Classification of Coffee
using Artificial Neural Network

Devil H.F. Yip and William W.H. Yu
Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong,
Pokfulam Road, Hong Kong.
Email : hfdyip@hkusua.hku.hk, wwhyu@hkucc.hku.hk

Abstract

The paper presents a method for classifying coffees according to their scents using artificial neural network (ANN). The proposed method of uses genetic algorithm (GA) to determine the optimal parameters and topology of ANN. It uses adaptive back-propagation to accelerate the training process so that the entire optimization process can be achieved in an accelerated time. The optimized ANN has successfully classified the coffees using as relatively small set of training data. The performance of the optimized ANN compare significantly better than the ones and the methods proposed by other researchers.

Keyword: Artificial Neural Network, Genetic Algorithms and Adaptive Back-Propagation.

1. Introduction

Coffee, alcohol and perfume are the products which are often classified not in terms of physical parameters such as colour and dimension, but odour. Traditionally, the human olfactory system is employed for classification of these products. The disadvantage is obvious. The results are obtained highly dependent on the subjective experience of the human classifier. It is difficult to establish a quantified standard with which to measure the quality of these products in the respect of their scents.

Tin oxide gas sensors have been used for quantifying the scents of this kind of products [1]. The data obtained from an array of these sensors have been used for the classification of alcohol by means of statistical techniques as well as using artificial neural network (ANN) [1,2].

From Gardner et al. [3], the discriminant and classification function analyses are performed on the array response to each of three commercial coffees. Multivariate function are calculated from the entire data set (90 samples) or alternatively half of it, to permit cross-validation. A success rate of 89.9% is achieved with the former procedure in classifying the three commercial coffee odours. This values falls to 81.1% when half of the data set is used for cross validation. The results still need to be improved.

When the method proposed by Gardner et al. [2] was employed for the classification of coffees, the results were less encouraging. Thirty sets of data were available for each of the 3 different types of coffee for the exercise. Taking 29 and 27 sets of data as the training data with 1 and 3 sets as the testing data respectively, both exercises resulted in one misclassification in the identification process. Taking 24 sets for training and 6 sets for testing, 9 misclassifications occurred.

It is possibly because the volume of data available for training is small and the quality of coffee as a product is subject to many factors, including species/varieties of beans, climatic and soil conditions during growth, storage conditions after both harvesting and toasting, and the parameters during the toasting process. Therefore, quite a variation of quality could be expected from different batches of final products.

The paper proposes a topologically optimized ANN for the classification of coffee and genetic algorithm (GA) is used as a tool for obtaining the optimal network. The proposed GA optimized networks have successfully classified the three different types of coffee using as little as 80% of the available data for training.

2. Experimental details

2.1 Measurement methods

The data of coffee [4] were generated by an array of 12 commercial tin oxide gas sensors. The details of the experimental procedure are described in [1].

2.2 Network architecture and training algorithm

Though Gardner et al.'s method [2] failed to classify the 3 different types of coffee correctly with one-hidden layer standard back-propagation network, it has been suggested by Lippman [5] that a three-layer network (two-hidden layer) has *sufficient computational degrees of freedom to solve any* problem for the multi-layer perceptron (MLP) and the ANN approach remains an obvious choice for the classification of coffees. It possesses several advantages over the conventional methods in terms of adaptability (learning, self-organization, generalization and training), noise tolerance and fault tolerance.

One of the critical problems with using ANN is in designing the network topology appropriate for a given problem. Design considerations include network architecture (number of hidden units, hidden layers, and interconnects), and the network parameters (learning rate, momentum, weight decay, initial weights, etc.,). Typically the topology and other parameters are determined through trials and experiments which could often be long and tedious. The temptation is to accept any solutions which may not be optimal. In some cases, the problem may appear insolvable after a long process of trials and experiments during which the right combination of parameters has not been reached.

Recently, GAs researchers have investigated GA as a tool for determining the optimal topology and parameters of ANN for a specific problem [6-8]. GAs has the advantage of not getting stuck in a local optima. When the search space is large and there are no *a priori* knowledge about the parameter space, the optimization process may take a long time to converge.

For classification of coffees, the ANN should have 12 nodes in input layer and 3 nodes in output layer representing the 12 sensor inputs and the confidence values for the 3 types of coffee respectively. Back-propagation network has been adopted because it is capable of approximating arbitrary mapping given a set of examples and adaptive back-propagation (ABP) technique [9] used to achieve fast convergence during training. Since the authors are using the ABP for training, the ANNs are restricted to a feed-forward MLP. The MLP may be randomly or fully connected. In a fully connected MLP every node in a layer is connected only to all the nodes in the next layer.

GAs has been used to determine the most effective number of hidden layers, number of hidden nodes per layer and interconnects. The parameters were coded into a 82 bits genetic string or chromosome and shown in Figure 1.

weight range (10 bits)
seed for random weight generation (10 bits)
offset range (10 bits)
seed for random offset generation (10 bits)
initial learning rate (10 bits)
initial momentum coefficient (10 bits)
number of hidden layers (2 bits)
number of nodes in each of hidden layers (10 bits)
connection of nodes to each other (10 bits)

Figure 1. The parameters of the chromosome of the network

For each jth genetic string or chromosome, the training of the network produces a fitness value given by:

$$f_j = \frac{1}{\sum_{i=1}^{8} W_i F_i(X_{ij})}$$

Where W_i is significance weighting factor for the ith parameter,
F_i is a function for the ith parameter and,
X_{ij} is the ith parameter value.

The parameters of the fitness values included are: the overall error, the maximum number of nodes in the network, the number of links in the network, the number of epochs required to train the network, maximum number of epoch specified, the number of misclassified pattern during training/recall, the number of node out of the tolerance limit during training/recall, and the total number of training/recall patterns. Each gene represents a specific network topology and the fitness value indicates its performance.

Crossover and/or mutation follows each training process and the genes compete for survival based on their fitness values. The optimization process is subject to some specified conditions. The network has a maximum of 2

hidden layers each with a maximum of 20 hidden nodes; 20 generations with population size 50; crossover probability 0.9 and mutation probability 0.01. Figure 2 illustrates the optimization process.
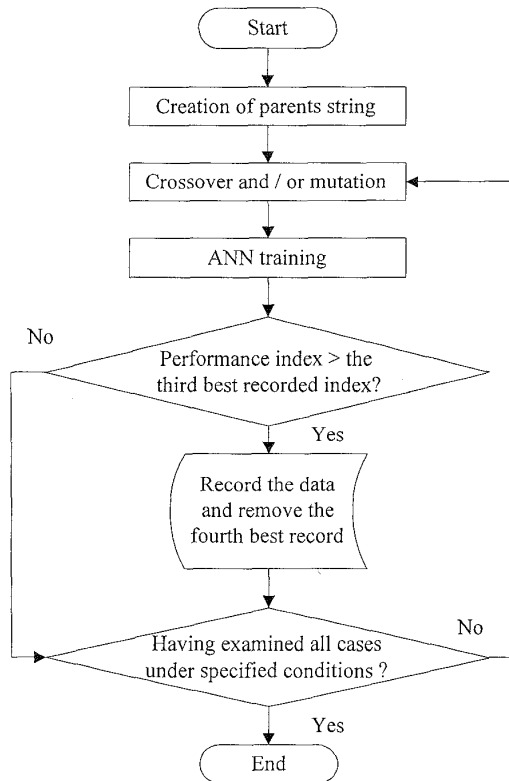

Figure 2. The optimization process

2.3 Classification experiments

In principle, in larger the training set is, the more robust the network. Due to the limited data available, 30 sets for each class, the maximum data for training is limited to 29 sets, leaving one set for testing the performance of the network. To test how effective the proposed method is for classifying the 3 types of coffees, experiments also have been conducted using less training data of 27 (90%), 24 (80%) and 21 sets (70%) with 3, 6 and 9 sets of data for testing the network respectively.

3. Experimental results

The ideal confidence value for the correctly recognized type of coffee is 1, otherwise 0. In practice, to be correctly classified into a selected type of coffees, the product under test should have a confidence value as close as possible

to 1, otherwise 0. A threshold must be chosen to determine if the product belongs to a specific type. A threshold of 0.6 is chosen for the product being positively identified as a specific type and 0.4 for certain rejection.

Results obtained by the proposed method from the four experiments using 29, 27, 24 and 21 sets of data for training are summarized (number of misclassification in set of 30) and compared with the results obtained from Gardner et al.'s method in Table 1.

Training data from 30 sets	29 sets (96.7 %)	27 sets (90 %)	24 sets (80 %)	21 sets (70 %)
Gardner et al.	1	1	9	10
Proposed method	0	0	0	1

Table 1. Comparison of classification results

The proposed method has successfully classified the coffee using a training set as small as 80% of the available data. The above results compare favourably with the work carried out using the discriminant and classification function analyses for the same data in [3] and the ANN proposed by Gardener et al. for the coffee data in [2], and indicates an improvement in reliability over the conventional techniques. Table 2 summaries the topologies parameters of the optimal networks for the experiments.

Proposed method	29 sets (96.7 %)	27 sets (90 %)	24 sets (80 %)	21 sets (70 %)
First hidden layer	3 nodes	3 nodes	13 nodes	13 nodes
Second hidden layer	7 nodes	7 nodes	4 nodes	4 nodes
Interconnects	108	108	171	171

Table 2. The summaries the topologies parameters of the optimal networks

4. Discussion and Conclusion

This paper has shown that GA has been successfully applied to obtain the optimal ANNs with marked improvement in the ability to correctly classify the 3 types of coffee using relatively small training sets of data. However, the experimental results also show that the optimal network topologies determined by applying GAs are not structurally simpler comparing with the network proposed by Gardner et al. [2]. Thus the improved ability in correctly classifying the coffees comes with the cost of increase in the topology complexity.

Using ABP has significantly improved the convergence speed. Typically the process converges around 150 iterations comparing to 15,000 to 20,000 reported by Gardner et al. [2]. The optimal network topologies were found at less than 20 generations for each set of training data. The improved performance of the ANN presented in this paper suggests that the commercial coffee products can be classified in an automated process using gas sensor array in conjunction with an ANN. The system can be developed to monitor the quality resulting in high quality products. With further studies and experiments, quantified standards for the scents of different products can be established.

Acknowledgment

References

[1]   H.V. Shurmer, J.W. Gardner and P. Corcoran, *'Intelligent Vapour Discrimination using a Composite 12-Element Sensor Array'*, Sensors and Actuators, Vol. B1, pp. 256-260, 1990.

[2]   J.W. Gardner, E.L. Hines and M. Wilkinson, *'Application of Artificial Neural Networks to an Electronic Olfactory System'*, Meas. Sci. Technol., Vol. 1, pp. 446-451, 1990.

[3]   J.W. Gardner, H.V. Shurmer and T.T. Tan, *'Application of an Electronic Nose to the discrimination of Coffees'*, Sensors and Actuators, Vol. B6, pp. 71-75, 1992.

[4]   J.W. Gardner and E.L. Hines, 'Integrated Sensor Array Processing in an Electronic Nose', IEE Colloquium number 1988/130, London, UK, December, 1988, pp. 7/1-7/4.

[5]   R.P. Lippman, 'An Introduction to Computing with Neural Nets', IEEE ASSP Magazine, Vol. 4, No. 2, pp. 4-22, 1987.

[6]   J.D. Schaffer, D. Whitley and L.J. Eshelman, *'Combinations of Genetic Algorithms and Neural Networks: A Survey of the State of the Art'*, International Workshop on Combinations of Genetic Algorithms and Neural Networks, pp. 1-37, 1992.

[7]   N.J. Radcliffe, *'Genetic Set Recombination and Its Application to Neural Network Topology Optimization'*, Neural Computing and Applications, Vol. 1, pp. 67-90, 1993.

[8]   Y. Ikuno, H. Kawabata, Y. Shirao, M. Hirata, T. Nagahara and Y. Inagaki, *'Application of an Improved Genetic Algorithm to the Learning of Neural Networks'*, IEICE Transactions Fundamentals, Vol. E77-A, No. 4, pp. 731-735, 1994.

[9]   F.M. Silva and L.B. Almeida, *'Acceleration Techniques for the Backpropagation Algorithm'*, Lecture Notes in Computer Science: Part 412, Springer-Verlag, 1990.