Implementation of Elbow Method to improve the Gases Classification Performance based on the RBFN-NSG Algorithm

Jin-Young Jeon, Jang-Sik Choi, and Hyung-Gi Byun

Abstract

Currently, the radial basis function network (RBFN) and various other neural networks are employed to classify gases using chemical sensors arrays, and their performance is steadily improving. In particular, the identification performance of the RBFN algorithm is being improved by optimizing parameters such as the center, width, and weight, and improved algorithms such as the radial basis function network–stochastic gradient (RBFN-SG) and radial basis function network–normalized stochastic gradient (RBFN-NSG) have been announced. In this study, we optimized the number of centers, which is one of the parameters of the RBFN-NSG algorithm, and observed the change in the identification performance. For the experiment, repeated measurement data of 8 samples were used, and the elbow method was applied to determine the optimal number of centers for each sample of input data. The experiment was carried out in two cases (the only one center per sample and the optimal number of centers obtained by elbow method), and the experimental results were compared using the mean square error (MSE). From the results of the experiments, we observed that the case having an optimal number of centers, obtained using the elbow method, showed a better identification performance than that without any optimization.

Keywords: RBFN-NSG, Elbow method, Chemical sensor array, K-means clustering, Pattern recognition

1. INTRODUCTION

At present, chemical sensor arrays, also known as electronic nose systems, are being widely used for the identification of certain chemicals, and the ability to classify specific pattern features from minimal information has led to an increasing interest in techniques related to the recognition of gases/odors. A radial basis function network (RBFN) and various other neural network algorithms are employed in the classification of odors using an electronic nose, and their performance is continually being improved [1,2]. In particular, the identification performance of the RBFN algorithm is being improved by optimizing its parameters such as the center, width, and weight. Some of the improved algorithms that have been announced are the radial basis function network–stochastic gradient (RBFN-SG) [3,4] and the radial basis function network–normalized stochastic gradient (RBFN-NSG) [5]. In this study, we optimized the number of centers, which is one of the parameters of the RBFN-NSG algorithm, and observed the change in identification performance. Repeated measurement data of 8 samples were used for the experiment, and the elbow method was applied to determine the optimal number of centers for each sample of input data [6,7]. The experiment was carried out for two cases – A (having only one center per sample) and B (having the optimal number of centers obtained by elbow method), and the experimental results were compared using mean square error (MSE).

2. EXPERIMENTAL

2.1 Experimental data

The data used in the experiments were provided by the University of Manchester (Prof. K.C. Persaud), UK. A sensor array consisting of 32 conducting polymer sensors (L = 32) was used to measure the gases (1 and 10% acetonitrile (acn1, acn10), 1% acetone (ac), 1% butanone (bu), 10% methanol (me), 1 and 10% propanol (pr1, pr10), and water(wa)). We divided the total 528 data samples (66 patterns for each gas) in a ratio of 8:2 for the measurement, and used 8 samples (423) for RBFN training and 2 (105) for the test.
2.2 Elbow method

The elbow method is one of the most popular methods used to determine the optimal number of clusters. In this study, we performed the elbow method using k-means, which is a type of clustering partitioning technique.

Clustering methods such as k-means minimize the total intra-cluster variation (known as total within-cluster variation or total within-cluster sum of square). The within-cluster sum of square (WSS) represents the compactness of the clustering, and it should be as small as possible.

The steps of the elbow method using k-means are given below [7].

1. Compute the clustering algorithm (k-means clustering) for different values of k. For instance, by varying k from 1 to 10 clusters
2. For each k, calculate the total WSS
3. Plot the curve of WSS against the number of clusters k.
4. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

Fig. 1 shows the curve of WSS against the number of clusters k (k is the number of centers). From this figure, the optimal number of centers can be determined as 3.

2.3 Radial Basis Function Network

Fig. 2 shows the basic RBFN. It consists of the input layer, the hidden layer, and the output layer. The number of nodes in the hidden layer is determined by the user, and this number is the same as the number of centers (or the number of clusters). There are weights that are connected to each other, between the hidden layer and the output layer, and the Gaussian function is applied to the hidden layer as its basis function.

2.3.1 Radial Basis Function Network - Stochastic Gradient

RBFN-SG is one of the variants of the basic RBFN that has an improved accuracy. This technique increases the classification accuracy by training the parameters of basic RBFN such as the center value (c), the width (σ) of the Gaussian function, and the weight (w) between the hidden layer and output layers [4].

2.3.2 Radial Basis Function Network – Normalized Stochastic Gradient

RBFN-NSG is a technique that normalizes RBFN-SG by applying Taylor's expansion. This method is less sensitive to various inputs than RBFN-SG, and possesses a better training speed and identification performance [5].

The RBFN-NSG technique showed the best performance during our experimental trials for this study.

3. RESULTS AND DISCUSSIONS

In this study, we applied the elbow method to each sample to obtain an optimal number of centers for each sample, and the experimental results were compared using MSE. The experiment was carried out for two cases: the first having only one center per sample and the second having the optimal number of centers obtained by the elbow method.
3.1 Determining the optimal number of centers

Eight kinds of samples were used for the experiment, and the elbow method was used eight times to determine the optimal number of centers for each sample. The elbow method was applied using the k-means function from the R stats library, and the WSS was calculated by increasing k from 1 to 20.

Figs. 3 and 4 show the WSS curves against the number of centers for each sample.

Fig. 3 shows the curves for the acn1, acn10, ac, and bu samples, and four, three, four, and three centers, respectively, seem to be the best choices.

Fig. 4 shows the curves for the me, pr1, pr10, and wa samples, and four, four, three and three centers, respectively, seem to be the best choices. Table 1 shows the number of centers selected for each sample. Twenty-eight centers were selected as the optimal number of centers.

| No. | Sample | The number of Centers |
|-----|--------|-----------------------|
| 1   | acn1   | 4                     |
| 2   | acn10  | 3                     |
| 3   | ac     | 4                     |
| 4   | bu     | 3                     |
| 5   | me     | 4                     |
| 6   | pr1    | 4                     |
| 7   | pr10   | 3                     |
| 8   | wa     | 3                     |

3.2 RBFN-NSG Training and Test

For the training and test of RBFN-NSG, 528 (423 training data and 105 recognition data) data were used.

Figs. 5, 6, and 7 depict the graphs of the training and test results.

In Fig. 5, (a) shows the result before optimization, and (b) shows the result after optimization. When graphs (a) and (b) are compared, it can be seen that the result of (b) is denser than that of (a). This can also be seen in Table 2.

From Table 2, it is seen that before optimization, the MSE was 0.013332, however, after optimization, it decreased to 0.004265.

Figs. 6 and 7 show the result of the recognition test based on the training result shown in Fig. 5. Fig. 6 shows the result before optimization.

Fig. 5. Results of 8 samples training: (a) before optimization, (b) after optimization.

Fig. 6. Results of 8 samples recognition before optimization.
optimization, while Fig. 7 shows the result after optimization. When the graphs in Fig. 6 and Fig. 7 are compared, it can be seen that the overall result shown in Fig. 7 is closer to the target value than that in Fig. 6. Likewise, this can be seen in Table 2. From Table 2, it is seen that before optimization, the MSE was 0.05703, however, after optimization, it was reduced to 0.019356.

### 4. CONCLUSIONS

In this study, to improve the identification performance of the RBFN-NSG algorithm, we performed an experiment to optimize the number of centers, which is one of the parameters of this algorithm. To optimize the number of centers for the sample data, we clustered and computed the WSS using the k-means algorithm. The calculated WSS was plotted as a graph against the number of centers, and the optimal number of centers for each sample was selected from the bend in the graph. The optimum number of centers for each sample was either 3 or 4, and 28 centers were selected.

Next, we carried out the training and recognition experiment by using RBFN-NSG for two cases: the first one having only one center per sample and the second one having the optimal number of centers. The results are shown both graphically and by using MSE. It was found that the MSE for the recognition before optimization was approximately 0.0193. The method of increasing the number of centers is a very simple technique for increasing the identification performance of the RBFN algorithm; however, the amount of computation required increases proportionately to the number of centers. Therefore, unconditionally increasing the number of centers would be inefficient.

However, if the elbow method used in this experiment is applied, it is possible to obtain a comparatively high identification performance with a minimal increase in computation.

### ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.R0126-16-1050, Olfactory Bio Data based Emotion Enhancement Interactive Content Technology Development)

### REFERENCES

[1] F.F. Soulie, P. Gallinari, “Industrial applications of neural network”, World Sci. pp. 85-90, 1998.
[2] H. Byun, K. Persaud, J. Kim, D. Lee, “Improvement of RBF Network’s performance for application to odour sensing systems”, Proc. of ISOEN’99, Tubingen, Germany, pp. 237-240, 1999.
[3] H. G Byun, “Post-processing Technique for Improving the Odor-identification Performance based on E-Nose System”, J. Sensor Sci. & Tech., Vol. 24, No. 6, pp. 368-372, 2015.
[4] I. Cha, S.A. Kassam, “Interference cancellation using radial basis function networks”, Signal Process., Vol. 47, pp. 247-268, 1995.
[5] N. Y. Kim, H. G. Byun, K. C. Persaud, “Normalization Approach to the Stochastic Gradient Radial Basis Function Network Algorithm for Odor Sensing Systems”, Sensors and Actuators B, Vol. 124, pp. 407-412, 2007.
[6] D. J. Ketchen, Jr. and C. L. Shook, “The application of cluster analysis in Strategic Management Research: An analysis and critique”, Strategic Management Journal, Vol. 17, No. 6, pp. 441-458, 1996.
[7] http://www.sthda.com/english/wiki/determining-the-optimal-number-of-clusters-3-must-known-methods-unsupervised-machine-learning/three-popular-methods-for-determining-the-optimal-number-of-clusters (retrieved on Nov. 28, 2016).