암진단시스템을 위한 Weighted Kernel 및 학습방법

Weighted Kernel and it's Learning Method for Cancer Diagnosis System

최규석*, 박종진**, 전병찬***, 박인규****, 안인석*****, 하남******

Gyoo-Seok Choi, Jong-Jin Park, Byoung-Chan Jeon, In-Kyu Park, Ihn-Seok Ahn, Ha-Nam Nguyen

요 약 많은 양의 데이터로부터 유용성있는 정보의 추출, 진단 및 예후에 대한 결정, 질병 치료의 응용 등은 바이오인포머틱스(Bioinformatics)분야에서 매우 중요한 문제들이다. 본 논문에서는 암진단시스템에 적용하기위해 support vector machine을 위한 weighted kernel fuction과 빠른 수렴성과 좋은 분류성능을 갖는 학습방법을 제안하였다. 제안된 kernel function에서 기본적인 kernel fuction의 weights는 암진단 학습단계에서 결정되고 분류단계에서 파리미터로 사용된다. 대장암 데이터와 같은 임상 데이터에 대한 실험결과에서 제안된 방법은 기존의 다른 kernel fuction들 보다 더 우수하고 안정적인 분류성능을 보여주었다.

Abstract One of the most important problems in bioinformatics is how to extract the useful information from a huge amount of data, and make a decision in diagnosis, prognosis, and medical treatment applications. This *paper proposes a weighted kernel function for support vector machine and its learning method with a fast convergence and a good classification performance. We defined the weighted kernel function as the weighted sum of a set of different types of basis kernel functions such as neural, radial, and polynomial kernels, which are trained by a learning method based on genetic algorithm. The weights of basis kernel functions in proposed kernel are determined in learning phase and used as the parameters in the decision model in classification phase. The experiments on several clinical datasets such as colon cancer indicate that our weighted kernel function results in higher and more stable classification performance than other kernel functions*

Key Words : *Support Vector Machine, Kernel Methods, Genetic Algorithm*

Ⅰ. Introduction

Support vector machine [1-3] (SVM) is a learning method that uses a hypothesis space of linear functions in a high dimensional feature space. This learning strategy, introduced by Vapnik [2], is a principled and powerful method. In the simplest and linear form, a SVM is the hyperplane that separates a set of positive samples from a set of negative samples with the largest margin. The margin is defined by the distance between the hyperplanes supporting the nearest positive and negative samples. The output formula of a linear case is

$$y = w \cdot x - b \tag{1}$$

*종신회원, 청운대학교 컴퓨터학과
**정회원, 청운대학교 인터넷학과
***정회원, 청운대학교 방송영상학과
****정회원, 중부대학교 컴퓨터학과
*****정회원, 위덕대학교 에너지전기공학부
******정회원, Hanoi National University

where w is a normal vector to the hyperplane and x is an input vector. The separating hyperplane is the plane y = 0 and two supporting hyperplanes parallel to it with equal distances are

$$H_1 : y = w \cdot x - b = +1, \quad H_2 : y = w \cdot x - b = -1 \quad (2)$$

Thus, the margin M is defined as

$$M = 2 / \|w\| \quad (3)$$

In order to find the optimal separating hyperplane having a maximal margin, a learning machine should minimize ||w|| subject to inequality constraints. This is a classic nonlinear optimization problem with inequality constraints. An optimization problem, which can be solved by the saddle point of the Lagrange function, is following

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha_i y_i ([w^T x + b] - 1) \quad (4)$$

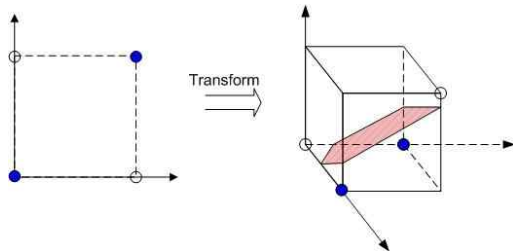where i 0 are Lagrange multipliers.

Fig. 1. An input space can be transformed into a linearly separable feature space by an appropriate kernel function

However, the limitation of computational power of linear learning machines was highlighted in the 1960s by Minsky and Papert [4]. It can be easily recognized that real-world applications require more extensive and flexible hypothesis space than linear functions. Such a limitation can be overcome by multilayer neural networks proposed by Rumelhart, Hinton and William. Kernel function also offers an alternative solution by projecting the data into high dimensional feature space to increase the computational power of linear learning machines.

Non-linear mapping from input space to high dimensional feature space can be implicitly performed by an appropriate kernel function (see Fig. 1). One of the advantages of the kernel method is that a learning algorithm can be exploited to obtain the specifics of application area, which simply can be encoded into the structure of an appropriate kernel function.

Genetic algorithm [5-7] is an optimization algorithms based on the mechanism of natural evolution procedure.

Most of genetic algorithms share a common conceptual base of simulating the evolution of individual structures via the processes of selection, mutation, and reproduction. In each generation, a new population is selected based on the fitness values representing the performances of the individuals belonging to the generation, and some individuals of the population are given the chance to undergo alterations by means of crossover and mutation to form new individuals. In this way, GA performs a multi-directional search by maintaining a population of potential solutions and encourages the formation and the exchange of information among different directions. GA is generally applied to the problems with a large search space. They are different from random algorithms since they combine the elements of directed and stochastic search.

Furthermore, GA is also known to be more robust than directed search methods. Recently, SVM and GA are combined for the classification of biological data related to the diagnosis of cancer diseases and achieved a good performance. GA.

In this paper, we propose weighted kernel function which is defined as the linear combination of basis kernel functions and a new learning method for the kernel function. In the proposed learning method, GA is applied to derive the optimal decision model for the classification of patterns, which consists of the set of the weights for basis kernels in our own kernel. The weighted kernel and the learning method were applied to classify three clinical data sets related to cancer

diagnosis and showed better performance and more stable classification accuracy than single basis kernels.

II. Weighted Kernel Function

A kernel function provides a flexible and effective learning mechanism in SVM, and the choice of a kernel function should reflect prior knowledge about the problem at hand. However, it is often difficult for us to exploit the prior knowledge on patterns to choose a kernel function, and it is an open question how to choose the best kernel function for a given data set. According to no free lunch theorem [1] on machine learning, there is no superior kernel function in general, and the performance of a kernel function rather depends on applications.

In our case, the proposed kernel function is defined as the weighted sum of the set of different basis kernel functions. This kernel function has the form of

$$K_c = \sum_{i=1}^{m} \beta_i \times K_i \qquad (5)$$

where $\beta_i \in [0,1]$ for $i = 1, ..., m$, $\sum_{i=1}^{m} \beta_i = 1$, and $\{K_i \mid i = 1, \cdots, m\}$ is the set of basis kernel functions to be combined. Table 1 shows the mathematical formula of the basis kernel functions used to construct weighted kernel function. It can be proved that (5) satisfies the conditions required for kernel functions by Mercer's the orem [1].

Table 1. Kernels are chosen to experiments in our study

Kernel function	Formula
Polynomial	$(\langle x.y \rangle + 1)^2$
Radial	$e^{(-\gamma \| x-y \|^2)}$
Neural	$\tanh(s \cdot \langle x,y \rangle - c)$

The coefficients β_i play the important role of fitting the proposed kernel function to a training data set. In the learning phase, the structure of a training sample space is learned by our weighted kernel, and the knowledge of a sample space is learned and embedded in the set of coefficients β_i. In the learning phase of our approach, GA technique is applied to obtain the optimal set of coefficients β_i that minimize the generalization error of classifier. At the end of learning phase, we obtain the optimal decision model, which is used to classify new pattern samples in classification phase.

III. The Learning Method

The overall structure for classification procedure based on the proposed kernel and the learning method is depicted in Fig. 2. The procedure consists of preprocessing, learning, and classification phases.

Firstly, in the preprocessing stage, feature selection methods were used to reduce the dimensionality of the feature space of the input data. Also, the training and testing sets consisting of a number of cancer and normal patterns are selected and passed to the learning phase.
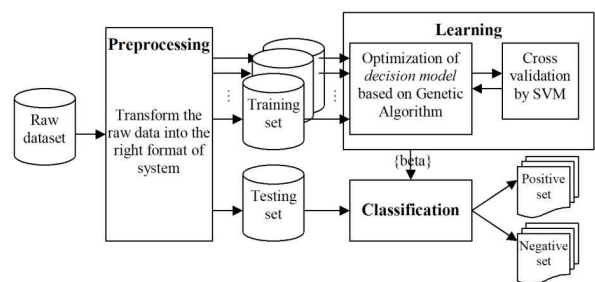

Fig. 2. Overall framework of proposed Method

Secondly, in the learning phase, we applied a learning method based on GA and SVM techniques to obtain the optimal decision model for classification. GA generates aset of chromosomes representing decision

models by evolutionary procedures. The fitness value of each chromosome is evaluated by measuring the accuracy from the classification with SVM containing the decision model associated with the chromosome.

An n-fold validation was used to evaluate the fitness of a chromosome to reduce overfitting [1]. In the evolutionary procedure of GA, only the chromosomes with a good fitness values are selected and given the chance to survive and improve in the further generations. Roulette wheel rule [5] is used for the selection of chromosome in our learning phase. Some of the selected chromosomes are given the chance to undergo alterations by means of crossover and mutation to form new chromosomes. In our approach, one-point crossover is used, and the probabilities for crossover and mutation are 0.8 and 0.015 in turn. The procedure is repeated for a predefined number of times.

At the end of GA procedure, the chromosome with the highest accuracy is chosen as the optimal decision model. Finally, the optimal decision model obtained in the learning phase is used to in SVM for the classification of new samples in the classification phase, and the performance of the model is evaluated against test samples.

IV. Simulation Results and Analysis

In this section, we show the results from the classification based on the model trained by the weighted kernel and the new learning method.

Furthermore, the performance of the classification model with our defined kernel is compared to the performances of the models with other kernels. All the experiments are conducted on a Pentium IV 1.8GHz computer. The experiments are composed preprocessing of samples, learning by GA to obtain the optimal decision model, and classification. For GA, we have used roulette wheel rule for selection method. Our proposed method was executed with 100 chromosomes

for 50 generations. Weighted kernel function and three other kernel functions in Table 1 are trained by GA in learning phase with training set. The three kernel functions are chosen since they were known to have good performances in the field of bioinformatics [4, 6].

We used 5-fold cross validation to measure the fitness to reduce overfitting[4]. The optimal decision model obtained after 50 generations of GA is used to classify the set of test samples.

The colon cancer dataset[8] contains gene expression information extracted from DNA microarrays. The dataset consists of 22 normal and 40 cancer tissue samples and each having 2000 features(Available at: http://sdmc.lit.org.sg/GEDatasets/Data/ColonTumor.zip).

42 samples were chosen randomly as training samples and the remaining samples were used as testing samples.

We chose 50 first features based on t-test statistic.

The Fig. 3 showed the feature importance of the first 15 features in decrease order. Each column represents the logarithm of p-value of all features in the data set that calculated by t-test procedure.
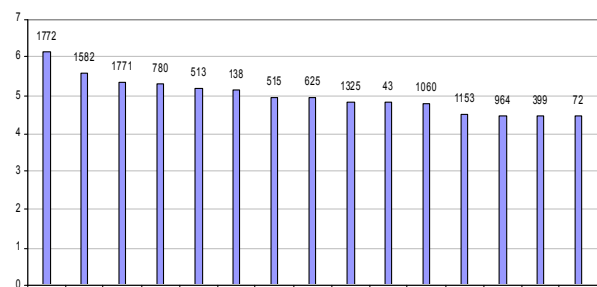

Fig. 3. The -lg(p) value of the first 15 features using the well-known hypothesis testing(t-test)

The value above each column represents the indexes of features in the data set. In case of colon dataset, the proposed method with weighted kernel function also showed much higher accuracy than other kernels (see Table 2).

Table 2. The comparison of average hit rate in classification phase of weighted kernel function case with other kernel functions though 50 trials

	Polyno-mial	Radial	Neural	Weighted Kernel
Pred. Acc.	72.84	82.94	62.41	86.23
S.D.	7.07	6.19	5.74	6.05

The comparison of our experiments and the results of previous studies [9] were depicted in Table 3.

Table 3. The best prediction rate of some studies in case of colon dataset

Type of classifier	Pred. rateS.D.(%)
GA\SVM [12]	84.79.1
Bootstrapped GA\SVM [13]	80.0
Weighted Kernel	86.236.05

Our experiments showed the accuracies comparable to the previous ones, and the standard deviation of the prediction rate for weighted kernel is less than GA\SVM (see Table 3). It is remarked that the new kernel and the learning method result in more stable classification accuracies than previous ones.

V. Conclusions

In this paper, we proposed a weighted kernel function by combining a set of basis kernel functions for SVM and its learning method based on GA technique to obtain the optimal decision model for classification. A kernel function plays the important role of mapping the problem feature space into a new feature space so that the performance of the SVM classifier is improved.

The weighted kernel function and the proposed learning methodwere applied to classify the clinical datasets to test their performance. In the comparison of the classifications by our defined kernel and other three kernel functions, the weighted kernel function achieved higher and more stable accuracies in classification phase than other kernels. Thus our kernel function has greater flexibility in representing a problem space than other kernel functions.

REFERENCES

[1] Richard O. Duda, Peter E. Hart, David G. Stork.: Pattern Classification (2nd Edition), John Wiley & Sons Inc., 2001.

[2] Joachims, Thorsten.: Making large-Scale SVM Learning Practical. In Advances in Kernel Methods -Support Vector Learning, chapter 11. MIT Press, 1999.

[3] Bernhard Schökopf, Alexander J. Smola.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), MIT press, 2002

[4] M.L. Minsky and S.A. Papert: Perceptrons, MIT Press, 1969.

[5] Z.Michalewicz.: Genetic Algorithms + Data structures = Evolution Programs, Springer-Verlag, 3 re rev. and extended ed., 1996.

[6] D. E. Goldberg.: Genetic Algorithms in Search, Optimization & Machine Learning, Adison Wesley, 1989.

[7] Melanie Mitchell.: Introduction to Genetic Algorithms, MIT press, fifth printing, 1999.

[8] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine.: Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays, Proceedings of National Academy of Sciences of the United States of American, vol 96, pp. 6745-6750, 1999.

[9] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander.: Molecular

Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, Science, vol. 286, pp. 531 - 537, 1999.

저자 소개

최 규 석(종신회원)

• 제8권 6호 참조
• 현 청운대학교 컴퓨터학과 교수
<주관심분야 : 이동통신, 인공지능, 인공생명, 지능형 교통체계(ITS), 이동 컴퓨팅>

전 병 찬(정회원)

• 제8권 6호 참조
• 현 청운대학교 방송영상학과 조교수
<주관심분야 : 컴퓨터구조, 홈 네트워크, 모바일, 마이크로프로세서>

안 인 석(정회원)

• 1987년 연세대학교(전기공학과)공학사
• 1989년 연세대학교(전기공학과)공학석사
• 1997년 연세대학교(전기공학과)공학박사
• 1989녀2월~1999년2월 ㈜삼성전자 생산기술연구소 연구원
• 현 위덕대학교 에너지전기공학부 교수
<주관심분야: 인공지능, 제어시스템설계, 임베디드시스템>

박 종 진(정회원)

• 제9권 1호 참조
• 현 청운대학교 인터넷학과 교수
<주관심분야 : 지능시스템, 임베디드시스템, 인터넷 프로그래밍>

박 인 규(정회원)

• 제9권 1호 참조
• 현 중부대학교 컴퓨터학과 교수
<주관심분야 : 영상처리, 네트워크>

Ha-Nam Nguyen(정회원)

• BA degree in Honoi National University(2000년)
• MS degree in Chungwoon University(2002년)
• PhD in Hankong University(2007년)
• Currently Lecturer in Honoi National University.
<주관심분야 : AI, Data Mining, Machine learning, Soft computing>