line denotes the feedback self-learning process.

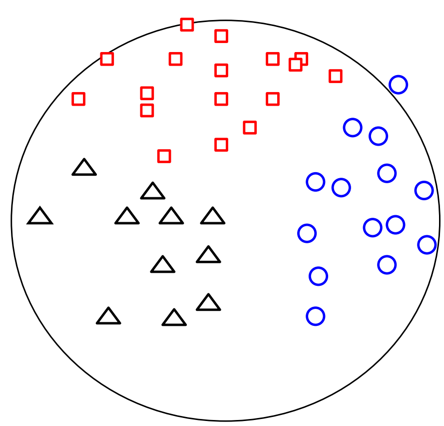


Fig. 2 Sphere Learner

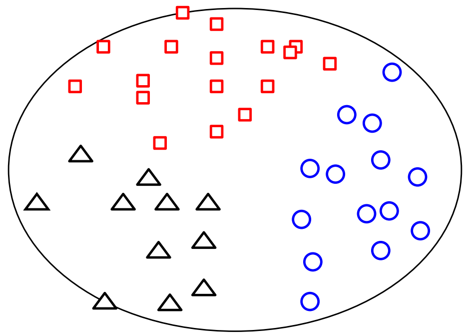


Fig.3 Ellipsoid Learner

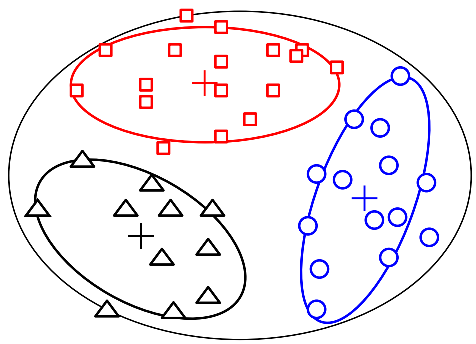


Fig.4 Combinational Ellipsoid Learner

Remarks: As for the above illustrated three type learner, we can see that the sphere learner generally has the biggest volume, and next is single Ellipsoid learner, and the combinational Ellipsoid learner has the smallest volume.

Recognizer. A cognitive recognizer is defined as the point detection and assignment algorithm.

The cognitive learner should own at least following features to get commendable performance:

A. regular and convenient to calculate

B. bounding with the minimum volume

C. convex bodies to guarantee optimality

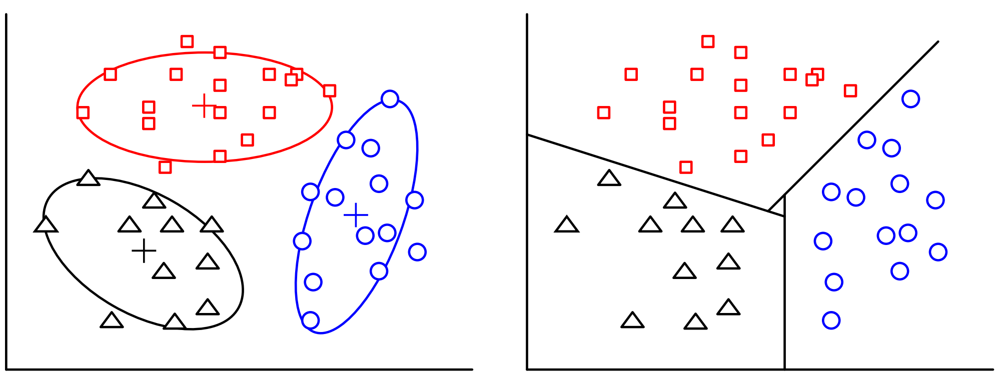
D. fault tolerant to assure generalization performance.

The basic geometric shapes are the best choices. Because they are all convex bodies and the operations like intersection, union or complement of the basic geometric shapes can be implemented using convex optimization methods easily. So we propose to use basic geometric shapes such as sphere, box or ellipsoid to serve as base learner.

The cognitive learner is then to use these geometric shapes to enclose all the given samples with the minimum volume objective in the feature space. This is the most important reason why we call this learning paradigm enclosing machine learning.

Cognitive Learning & Classification algorithms

We first investigate the difference between enclosing machine learning and other feature space partition based methods. Fig.5 gives a geometric illustration of the differences. For the cognizing (or learning) process, each class is described by a cognitive class description learner. For the recognizing (or classification) process, we only need to check which bounding learner the testing sample locates inside.



(a) (b)

Fig.5. A geometric illustration of learning a three class samples via enclosing machine learning vs. feature space partition learning paradigm. (a) For the depicted example, the cognitive learner is the bounding minimum volume ellipsoid, while the cognitive recognizer is actually the point location detection algorithm of the testing sample. (b) All the three classes are separated by three hyperplanes.

But for the partition based learning paradigm, among the learning process, each two classes are separated via a hyperplane (or other boundary forms, such as hypersphere etc.). While among the classification process, we need to check whether it is located on the left side or the right side of the hyperplane and then assign the corresponding class indicator. We can see that the feature space partition learning paradigm in fact amplify the real distribution regions of each class. But the enclosing machine learning paradigm obtains more reasonable distribution region of each class.

In enclosing machine learning, the most important step is to obtain a proper description of each single class of samples. From mathematic point of view, our cognitive class description methods actually are the so-called one class classification method (OCC) (Scholkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001). OCC can recognize the new samples that resemble the training set and detect uncharacteristic samples, or outliers, to avoid the ungrounded classification.

By far, the well-known examples of OCC are studied in the context of SVM. For this problem, One Class Support Vector Machines (OCSVM) (Tax & Duin, 1999) is firstly proposed. The OCSVM first maps the data from the original input space to a feature space $\mathbb{F}$ via some map Φ, and then construct a hyperplane in $\mathbb{F}$ which separate the mapped patterns from the origin with maximum margin. The one-class SVM proposed by Tax (Tax, 2001) is named support vector domain description (SVDD), which seeks the minimum hypersphere that encloses all the data of the target class in a feature space. In this way, it finds the descriptive area that covers the data and excludes the superfluous space that results in false alarms.

However, both OCSVM and SVDD depend on the Euclidean distance, which is often sub-optimal. An important problem in Euclidean distance based learning algorithm is the scale of the input variables. And thus Tax et al (Tax & Juszczak, 2003) proposes a KPCA based techniques to rescale the data in a kernel feature space to unit variance in order to reduce the input variable scale influences to minima. And People proposed to maximize the Mahalanobis distance of the hyperplane to the origin instead, which is the core idea of the One Class Minimax Probability Machine (OCMPM) (Lanckriet, Ghaoui & Jodan, 2002) and the Mahalanobis One Class Support Vector Machines (MOCSVM) (Tsang, Kwok, & Li, S., 2006). Because the Mahalanobis distance is normalized by the covariance matrix, it is linear translation invariant. Therefore, we need not worry about the scales of input variables.

What's more, to alleviate the undesirable effects of estimation error in the covariance matrix, we can easily incorporate a priori knowledge with an uncertainty model and then address it as a robust optimization problem.

Because Ellipsoid and the accompanying Mahalanobis distance own many commendable virtues mentioned above, we proposed to incorporate Ellipsoid and Mahalanobis into class learning. And then currently our main progress towards class learning or "cognizing" is that we proposed a new minimum volume enclosing ellipsoid learner and several Mahalanobis distance based OCC methods. In our previous works, we proposed a QP based Mahalanobis Ellipsoidal Learning Machine (QP-MELM) (Wei, Huang & Li, 2007A) and QP based Mahalanobis Hyperplane Learning Machine (QP-MHLM) (Wei, Huang & Li, 2007B) via solving the dual form, and applications to real world datasets show promising performances. However, as is suggested (Boyd, & Vandenberghe, 2004), if both the primal form and dual form of an optimization problem is feasible, then the primal form is more preferable. Therefore, we proposed a Second Order Cone Programming representable Mahalanobis Ellipsoidal Learning

Machine (SOCP-MELM) (Wei, Li, Feng & Huang, 2007A). And according to this new learner, we developed several useful learning algorithms.

Minimum Volume Enclosing Ellipsoid Learner

In this new algorithm, we summarize several solutions (see Kumar, Mitchell, & Yildirim, 2003; Kumar, P. & Yildirim, 2005; Sun & Freund, 2004). As for the SDP solution, we can directly solve its primal form using Schur complement theorem. As for the lndet solution, we can solve its dual efficiently in polynomial time. As for the SOCP solution (Wei, Li, Feng, & Huang, 2007B), we can also efficiently solve its primal form in polynomial time. We suppose all the samples are centered firstly. So we only give results for minimum volume enclosing ellipsoid center at the origin for this case. But it is straightforward for lift the ellipsoid with center in a d dimension space to a generalized ellipsoid with center at the origin in a $d+1$ dimension space , for more detail, the reader may check the paper (Wei, Li, Feng & Huang, 2007A) for more detail.

Given samples $X \in R^{m\times n}$, suppose $\mathrm{E}(c,\Sigma) := \{x : (x-c)^T \Sigma^{-1}(x-c) \le 1\}$ is the demanded ellipsoid, then the minimum volume problem can be formulated as

$$\begin{aligned} &\min_{A,b} -\ln\det\Sigma^{-1} \\ &s.t. \begin{cases} (x_i-c)^T \Sigma^{-1}(x_i-c) \le 1 \\ \Sigma^{-1} \succ 0 \end{cases} \end{aligned} \tag{1}$$

However, this is not a convex optimization problem. Fortunately, it can be transformed into following convex optimization problem

$$\begin{aligned} &\min_{A,b} -\ln\det A \\ &s.t. \begin{cases} (Ax_i-b)^T (Ax_i-b) \le 1 \\ A \succ 0, \forall i = 1,2,\cdots,n \end{cases} \end{aligned} \tag{2}$$

using matrix transform $\begin{cases} A = \Sigma^{-\frac{1}{2}} \\ b = \Sigma^{-\frac{1}{2}} c \end{cases}$.

In order to allow errors, using Schur Complete theorem, (2) can be represented in following SDP form

$$\begin{aligned} &\min_{A,b,\zeta_i} - \ln \det A + \Theta \sum_{i=1}^{n} \zeta_i \\ &s.t. \begin{bmatrix} I & (Ax_i - b) \\ (Ax_i - b)^T & 1 + \zeta_i \end{bmatrix} \geq 0 \end{aligned} \tag{3}$$

Solving (3), we can then obtain the minimum volume enclosing ellipsoid. Yet, SDP is quite demanded especially for large scale or high dimensional data learning problem.

As for $\mathrm{E}(c,\Sigma) := \{x : (x-c)^T \Sigma^{-1} (x-c) \leq R^2\}$, we can reformulate the primal form of minimum volume enclosing problem as following SOCP form:

$$\begin{aligned} &\min_{R,\xi_i \geq 0,\mathbf{c}} R + \Theta \sum_{i=1}^{N} \xi_i \\ &s.t. \begin{cases} \sqrt{(\mathbf{x}_i - \mathbf{c})^T \mathbf{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{c})} \leq R + \xi_i, \\ R > 0, \xi_i \geq 0, i = 1, 2, \cdots, N. \end{cases} \end{aligned} \tag{4}$$

Accordingly, it can be kernelized as

$$\begin{aligned} &\min_{\mathbf{w},R,\xi_i \geq 0} R + \Theta \sum_{i=1}^{n} \xi_i \\ &s.t. \begin{cases} \sqrt{n} \left\| \mathbf{\Omega}^{-1} \mathbf{Q} (\mathbf{k} - \mathbf{K}\mathbf{w}) \right\|_2 \leq R + \xi_i, \\ R > 0, \xi_i \geq 0, i = 1, 2, \cdots, n. \end{cases} \end{aligned} \tag{5}$$

Where $\mathbf{c}$ the center of the ellipsoid is, R is the generalized radius, n is number of samples, and $\mathbf{K}_C = \mathbf{Q}^T \mathbf{\Omega} \mathbf{Q}$.

So as to obtain more efficient solving method, except above primal form based methods, we can also reformulate the minimum volume enclosing ellipsoid centered at origin as following optimization problem:

$$\begin{aligned} &\min_{U,\zeta_i} -\ln\det\Sigma^{-1} + \Theta\sum_{i=1}^{n}\zeta_i \\ &s.t.\begin{cases} x_i^T\Sigma^{-1}x_i \le 1+\zeta_i \\ \zeta_i \ge 0, \forall i = 1,2,\cdots,n \end{cases} \end{aligned} \tag{4}$$

Where Θ balances the volume and the errors, $\zeta_i \ge 0$ is slack variable.

Actually via optimization conditions and KKT conditions, this problem can be efficiently solved using following dualized representation form:

$$\begin{aligned} &\max_{\alpha} \ln\det\sum_{i=1}^{n}\alpha_i x_i x_i^T - \sum_{i=1}^{n}\alpha_i \\ &s.t.\begin{cases} 0 \le \alpha_i \le \Theta \\ i = 1,\cdots,n \end{cases} \end{aligned} \tag{5}$$

Where α is the dual variable.

We see that (5) cannot be kernelized directly, therefore we need to use some tricks [] to kernelized its equivalent counterpart

$$\begin{aligned} &\max_{\alpha} \ln\det \mathrm{K}^{\frac{1}{2}}\Gamma\mathrm{K}^{\frac{1}{2}} - \sum_{i=1}^{n}\alpha_i \\ &s.t.\begin{cases} 0 \le \alpha_i \le \Theta \\ i = 1,\cdots,n \end{cases} \end{aligned} \tag{6}$$

Where α is the dual variable, $\Gamma := \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{bmatrix}$, $\mathrm{K} := \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_n,x_1) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{bmatrix}$

Multiple Class Classification algorithms

As is pointed out previously, cognitive learning is actually to use minimum volume geometric shapes to enclose each class samples for imitating the learning process of human brain. Thus for multiple class classification problem, a naturally solution is firstly to use minimum volume geometric shapes to approximate each class samples' distribution, and then for giving unknown samples, we only need to check whether they are inside a learner or not. But these are for ideal cases, where no overlaps occur in each single class distributions. When overlaps occur, we proposed two algorithms to handle this case (Wei, Huang & Li, 2007C).

For the first solution, we use a distance based metric, we would like to assign it to the closest class. This algorithm can be summarized as

$$f(x) = \arg \min_{k \in \{1,2,\cdots,m\}} \|x - c\|_{\Sigma} - R \tag{7}$$

Where $\|\bullet\|_{\Sigma}$ denotes Mahalanobis norm.

Another way is to use optimum Bayesian decision theory, and assign its indicator to the class with maximum posterior probability:

$$f(x) = \arg \max_{k \in \{1,2,\cdots,m\}} \frac{P_k}{(2\pi R_k^2)^{\frac{d}{2}}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_k\|_M^2}{R_k^2}\right) \tag{8}$$

where d is the dimension of the feature space and $P_k = \frac{1}{N}\sum_{i=1}^{N} 1_{y_i = k}$ is the prior distribution of class k. According to (8) the decision boundary between class 1 and 2 is given by

$$\frac{P_1 (2\pi R_1^2)^{-\frac{d}{2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_1\|_M^2}{R_1^2}\right)}{P_2 (2\pi R_2^2)^{-\frac{d}{2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_2\|_M^2}{R_2^2}\right)} = 1 \tag{9}$$

And this is equivalent to

$$\frac{\|\mathbf{x}-\mathbf{c}_1\|_M^2}{R_1^2}+T_1=\frac{\|\mathbf{x}-\mathbf{c}_2\|_M^2}{R_2^2}+T_2 \tag{10}$$

Therefore we can give a new decision rule

$$f(x)=\arg\max\left(\frac{\|\mathbf{x}-\mathbf{c}_k\|_M^2}{R_k^2}+T_k\right) \tag{11}$$

where $T_k = d\log R_k - \log P_k$ can be estimated from the training samples.

Remarks. Actually, we also proposed a single MVEE learner based two class classification algorithm (Wei, Li, Feng & Huang, 2007A), which owns both features of MVEE description and SVM discrimination. Then using One Vs One or One Against One, we can also get a multiple class classification algorithm. Except this, we are now working on a multiple class classification algorithm at complexity of a single MVEE based two class classification algorithm, which is expected to obtain promising performances.

Gap tolerant SVM design

Here we briefly review the new gap tolerant SVM design algorithm. This new algorithm is based on the minimum volume enclosing ellipsoid learner for assuring a compact description of all the samples. We firstly find the MVEE around all the samples and thus obtain a Mahalanobis transform. We then use the Mahalanobis transform to whiten all the samples and thus map them to a sphere distribution. Then we construct standard SVMs in this whiten space.

The MVEE gap tolerant classifier design algorithm can be summarized as

Step1, Solve MVEE and obtain Σ and center c

Step2, Whiten data using Mahalanobis transform $t_i = \Sigma^{-\frac{1}{2}}(x_i - c)$ and get new sample pairs $\left(t_i, y_i\right)_{i=1}^n$

Step3, Solve standard SVM and get the decision function $y(x)=\mathrm{sgn}(w^T t+b)$.

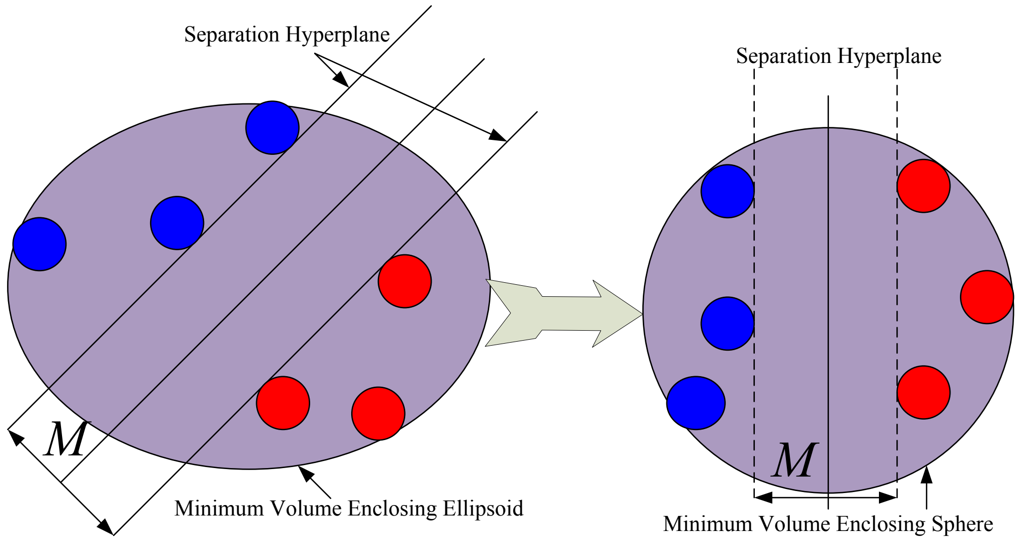


Fig 6 MVEE gap tolerant classifier illustration

Remarks. This algorithm is very concise and has several commendable features worth noting. The classifier designed using this algorithm has less VC dimension compared with traditional ones. Also this algorithm is scale invariant. For more details, the reader should refer to (Wei, Li & Dong, 2007).

FUTURE TRENDS

In the future, more learner algorithms will be developed. Another important direction is to develop set based combinational learner algorithm (Wei, & Li, 2007; Wei, Li, & Li, 2007B). Also more reasonable classification algorithms will be focused. Except theoretical developments, we will also focus on applications such as novelty detection (Dolia, Page, White & Harris, 2004), face detection, industrial process condition monitoring, and many other possible applications.

CONCLUSION

In this article, we introduced enclosing machine learning paradigm. We focused on its concept definition, and progresses in modeling the cognizing process via minimum volume enclosing ellipsoid. We then introduced several learning and classification algorithms based on

MVEE. And we also report a new gap tolerant SVM design method based MVEE. Finally, we give future development directions.

REFERENCES

Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*, 1st edn, Oxford: Oxford University Press

Boyd, S. & Vandenberghe, L. (2004) *Convex Optimization*, 2nd edn, Cambridge: Cambridge University Press

Dolia, A. N., Page, S. F., White, N. M. & Harris, C.J. (2004) D-optimality for Minimum Volume Ellipsoid with Outliers. In: *Proceedings of the Seventh International Conference on Signal/Image Processing and Pattern Recognition (UkrOBRAZ'2004)*, 73-76

Fischer, K., Gartner, B., & Kutz, M. (2003) Fast Smallest-Enclosing-Ball Computation in High Dimensions, In: *Algorithms - ESA 2003*, LNCS 2832, 630-641

Gerstner, W. & Kistler, W. M. (2002) *Spiking Neuron Models: Single Neurons, Populations, Plasticity*, 1st edn, Cambridge: Cambridge University Press

Haykin, S. (1994) *Neural Networks: A Comprehensive Foundation*, 1st edn, NJ: Prentice Hall Press

Löfberg, J. (2004) YALMIP: A Toolbox for Modeling and Optimization in MATLAB. http://control.ee.ethz.ch/ ~joloef

Kumar, P., Mitchell, J. S. B. & Yildirim, E. A. (2003) Approximate minimum enclosing balls in high dimensions using core-sets, *The ACM Journal of Experimental Algorithmics*, 8(1), 1-29

Kumar, P. & Yildirim, E. A. (2005) Minimum volume enclosing ellipsoids and core sets, *Journal of Optimization Theory and Applications*, 126(1), 1–21

Lanckriet, G. Ghaoui, L. E. & Jodan, M. (2002) Robust Novelty Detection with Single-Class MPM. In: Becker, S., Thrun, S. & Obermayor, K. (Eds.): *NIPS 15*

Li, Y. H., Wei, X. K & Liu, J. X. (2004) *Engineering Applications of Support Vector Machines*. 1st edn. Beijing: Weapon Industry Press

Li, Y. H. & Wei, X. K. (2005) Fusion Development of Support Vector Machines and Neural Networks, *Journal of Air Force Engineering University*, 4, 70-73

Scholkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., & Williamson R. (2001) Estimating the Support of a High Dimensional Distribution, *Neural Computation*, 13(7), 1443-1471

Scholkopf, B. & Smola, A. (2001) *Learning with Kernels*, 1st edn, Cambridge, MA: MIT Press

Shawe-Taylor, J., Williams, C., Cristianini, N. & Kandola, J. S. (2002) On the Eigenspectrum of the Gram Matrix and Its Relationship to the Operator Eigenspectrum. In: N. CesaBianchi et al. (Eds.): *Proceedings of the 13th International Conference on Algorithmic Learning Theory (ALT2002)*, LNAI 2533, 23-40

Sturm, J. F. (1999) Using SeDuMi, a Matlab Toolbox for Optimization over Symmetric Cones, *Optimization Methods and Software*, 11-12, 625-653

Sun, P. & Freund, R. M. (2004) Computation of Minimum-Volume Covering Ellipsoids, *Operations Research*, 5, 690-706

Tax, D. M. J. (2001) *One-class Classification: Concept-learning in the Absence of Counter-Examples*, PhD Thesis, Delft University of Technology

Tax, D. M. J. & Duin, R. P. W. (1999) Support Vector Domain Description, *Pattern Recognition Letters*, 20, 1191-1199

Tax, D. M. J. & Juszczak, P. (2003) Kernel Whitening for One-Class Classification, *International Journal of Pattern Recognition and Artificial Intelligence*, 17(3), 333-347

Toh, K.C., Todd, M.J. & Tutuncu, R.H. (1999) SDPT3-a Matlab Software Package for Semidefinite Programming, *Optimization Methods and Software*, 11, 545-581

Tsang, Ivor W. Kwok, James T. & Li, S. (2006) Learning the Kernel in Mahalanobis One-class Support Vector Machines, in *Proceeding of IJCNN 2006 Conference*, 1169-1175

Vapnik, V. N. (1995) *The Nature of Statistical learning theory*, 1st edn, New York: Springer-Verlag

Wei, X. K., Huang, G. B. & Li, Y. H. (2007A) Mahalanobis Ellipsoidal Learning Machine for One Class Classification. In *Proceeding of ICMLC 2007*, Accepted

Wei, X. K., Huang, G. B. & Li, Y. H. (2007B) A New One Class Mahalanobis Hyperplane Learning Machine based on QP and SVD. In *Proceeding of LSMS2007*, DCDIS journal, Accepted

Wei, X. K., Huang, G. B. & Li, Y. H. (2007C) Bayes Cognitive Ellipsoidal Learning Machine for Recognizing Process Imitation. In *Proceeding of LSMS2007*, DCDIS journal, Accepted

Wei, X. K., Li, Y. H & Feng, Y. (2006) Comparative Study of Extreme Learning Machine and Support Vector Machines. In: Wang, J. et al. (Eds.): *Advances in Neural Networks - ISNN 2006*, LNCS 3971, 1089-1095

Wei, X. K., Li, Y. H., Feng, Y. & Huang, G.B. (2007A) Minimum Mahalanobis Enclosing Ellipsoid Machine for Pattern Classification. In: Huang, D.-S., Heutte, L. & Loog, M. (Eds.): *ICIC 2007*, CCIS 2, 1176–1185

Wei, X. K., Li, Y. H, Feng, Y. & Huang, G.B. (2007B) Solving Mahalanobis Ellipsoidal Learning Machine via Second Order Cone Programming. In: Huang, D.-S., Heutte, L. & Loog, M. (Eds.): *ICIC 2007*, CCIS 2, 1186–1194

Wei, X. K. & Li, Y. H. (2007) Linear Programming Minimum Sphere Set Covering for Extreme Learning Machines, *Neurocomputing*, DOI:10.1016/j.neucom.2007.07.021

Wei, X.K., Li, Y.H., and Dong, Y. (2007) A New Gap Tolerant SVM Classifier Design based on Minimum Volume Enclosing Ellipsoid, In China Conference of Pattern Recognition 2007, Accepted.

Wei, X. K., Li, Y. H, & Li, Y. F. (2007A) Enclosing machine learning: Concepts and Algorithms, *Neural computing and Applications*, DOI 10.1007/s00521-007-0113-y 2007.

Wei, X. K., Li, Y. H. & Li, Y. F. (2007B) Optimum Neural Network Construction Via Linear Programming Minimum Sphere Set Covering. In: Alhajj, R. et al. (Eds.): *ADMA 2007*, LNAI 4632, 422–429

Wei, X. K., Löfberg, J., Feng, Y., Li, Y. H., & Li, Y.F. (2007) Enclosing Machine Learning for Class Description. In: Liu, D. et al. (Eds.): *Advances in Neural Networks - ISNN 2007*, LNCS 4491, 424–433

KEY TERMS AND THEIR DEFINITIONS

Enclosing Machine Learning: It is a new machine learning paradigm which is based on function imitation of human being's cognizing and recognizing process.

Cognitive Learner: A cognitive learner is defined as the bounding boundary of a minimum volume set which encloses all the given samples to imitate the learning process.

Cognitive Recognizer: A cognitive recognizer is defined as the point detection and assignment algorithm to imitate the recognizing process.

MVEE Gap Tolerant Classifier: A MVEE Gap Tolerant Classifier is specified by the shape matrix and location of an ellipsoid, and by two hyperplanes, with parallel normals. The set of points lying in between (but not on) the hyperplanes is called the margin set. Points that lie inside the ellipsoid but not in the margin set are assigned a class, $\{\pm 1\}$, depending on which side of the margin set they lie on. All other points are defined to be correct: they are not assigned a class. A MVEE gap tolerant classifier is in fact a special kind of Support Vector Machine which does not count data falling outside the ellipsoid containing the training data or inside the margin as an error.