EVOLUTIONARY BASED SMOOTHING PARAMETER OPTIMIZATION OF PROBABILISTIC NEURAL NETWORK AND ITS USAGE AS A CLASSIFIER IN ODOR RECOGNITION SYSTEM

Benyamin Kusumoputro, Herry
Faculty of Computer Science, University of Indonesia

ABSTRACT

Probabilistic Neural Network has received considerable attention nowadays and obtained many successful application. This type of neural system has shown marvelous higher recognition capability compare with that of Back-Propagation neural system. However, this neural has shown some drawbacks, especially on determining the value of its smoothing parameter and its neural structure optimization when large number of data is necessary. Supervised-structure determination of PNN is an algorithm to solve these problems by selecting a set of valuable neurons using Orthogonal Algorithm and determining the optimal smoothing parameter value using Genetic Algorithm. In this paper an experimental set up for comparison of the Supervised-structure determination of PNN with that of the Standard PNN as a neural classifier on the Odor Recognition System is conducted. Experimental results show that the Supervised-structure determination of PNN performed higher recognition rate compare with that of Standard PNN, even using lower number of neurons.

Keywords: Neural Network, Probabilistic Neural Network, Genetic Algorithm, Odor Recognition System

1. INTRODUCTION

Recently, neural networks is widely used as a pattern classifier in the fields of pattern recognition system. As one of the promising neural system that shown superior recognition capability compare with that of backpropagation neural system [1][2], Probabilistic Neural Networks (PNN) [3][4] has relied its capability on its neural structure determination as a consequence with its learning algorithm. Probabilistic Neural Network (PNN) is a type of neural networks that developed based on Bayesian Classification theory that has widely used in pattern classification problems. However, PNN has its drawback, especially on its network structure determination, i.e. determining network size and choosing the value of smoothing parameter.

As in PNN training algorithms, every new training data will be represented by a new neuron, which is added on its pattern layer [3][4]. As the consequence, the neural size of PNN will increase according to the increment of the used training data, which in return will affect on its calculation cost. Another problem that should encountered in the PNN is on determining the exact value of the smoothing parameter, where the optimal value depends directly on the characteristics of the known training data.

These two problems mentioned above are well realized by researchers and some algorithms are proposed to reduce the training sample, such as vector quantization method for grouping the training sample to find the cluster centers as reduced data [5][6][7]. As the reduction of the neural structure is done based on clustering approach, these reduction algorithms are called unsupervised structure determination methods. Mao et al [8] have then developed a reduction structure of the PNN by directly used in its neuron selection in which at the same time also looking for the best smoothing parameter through the use of Genetic Algorithms [9]. This method is then called supervised-structure determination [8].

In this paper, we will review the performance of the PNN algorithm on our developed Odor Recognition System as a neural classifier for determining two-mixture odors, which is not properly discriminated when using backpropagation neural system.

2. PROBABILISTIC NEURAL NETWORK

Probabilistic Neural Network (PNN) was introduced by Donald Specht at 1990 [3]. As shown in Fig. 1, the architecture of PNN is consists of 4 layers, i.e. an input layer, a pattern layer, a summation layer, and a decision layer. The main function of input layer is just for distributing a new data to each neuron in pattern layer. Pattern layer consists of neurons where each neuron represents one training vector data. As its consequence, the more training data is added to the neural system, the more number of neurons should be created accordingly.

The activation function for neuron x_{ij} (j^{th} neuron of i^{th} class) for an incoming data x can be calculated through

$$\phi_{ij}(x) = \frac{1}{(2\pi)^{d/2}\sigma^{d}} \exp\left[-\frac{(x - x_{ij})^{T}(x - x_{ij})}{2\sigma^{2}}\right] \quad (1)$$

where i denotes class index, j the neuron index, d the vector's dimension, σ denotes the determined smoothing

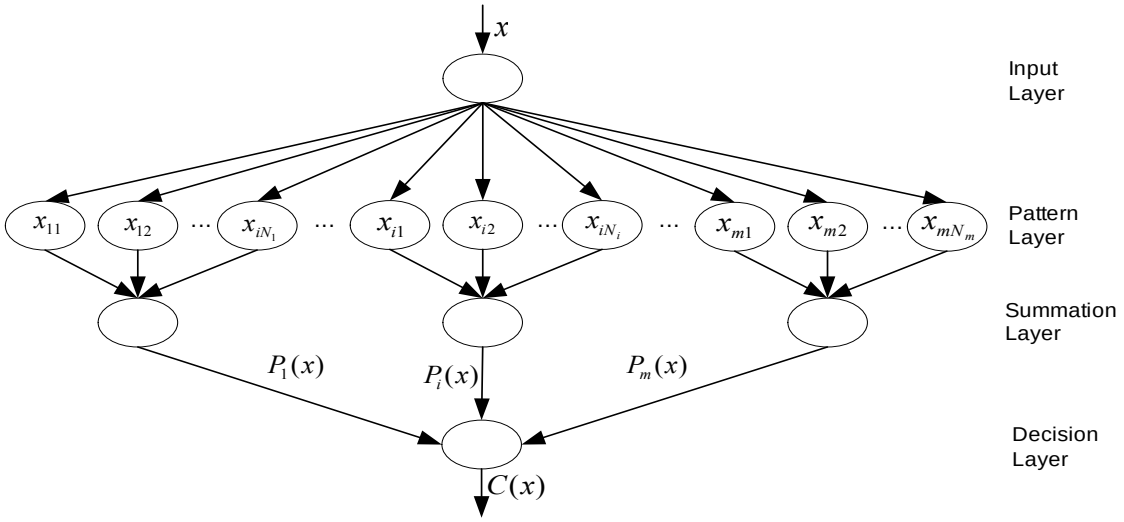

Figure 1. Architecture of PNN

parameter, x the testing vector, and x_{ij} the training vector for j^{th} neuron of i^{th} class.

In the summation layer, the maximum probability of the unlearn-testing data to be classified as a member of a determined class is computed by summarizing and averaging the calculated values from Eq. 1 for all neurons in the same class, through:

$$P_i(x) = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi_{ij}(x) \qquad (2)$$

where P_i is maximum probability of testing vector x as member of i^{th} class and N_i is number of neuron at pattern layer which member of i^{th} class.

In the next step, one neuron in decision layer will decide the selected class of testing vector by choosing the class that has highest likelihood compare with that of the other classes. The activation function of this neuron is,

$$\hat{C}(x) = \text{argmax}\{ P_i(x) \} \quad i = 1, 2, 3, \ldots, m \qquad (3)$$

where $\hat{C}(x)$ is selected class of testing vector x and m is number of class.

3. SUPERVISED PNN STRUCTURE DETERMINATION

Supervised-structure determination of PNN is firstly introduced by Mao et al [8], to discriminate their method with that already developed from other researchers [5][6][7]. The main goal of this algorithm is to construct an optimal PNN structure to achieve higher recognition rate even using minimum size of network topology. This algorithm can be divided in two parts, with the first part is searching the optimal smoothing parameter value using Genetic Algorithm, while in the second part which is looking for the most valuable neurons, is done through selection process based on Orthogonal Algorithm.

3.1. SMOOTHING PARAMETER VALUE SELECTION USING GENETIC ALGORITHM

In PNN, the determination of the smoothing parameter is a critical aspect, due to its direct connection with the recognition accuracy of the neural system. An appropriate smoothing parameter is often data dependent; therefore the selection of the smoothing parameter selection is an essential step to be taken in every usage of the PNN. Suppose we already have the network size of the PNN, with n is the number of selected neuron in pattern layer, then the problem can be defined as constrained optimization problem of *min{n}* subject to:

$$\mu < \delta \qquad (4)$$

where μ is error rate of classification, and δ is a given maximum limit of error rate tolerance of classification. Since quantitative relation between network size, error rate of classification, and smoothing parameter (σ) is not exist, then genetic algorithm (GA) is used to solve above optimization problem.

GA is a searching algorithm, which developed based on nature selection of genetics, introduced by John Holland [9]. The usage of GA in this step is done by applying GA's processes, i.e. encoding, fitness value calculation, reproduction, crossover, and mutation. The last three operators are commonly used and theoretically

are fully explored somewhere, and will not explained here, however, the first two parameters will be explained below.

1) Encoding the Smoothing Parameter

Genetic Algorithms is searching algorithm that directly applied in the smoothing parameter (σ) searching space, and since normalized data is used in this encoding process, the value of σ is smaller then one. In this paper, 4 decimal bits encoding is being used, therefore only 4 decimal bits behind comma will be encoded. For example, if an individual $\sigma = 0.3279$, then this value is represented as string of decimals as $b_1{:}3,\ b_2{:}2\ b_3{:}7\ b_4{:}9$

2) Fitness Value Calculation

In the used algorithm, every individual represents a smoothing parameter value. Using every possible value of σ, a set of network structure candidate is then determined through procedure that will be explained in the next section. Since the objective function of the system to be minimized is the size of the neural networks, the fitness function can be calculated with the following scheme:

$$\rho_i = \rho_{max} - \frac{\rho_{max} - \rho_{min}}{\eta_{max} - \eta_{min}}(\eta_i - \eta_{min}) \qquad (5)$$

where,

ρ_i = fitness value of i^{th} individual

η_{min} = minimum network size of current candidates population

η_{max} = maximum network size of current candidates population

ρ_{min} = maximum fitness value

ρ_{max} = minimum fitness value

In all of the experiments conducted here, values of ρ_{min} and ρ_{max} are chosen to be *0.5* and *1*, respectively.

3.2. NEURON SELECTION USING ORTHOGONAL ALGORITHM

The main goal of this part is to select the most representative neurons in pattern layer. If vector x_{ik} is training vector of k^{th} neuron of class C_i, then the maximum probability of vector x_{ik} if classified as member of C_i is,

$$p_i(x_{ik}) = \frac{1}{(2\pi)^{d/2}\sigma^d}\frac{1}{N_i}\sum_{j=1}^{N_i}\exp\left[-\frac{(x_{ik}-x_{ij})^T(x_{ik}-x_{ij})}{2\sigma^2}\right]$$

$$= \sum_{j=1}^{N_i}\phi_{ij}(x_{ik}) \qquad (6)$$

where

$$\phi_{ij}(x_{ik}) = \frac{1}{(2\pi)^{d/2}\sigma^d}\frac{1}{N_i}\exp\left[-\frac{(x_{ik}-x_{ij})^T(x_{ik}-x_{ij})}{2\sigma^2}\right] \qquad (7)$$

with $p_i(x_{ik})$ is a nonlinear function of σ and vector x_{ik}. But if value of σ had been determined before and output of each neuron $\phi_{ij}(x_{ij})$ is treated as a variable, then $p_i(x_{ik})$ will become a linier combination of $\phi_{ij}(x_{ij})$ as shown in Eq. 6. These variables will then be used to evaluate the degree of importance of every neuron in pattern layer.

Equation (6) can also be written in matrix form as,

$$P = \Phi\,\theta \qquad (8)$$

where

$$\theta = [1,1,...,1]^T$$

$$P = [p_i(x_{i1}), p_i(x_{i2}),..., p_i(x_{iN_i})]^T$$

$$\Phi = \begin{bmatrix} \phi_{i1}(x_{i1}) & \phi_{i2}(x_{i1}) & \dots & \phi_{iN_i}(x_{i1}) \\ \phi_{i1}(x_{i2}) & \phi_{i2}(x_{i2}) & \dots & \phi_{iN_i}(x_{i2}) \\ \dots & \dots & \dots & \dots \\ \phi_{i1}(x_{iN_i}) & \phi_{i2}(x_{iN_i}) & \dots & \phi_{iN_i}(x_{iN_i}) \end{bmatrix}$$

Transforming matrix Φ using orthogonal transformation will give,

$$\Phi = QR = [Q_1, Q_2, \dots, Q_{N_i}]R \qquad (9)$$

where $Q_1, Q_2, \dots, Q_{Ni}$ are an orthogonal basis and R is an upper triangular matrix. Degree of importance (Γ_j) of j^{th} candidate neuron that member of class C_i is calculated based on norm of vector Q_j [1], i.e.

$$\Gamma_j = Q_j^T Q_j \qquad (10)$$

In condition where all neurons have same smoothing parameter value, as higher value of Γ_j as more important j^{th} neuron.

The summarize of procedure to calculate neuron's degree of importance as follows,

1) Choose the most representative neuron from N_i neuron of class C_i by searching neuron that has highest degree of importance. Then use selected neuron to calculate Q_1.

   $Q_1^{(\alpha)} = \phi_\alpha, \qquad \alpha = 1, 2, \dots, N_i$

   $\phi_\alpha = [\phi_\alpha(1), \phi_\alpha(2),...., \phi_\alpha(N_i)]^T$

   Degree of importance is calculated as,

   $\Gamma_1^{(\alpha)} = [Q_1^{(\alpha)}]^T Q_1^{(\alpha)}, \qquad \alpha = 1, 2, \dots, N_i$

2) Choose j^{th} most representative neuron from remaining neurons (total remaining neurons = $N_i - j + 1$). Neuron that has highest degree of importance is selected as j^{th} most representative neuron.

   $Q_j^{(\alpha)} = \phi_{k_\alpha} - \sum_{l=1}^{j-1} r_{l\alpha}^{(\alpha)} Q_1, \quad \alpha = 1, 2, \dots, N_i - j + 1$

   $r_{l\alpha}^{(\alpha)} = Q_l^T \phi_{k_\alpha} / Q_l^T Q_1, \quad \alpha = 1, 2, \dots, N_i - j + 1, \ 1 < i$

By combining two methodologies as written in Section 3.1 and Section 3.2 respectively, the algorithm of supervised PNN structure determination can be summarized as follows:

1. Generate initial population of *P* that consist of n individual where each individual represents a value of σ. Set number of generation $i = 1$.
2. By using neuron selection procedure with orthogonal algorithm and a set value of σ, a set of candidate network structure can be obtained. Calculate fitness value for each candidate. Build mating pool *M* using all population in *P*, where fitness value of each candidate is used as probability.
3. Select randomly *2* individuals as a couple parents from *M*. Do crossover procedure that will create *2* offspring and put them in *O*. Crossover procedure is repeated until number of offspring in *O* is in the same value with that in *P*.
4. Do mutation procedure for offspring in *O*. Build the network structure for each individual in *O* and calculate fitness value for each structure.
5. Select n highest fitness value of individual from *P* and *O*.
6. Empty *P*, and set *P* with selected individual based on step 5. Set number of generation $I = I + 1$, then empty *O*.
7. Repeat step 2-6 until number of generated reach specified value.

4. EXPERIMENTS ON ODOR RECOGNITION SYSTEM

The experiments are designed to elaborate the capability of the Supervised PNN system and compare its results with the Standard PNN in our developed odor recognition system. The main task of this odor recognition system is to recognize and determine the-unlearn of two mixture odors, which can not be properly recognized using Backpropagation neural system.

There are 3 types of two-mixed odors that used in all of the experiments, i.e. citrus (*Ci*), canangga (*Can*) and rose (*Ro*), which is diluted in 6 grades of alcohol (Alch) concentrations. As can be seen in the Table 1, every type of two-mixed odors is consists of 6 classes of odors, thus constructs 18 classes of odors. Label of CiAlch25% means that a citrus odor with alcohol with 25% concentration, which is prepared by mixing a *50%* of odor and *50%* of alcohol with various gradient concentrations from *0%* to *70%*..

Using three types of odors as depicted in Table 1, the experiments are divided into 5 datasets of classes, to observe the relationship between the numbers of the used neuron with its recognition capability. First dataset consists of 6 classes of two-mixed odors, while second dataset consists of 8 classes of two-mixed odors, and so on.

Table 1. List of Types of Odor

Types of Odor	Label	Description
Type 1	CiAlch0%, CiAlch15%, CiAlch25%, CiAlch35%, CiAlch45%, CiAlch70%	Composition 1:1 (10ml : 10ml)
Type 2	CanAlch0%, CanAlch15%, CanAlch25%, CanAlch35%, CanAlch45%, CanAlch70%	
Type 3	RoAlch0%, RoAlch15%, RoAlch25%, RoAlch35%, RoAlch45%, RoAlch70%	

As it is the drawback of the conventional PNN system for determining its smoothing parameter, two selected values of smoothing parameter, i.e. *0.1* and *0.003* are determined as example. In the Supervised-PNN, however, the smoothing parameter and its numbers of neuron are determined through evolutionary processes through genetic algorithms, until the neural system could give high recognition rate.

Parameters that are used in Supervised PNN are error rate tolerance of training data recognition of 0%, the number of population for each generation is 30, the size of mating pool is 30, the number of generations is 30, and the mutation probability is 0.1 respectively.

Table 2. Average Recognition Rate of Training Data

Number of Class	Recognition Rate of Training Data		
	Standard PNN (σ = 0,01)	Standard PNN (σ = 0,003)	Supervised PNN
6	79.43%	99.92%	100.00%
8	81.74%	99.88%	100.00%
9	90.93%	100.00%	100.00%
12	79.56%	99.91%	100.00%
18	78.25%	99.92%	100.00%
	81.98%	**99.93%**	**100.00%**

Table 3. Average Recognition Rate of Testing Data

Number of Class	Recognition Rate of Testing Data		
	Standard PNN (σ = 0,01)	Standard PNN (σ = 0,003)	Supervised PNN
6	75.24%	98.06%	98.31%
8	78.61%	97.55%	97.67%
9	89.61%	99.53%	99.51%
12	77.12%	98.13%	98.02%
18	76.40%	98.07%	98.18%
	79.39%	**98.27%**	**98.34%**

Table 2 shows the recognition rate of training data As shown in this table, the recognition rate using Standard PNN with $\sigma = 0.01$ is lower that that of the Standard PNN

with σ = 0.003. However, the recognition rate of the Supervised PNN has shown superior recognition capability, which can gave maximum tolerance of error rate, i.e. 0%. This table also shows that smoothing parameter value has a major influence on the recognition rate of the neural system.

Table 3 shows the recognition rate of the overall PNN systems when using the testing data. As depicted in this table, the recognition rate of the Standard PNN with σ = 0.01 shows a lower result as can be expected from the result on Table 1. When using the Standard PNN with σ = 0.003, however, the recognition rate is higher and comparable with that of Supervised PNN system. Even the value of σ for Standard PNN is set to be the same for overall of every dataset, in the Supervised PN, however, the value of σ is rarely to be the same value for every dataset. The value of σ is solely depends on the determination result using GA, and usually differ from dataset to other dataset.

As can be seen from Table 2 and Table 3, respectively, it can be concluded that the number of classes is not affecting the recognition rate of the overall neural systems. This is also the superiority of the PNN system compare with Back propagation neural system as reported in [9], where the Backpropagation neural system shows lower recognition rate when more number of odor are used.

Table 5 shows the comparison of number of neurons that are used in the Standard PNN and in the Supervised PNN. Standard PNN with σ = 0.01 and σ = 0.003 are used the same number of neurons, which is equal with number of training data to be inputted to the system. As a contrary, the Supervised PNN always used a lower number of the used neurons, which differs for each datasets. Since the computation cost has a relationship with the number of the used neurons at pattern layer, it can be concluded that the Supervised PNN always performs lower computation cost compare with that of Standard PNN.

Figure 3 shows the percentage of the used neuron in the Supervised PNN system compare with that of Standard PNN system. It shows that even with only about 20% of using number of neurons as in the Standard PNN system, the Supervised PNN can achieve higher recognition capability.

Table 4. Average Number of Neuron at Pattern Layer

Number of Class	Rate Number of Neuron at Pattern Layer		
	Standard PNN (σ = 0,01)	Standard PNN (σ = 0,003)	Supervised PNN
6	583.33	583.33	108.31
8	766.67	766.67	156.47
9	875.00	875.00	97.75
12	1162.50	1162.50	252.75
18	1750.00	1750.00	404.08

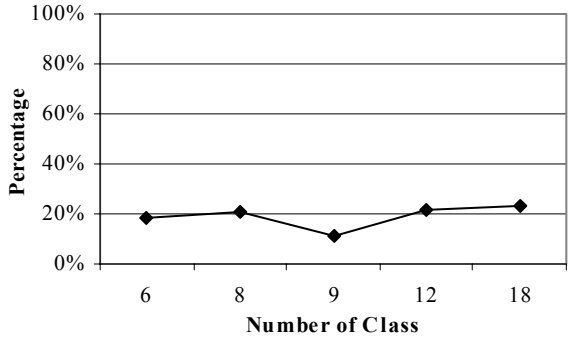

Figure 3. Percentage of Number of Neuron of Supervised PNN Compare with Standard PNN

5. CONCLUSIONS

Probabilistic Neural Network has widely used in pattern classification problems with high recognition capability compare with that of backpropagation neural system, however, PNN has its drawback, especially on its network structure determination, i.e. determining network size and choosing the value of smoothing parameter.

As the smoothing parameter has a big influence to the recognition rate of PNN system, selection of the optimal value is an important issue to have higher recognition rate. An evolutionary algorithm such as Genetic Algorithms is then used to determined this smoothing parameter, and with reducing the number of neuron at pattern layer is also be done through orthogonal algorithm. These methodologies are then applied to determine a mixture-odor that could not be properly solved by backpropagation neural system. It is shown that the Supervised PNN has higher recognition rate compare with that of Standard PNN even with lower number of hidden neurons. Reduction of the hidden neurons also reduces the computation cost of PNN, however, it is not affecting the recognition rate of the system.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Ministry of Research and Technology of Indonesia for their support funding of this research through RUT X under contract number 14.06/SK/RUT/2003

REFERENCE

[1]. B.Kusumoputro and M.Rivai, *Discrimination of Fragrance Odor by Arrayed Quartz Resonator and a Neural Network,* Int. Conf. on Comput. Intelli. and Multimed. Application H.Selvaraj and B. Verma (Eds), World Scientific, 264-269, 1998

[2] B. Kusumoputro, Comparison of FALVQ and modified backpropagation in artiifcial odor discrimination system, IASTED Proc. Int. Conf. on Modelling, Identification and Control, February 15-19, Innsbruck, Austria, 1999, 434 – 437.

[3] D. F. Specht, Probabilistic neural networks and the polynomial adaline as complementary techniques for classification, *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 111- 121, 1990

[4] Masters, Timothy, *Advanced Algorithms For Neural Networks: A C++ Sourcebook*, John Wiley & Sons, 1995

[5] Raghu, P. P., and B. Yegnanarayana, "Supervised Texture Classification Using a Probabilistic Neural Network and Constraint Satisfaction Model", *IEEE Transaction on Neural Networks*, vol. 9, no. 3, May 1998.

[6]. P. Burrascano, Learning vector quantization for the probabilistic neural networks', IEEE Trans. On Neural Networks, vol 2, 1991, pp 458 -461, July

[7] A. Zaknich," A vector quantization reduction method for the propabilistic neural networks," in Proc. IEEE Int. Conf. Neural Networks, Piscataway, NJ, 1997, pp 1117 – 1120

[8] Mao, K.Z., K.-C. Tan, and W. Ser, "Probabilistic Neural-Network Structure Determination for Pattern Classification", *IEEE Transaction on Neural Networks*, vol. 11, no. 4, July 2000.

[9]. Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989.