Kernelized Radial Basis Probabilistic Neural Network for Classification of River Water Quality

Lim Eng Aik[a], Zarita Zainuddin[b]

[a]*Institute of Engineering Mathematic*
Universiti Malaysia Perlis, 02000 Kangar, Perlis
Tel : 04-9855845
E-mail : ealim@unimap.edu.my

[b]*School of Mathematical Science*
Universiti Sains Malaysia, Minden, Penang
E-mail : zarita@cs.usm.my

ABSTRACT

Radial Basis Probabilistic Neural Network (RBPNN) demonstrates broader and much more generalized capabilities which have been successfully applied to different fields. In this paper, the RBPNN is extended by calculating the Euclidean distance of each data point based on a kernel-induced distance instead of the conventional sum-of squares distance. The kernel function is a generalization of the distance metric that measures the distance between two data points as the data points are mapped into a high dimensional space. Through comparing the four constructed classification models with Kernelized RBPNN, Radial Basis Function networks, RBPNN and Back-Propagation networks as intended, results showed that, model classification on River water quality of Langat river in Selangor, Malaysia by Kernelized RBPNN exhibited excellent performance in this regard.

Keywords
Kernel function, Radial Basis Probabilistic Neural Network, Water Quality

1.0 INTRODUCTION

In recent years, water quality management and assessment model is a very popular and important research topic. The so-called management and assessment means to use the related water information to evaluate the level of water quality of a river. However, the water quality status of a river is usually what the public concerns about and is usually the most important key for the healthy water supply to household (Ginoris *et al*, 2007). Therefore, if the water quality of a river can be predicted much earlier, the hazardous causes to the public can be greatly reduced.

In the past literature, some researchers use conventional statistical model for model construction, for example, the Discriminant Analysis of (Altman, 1968), the Probit of (Ohlson, 1980). However, in recent years, some researchers found that the model constructed by data mining technique has better prediction accuracy than that of the conventional statistical model; some of them are, for example, the Back-Propagation Networks model of (Odom, 1990) and the decision tree classification model of (Breiman, 1996). Another recent approach done by (Zarita *et al*, 2004) using backpropagation method in classification of river water quality. The approach provided a good result in classification, but, it takes a long times for training the network. In this paper, Radial Basis Probabilistic Neural Networks (RBPNN) which is known to have good generalization properties and is trained faster than the backpropagation network (Ganchev, 2003) is adopted for the construction of water quality level early warning classification prediction model.

Our proposed method, used kernel-function to increase the separability of data by working in a high dimensional space. Thus, the proposed method is characterized by higher classification accuracy than the original RBPNN. The kernel-based classification in the feature space not only preserves the inherent structure of groups in the input space, but also simplifies the associated structure of the data (Muller, 2001). Since Girolami first developed the kernel *k*-means clustering algorithm for unsupervised classification (Girolami, 2002), several studies have demonstrated the superiority of kernel classification algorithms over other approaches to classification (Zhang, 2002; Wu & Xie, 2003).

In this paper, we evaluate the performance of our proposed method, kernelized RBPNN, with a comparison to three well-known classification methods: the Back-Propagation (BP), Radial Basis Function Network (RBFN), and RBPNN; meanwhile, comparison and analysis on the classification power is done to these three methods. These methods performance are compared using river water quality data sets from Langat River of Selangor, Malaysia.

2.0 RADIAL BASIS PROBABILISTIC NEURAL NETWORKS

Radial Basis Probabilistic Neural Network (RBPNN) architecture is based on based on Bayesian decision theory and nonparametric technique to estimate Probability Density Function (PDF). The form of PDF is a Gaussian distribution. (Specht, 1990) had proposed this function (Yeh, 1998):

$$f_k(X) = \frac{1}{N_k} \frac{1}{(2\pi)^{m/2}} \frac{1}{\sigma^m} \sum_{j=1}^{N_k} \exp\left[\frac{\|X_k - X_j\|^2}{2\sigma^2} \right] \quad (1)$$

Since RBPNN is applicable to general classification problem, and assume that the eigenvector to be classified must belong to one of the known classifications, then the absolute probabilistic value of each classification is not important and only relative value needs to be considered, hence, in equation (1),

$$\frac{1}{(2\pi)^{m/2}} \frac{1}{\sigma^m}$$

Can be neglected and equation (1) can be simplified as

$$f_k(X) = \frac{1}{N_k} \sum_{j=1}^{N_k} \exp\left[\frac{\|X_k - X_j\|^2}{2\sigma^2} \right] \quad (2)$$

In equation (2), σ is the smoothing parameter of RBPNN. After network training is completed, the prediction accuracy can be enhanced through the adjustment of the smoothing parameter σ, that is, the larger the value, the smoother the approaching function. If the smoothing parameter σ is inappropriately selected, it will lead to an excessive or insufficient neural units in the network design, and over fitting or inappropriate fitting will be the result in the function approaching attempt; finally, the prediction power will be reduced.

Let $$d_{kj} = \|X_k - X_j\|^2$$

Be the square of the Euclidean distance of two points X_k and X_j in the sample space, and equation (2) can be re-written as

$$f_k(X) = \frac{1}{N_k} \sum_{j=1}^{N_k} \exp\left[\frac{d_{kj}}{2\sigma^2} \right] \quad (3)$$

In equation (3), when smoothing parameter σ approaches zero,

$$f_k(X) = \frac{1}{N_k}$$

If $X_k = X_j$, then

$$f_k(X) = 0$$

At this moment, RBPNN will depend fully on the non-classified sample which is closest to the classified sample to decide its classification. When smoothing parameter σ approaches infinity,

$$f_k(X) = 1$$

At this moment, RBPNN is close to blind classification. Usually, the researchers need to try different σ in certain range to obtain one that can reach the optimum accuracy. (Specht, 1992) had proposed a method that can be used to adjust smoothing parameter σ ,that is, assign each input neural unit a single σ; during the test stage, σ that have the optimum classification result are taken through the fine adjustment of each σ.

RBPNN is a three-layer feedforward neural network (as in Figure 1). The first layer is the input layer and the number of neural unit is the number of independent variable and receives the input data; the second hidden layer in the middle is Pattern Layer, which stores each training data; the data sent out by Pattern Layer will pass through the neural unit of the third layer Summation Layer to correspond to each possible category, in this layer, the calculation of equation (3) will be performed. The fourth layer is Competitive Layer; the competitive transfer function of this layer will pick up from the output of the last layer the maximum value from these probabilities and generate the output value. If the output value is 1, it means it is the category you want; but if the output value is 0, it means it is other unwanted category.

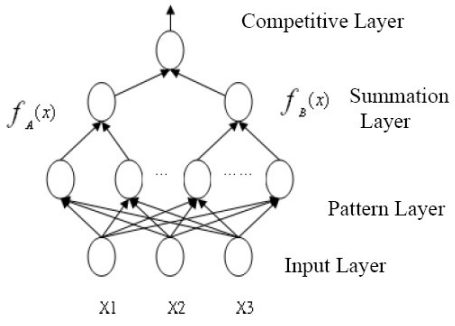

Figure 1: RBPNN architecture

3.0 KERNELIZED RADIAL BASIS PROBABILISTIC NEURAL NETWORKS

3.1 Kernel-Based Approach

Given an unlabeled data set $X = \{x_1, \ldots, x_n\}$ in the d-dimensional space R^d, let $\Phi : R^d \rightarrow H$ be a non-linear mapping function from this input space to a high dimensional feature space H. By applying the non-linear mapping function Φ, the dot product $x_i \bullet x_j$ in the input space. The key notion in kernel-based learning is that the mapping function Φ need not be explicitly specified. The dot product $\Phi(x_i) \bullet \Phi(x_j)$ in the high dimensional feature space can be calculated through the kernel function $K(x_i, x_j)$ in the input space R^d (Scholkopf & Smola, 2002).

$$K(x_i, x_j) = \Phi(x_i) \bullet \Phi(x_j) \tag{4}$$

Three commonly used kernel functions (Scholkopf & Smola, 2002) are the polynomial kernel function,

$$K(x_i, x_j) = (x_i \bullet x_j + c)^d \tag{5}$$

where $c \geq 0$, $d \in N$; the Gaussian kernel function,

$$K(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{6}$$

where $\sigma > 0$; and the sigmoidal kernel function,

$$K(x_i, x_j) = \tanh(\kappa(x_i, x_j) + \gamma) \tag{7}$$

where $\kappa > 0$ and $\gamma < 0$.

3.2 Formulation

Given a data point $x_i \in R^d (1 \leq i \leq n)$ and a non-linear mapping $\Phi : R^d \rightarrow H$, the Probability Density Function at data point x_i is defined as

$$f_i(x) = \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left(\frac{\|\Phi(x_i) - \Phi(x_j)\|^2}{2\sigma^2}\right) \tag{8}$$

where $\|\Phi(x_i) - \Phi(x_j)\|^2$ is the square of distance between $\Phi(x_i)$ and $\Phi(x_j)$. Thus a higher value of $f_i(x)$ indicates that x_i has more data points x_j near to it in the feature space. The distance in feature space is calculated through the kernel in the input space as follows:

$$\begin{aligned} \|\Phi(x_i) - \Phi(x_j)\|^2 &= (\Phi(x_i) - \Phi(x_j)) \bullet (\Phi(x_i) - \Phi(x_j)) \\ &= \Phi(x_i) \bullet \Phi(x_i) - 2\Phi(x_i) \bullet \Phi(x_j) + \Phi(x_j) \bullet \Phi(x_j) \\ &= K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j) \end{aligned} \tag{9}$$

Therefore, Equation (8) can be rewritten as

$$f_i(x) = \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left(\frac{K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j)}{2\sigma^2}\right) \tag{10}$$

The rest of the computation procedure is similar to that of the RBPNN method.

4.0 CONSTRUCTION OF CLASSIFICATION PREDICTION MODEL AND ROC CURVE ANALYSIS

A sequence of 323 rows of data consists of Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammonia-Nitrate level (AN), Suspended Substances level (SS), Dissolved Oxygen (DO), and Acidity (pH) that constitutes to the dataset. The data to included in the database is collected by Department of Environment of Malaysia and contains the records of Langat River water quality level which also reported in (Zarita *et al*, 2004). All of data sets of the six parameters have been Z-transformed for classification purposes and the Gaussian kernel was used in training of Kernelized RBPNN. The Z-transform is performed as a preprocessing step to minimize the gap of minimum and maximum of data values. Thus, we divided 323 rows of data into two separate parts, 200 rows for training and 123 rows remaining for testing. So, the ratio of train to test is approximately 2:1. All training is done in a desktop PC of Intel duo core 2GHz with 1Gb RAM memory.

In the Kernelized RBPNN, MATLAB is used to self-write the program for the model construction and we fixed value to the Smoothing Parameter σ to 0.1 according to (Sansone & Vento, 2001). In the test phase, each classifier has been tested with data sets mentioned above. Performance of different classifiers in the test phase can be observed in Table 1. The proposed classifier, which referred as Kernelized RBPNN, achieves the best result amongst all other methods. It can be concluded from Table 1 that the performance of Kernelized RBPNN is better than RBPNN, RBFN and BP. In term of time, Kernelized

RBPNN is about 89 times faster than RBFN and 2299 times faster than GD. The classification accuracy obtained using Kernelized RBPNN is 6% higher than RBPNN. It is also obvious that Kernelized RBPNN consumed less time in comparison to RBPNN, RBFN and BP. RBPNN outperform RBFN in classification accuracy due to its network architecture that implemented Bayesian decision and nonparametric technique which make it more applicable to general classification problem, this bring RBFN has the worst performance among all models. Although the classification accuracy of the well known classification method, BP, is high, but as we know, high processing time of BP makes it undesirable for many on-line recognition applications.

The true positive rate in Figure 2 means the percentage of the number of class with prediction result of 0 to the number of class with real value of 0; false positive rate means the percentage of the number of class with prediction result of 1 to the number of class with real value of 1. In this paper, the mutual verification of the data among the models is performed. The Kernelized RBPNN is verified as the best performed model and there are 123 rows of data in the test results. The result is drawn as Receiver Operator Characteristic (ROC) curve as in Figure 2. In the figure, the farther the ROC curve above the reference line, the larger the area under the ROC curve (AUC), which also means the higher the classification prediction power of this model (Bradley, 1997). Table 2 shows the mutual verification results of the river water quality. Figure 2 shows that the area under the ROC curve of the Kernelized RBPNN model of the river water quality data are larger than the other three models; from the observation of Table 2, it seems that the specificity of Kernelized RBPNN model is less than other three models, but the AUC value are higher than those of BP, RBPNN and RBFN models; therefore, it can be concluded that Kernelized RBPNN model has very good classification prediction capability.

Table 1: *Classification Prediction accuracy of Langat river water quality.*

Method	Classification Accuracy (%)	CPU Time (s)	Epoch
Kernelized RBPNN	88.62	0.047	9
RBPNN	82.93	0.062	11
RBFN	47.15	4.188	200
GD	84.55	108.062	30000

Table 2: *Mutual Verification with AUC.*

Method	AUC
Kernelized RBPNN	0.7195
RBPNN	0.6650
RBFN	0.4717
GD	0.6229

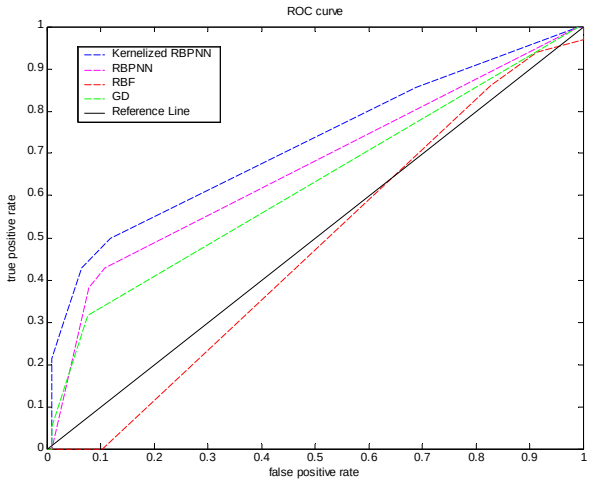

Figure 2: The Mutual verified ROC curve of test data of Langat river water quality

5.0 CONCLUSION

In this paper, a new RBPNN which consists of kernel-based approach has been proposed. Not like the conventional RBPNN that only based on Euclidean distance computation, the Kernelized RBPNN applied the kernel-induced distance computation which can implicitly map the input data to a high dimensional space in which data classification is easier. The network is applied to river water classification problem and showed an increased in classification accuracy in comparison with other well-known classifiers. The metrics for considering performance of a classifier in this work were the classification accuracy and processing time of the classifier. In future work, we plan to improve the classification rate by employing different kernel function such as wavelet function.

REFERENCES

Altman, E. I. (1968). Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609.

Bradley, A. P. (1997). The use of the area under ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30 (7), 1145-1159.

Breiman, L. (1996). Bagging predictors. *Machine Learning,* 24*(*2), 123-140.

Ganchev, T., Tasoulis, D. K. & Vrahatis, M. N. (2003). Locally Recurrent Probabilistic Neural Network for Text-Independent Speaker Verification. *In Proc. Of the EuroSpeech,* Vol 3, 762-766.

Ginoris, Y. P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007). Development of an image analysis procedure for identifying protozoa and metazoan typical of activated sludge system. *Water Research*, 41, 2581-2589.

Girolami, M. (2002). Mercer kernel-based clustering in feature space. *IEEE Trans. Neural Networks*, 13(3), 780-784.

Muller, K. R. (2001). An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12(2), 181-202.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* 18(1), 109-131.

Odom, M. D. & Sharda, R. (1990). A neural network model for bankruptcy prediction. *IEEE INNS IJCNN*, 2, 163-168.

Specht, D. F. (1990). Probabilistic neural network and the polynomial adaline as complementary techniques for classification. *IEEE Trans. Neural Networks*, 1 (1), 111-121.

Scholkopf, B. & Smola, A. J. (2002). *Learning with Kernels.* Cambridge: MIT Press.

Sansone, C. & Vento, M. (2001). A classification reliability drive reject rule for multi-expert system. *International Journal of Pattern Recognition and Artificial Intelligent,* 15(6), 1-19.

Wu, Z. D. & Xie, W. X. (2003). Fuzzy *c*-means clustering algorithm based on kernel method. *The Fifth International Conference on Computational Intelligent and Multimedia Applications,* 1-6.

Yeh, Y. C. (1998). Application of neural network (Scholars Book, Taiwan Taipei, 1998). *Scholars Books*. Taiwan.

Zarita, Z., Sharifuddin, M. Z., Lim, E. A. & Hafizan, J. (2004). Application of neural network in river water quality classification. *Ecological and Environmental Modeling (ECOMOD 2004)*, 2 (1), 24-31.

Zhang, R. & Rudnicky, A. I. (2002). A large scale clustering scheme for kernel *k*-means. *The 16th International Conference on Pattern Recognition,* 289-292.