Artificial Neural Network for Quantitative Posture Classification in Thai Sign Language Translation System

Kumphol Wasanapongpan, and Nopporn Chotikakamthorn

Faculty of Information Technology &
Research Center for Communications and Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand
(E-mail: s3067019@kmitl.ac.th)

Abstract: In this paper, a problem of Thai sign language recognition using a neural network is considered. The paper addresses the problem in classifying certain signs conveying quantitative meaning, e.g., large or small. By treating those signs corresponding to different quantities as derived from different classes, the recognition error rate of the standard multi-layer Perceptron increases if the precision in recognizing different quantities is increased. This is due the fact that, to increase the quantitative recognition precision of those signs, the number of (increasingly similar) classes must also be increased. This leads to an increase in false classification. The problem is due to misinterpreting the amount of quantity the quantitative signs convey. In this paper, instead of treating those signs conveying quantitative attribute of the same quantity type (such as 'size' or 'amount') as derived from different classes, here they are considered instances of the same class. Those signs of the same quantity type are then further divided into different subclasses according to the level of quantity each sign is associated with. By using this two-level classification, false classification among main gesture classes is made independent to the level of precision needed in recognizing different quantitative levels. Moreover, precision of quantitative level classification can be made higher during the recognition phase, as compared to that used in the training phase. A standard multi-layer Perceptron with a back propagation learning algorithm was adapted in the study to implement this two-level classification of quantitative gesture signs. Experimental results obtained using an electronic glove measurement of hand postures are included.

Keywords: Sign Language, Gesture Recognition, Artificial Neural Network, Quantitative Posture

1. INTRODUCTION

Gesture classification is a form of pattern recognition problems. Like other pattern recognition problems, neural networks have also been applied to solve gesture recognition problems. For example, in [1] a neural net was applied for isolated American Sign Language (ASL) recognition. Their system obtained gesture data from an electronic (data) glove and achieved the accuracy level of 86%. In [2], a gesture-recognition system called Glove–Talk was developed, using a neural network for linking information from a data glove and a speech synthesizer. In [3], a neural network was applied to recognition a small number of hand postures representing English alphabets. In [4], a problem of segmenting continuous Thai signs is considered. The Bayesian estimator is applied in the paper for sign recognition.

One of the most popular neural network architecture is a multi-layer Perceptron. However, the network has a weakness in terms of the difficulty to strictly confine the hidden-layer multi-dimensional space regions of similar patterns, as constructed by the network during the training stage, to be close to one another. As a result, similar patterns may be recognized as totally different patterns. This is likely to occur for patterns that are not included in the training set (or those that are not well represented by the training patterns).

In Thai sign language, there are some hand postures that convey quantitative information. Examples of such signs include those that represent 'size' and 'amount' quantity type. For each quantity type, a sign can describe different levels of quantitative attribute by slightly changing hand posture in a well-defined manner. Figure 1 shows an example of such quantitative hand posture signs.

To apply a standard multi-layer Perceptron to recognize signs in which those quantitative postures are included, hand postures corresponding to the same quantity type may be divided into some number of classes. Each class contains the hand postures that convey the same level of quantitative attribute. For example, there may be three quantitative classes for the hand postures representing different quantitative levels of 'size' (as may be labeled by 'small', 'medium', 'large'). This approach does not need any modification made to the network structure, and standard back propagation algorithm can be directly applied. However, the approach has two weaknesses. First, the number of classes to be recognized is increased. The higher precision in recognizing different quantitative levels is needed, the higher number of classes is required. This can result in an increase in the complexity of the network. With the fact that the number of such quantitative signs as found in Thai sign language is relatively low, this complexity problem may not be a serious concern.

The second problem with the approach as described is due to similarity among quantitative and non-quantitative hand postures. Firstly, same-quantitative-type hand postures conveying similar quantitative levels are difficult to differentiate due to the similarity of their hand shapes. Secondly, insufficient degree of freedom in a network model, as well as insufficient number of the hand postures used in the training set, can result in some quantitative hand postures being misclassified as non-quantitative postures and vice versa. The latter is due to the difficulty in controlling the Perceptron to strictly confine the hidden-layer multi-dimensional space regions of similar patterns to be close to one another. In any of the two cases, the recognition rate is decreased.

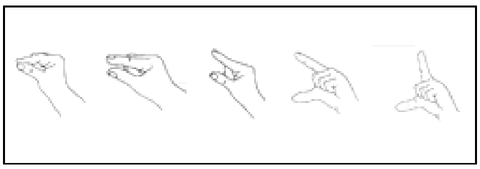

Fig. 1. Example of quantitative hand postures

This paper proposes to use a two-level classification for quantitative hand posture recognition to solve the problems as described. The first level is used to classify those quantitative postures according to the quantity types they represent. The second level is used to identify the posture's quantitative level. The paper is organized as follows. First, a theory and notation of multi-layer Perceptron neural network as applied to Thai sign language recognition are introduced in Section 2. Next, the proposed two-level quantitative posture classification is described in Section 3. Experiment results, obtained from a real measurement using an electronic glove, are provided with some discussion in Section 4. A conclusion remark is given in Section 5.

2. MULTI-LAYER PERCEPTRON FOR SIGN LANGUAGE RECOGNITION

The MLP is a supervised neural network. A general structure of a multi-layer Perceptron neural network is shown in Figure 2.

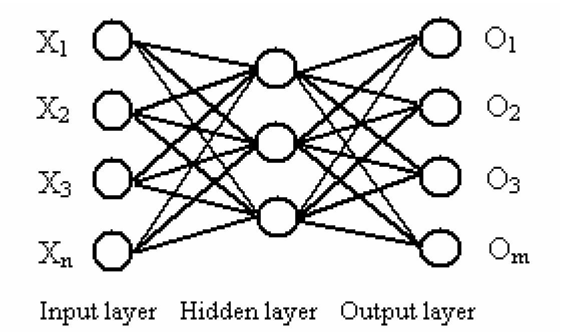

Fig. 2. Standard multilayer Perceptron neural network

The MLP use the back-propagation (BP) algorithm, [5-7] for training. The algorithm consists of 2 phases : a feed-forward process and a back-propagation process. For the initial stage, the weights of the network are randomly selected. The learning rate η is pre-set before the learning phase. Given $\mathbf{x}_i$ is the vector of n dimension. It is fed to the network input layer. The i^{th} input layer is connected to the j^{th} node of the hidden layer through the weights $\mathbf{W}_{ji}$. i = 1, 2, …, n, and j = 1, 2, …, s. The j^{th} hidden layer is connected to m^{th} node of the output layer through the weights $\mathbf{Z}_{kj}$. k = 1, 2, …, m.

Where
- n = number of input layer node
- s = number of hidden layer node
- m = number of output layer node

The output of the hidden layer is obtained by passing the summation of the multiplication between the inputs and the weights ($\mathbf{W}$), through the activation function, as described by

$$\mathbf{h}_j = f(\mathbf{net}_j) \tag{1}$$

$$\mathbf{net}_j = \sum_{i=1}^{n} \mathbf{W}_{ji}\mathbf{x}_i \tag{2}$$

where $f(\circ)$ is the activate function or transfer function . Here , $f(\circ)$ is chosen as a sigmoidal function given below.

$$f(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{x})} \tag{3}$$

Next, the output from the hidden layer is fed as an input to the output layer. The output of the output layer is obtained by passing the summation of the multiplication between the hidden output and the weights ($\mathbf{Z}$), through the activation, as described by

$$\mathbf{o}_k = f(\mathbf{net}_k) \tag{4}$$

$$\mathbf{net}_k = \sum_{j=1}^{s} \mathbf{Z}_{kj}\mathbf{h}_j \tag{5}$$

The output vector ($\mathbf{o}_k$), $\mathbf{o}_k = [o_1, o_2, \ldots o_m]$ generated from the feed-forward process is compared with the desire response vector ($\mathbf{t}_k$), $\mathbf{t}_k = [t_1, t_2, \ldots t_m]$. The cost function or error (e) used here is a square error function, which is given by

$$e = \frac{1}{2}\sum_{k=1}^{m}(\mathbf{t}_k - \mathbf{o}_k)^2 \tag{6}$$

Back-propagation process is then using error to adjust weights as follows:

$$\mathbf{Z}_{(t+1)} = \eta\boldsymbol{\delta}_k\mathbf{h}_j + \mathbf{Z}_{(t)} \tag{7}$$

where

$$\boldsymbol{\delta}_k = \mathbf{o}_k(1-\mathbf{o}_k)(\mathbf{t}_k - \mathbf{o}_k) \tag{8}$$

$$\mathbf{W}_{(t+1)} = \eta\boldsymbol{\beta}_j\mathbf{x}_i + \mathbf{W}_{(t)} \tag{9}$$

where

$$\boldsymbol{\beta}_j = \mathbf{h}_j(1-\mathbf{h}_j)\sum_k \boldsymbol{\delta}_k\mathbf{W}_{kj} \tag{10}$$

The network as described is here applied to the sign language recognition problem. Only the static and isolated postures are considered here. The problems of static/dynamic segmentation, as well as dynamic gesture classification were discussed elsewhere (see [4], for example). Under this scope, X 's are data as measured from an electronic glove sensors. In this study, a total of 22 sensors were placed on the glove. Thus, the number of input signals n = 22. The outputs O correspond to the hand posture classes.

3. TWO-LEVEL METHOD FOR QUANTITATIVE POSTURE CLASSIFICATION

In this section, the proposed two-level method for quantitative hand posture classification is described.

3.1 Network structure

A structure of the multi-layer Perceptron neural network, modified to accommodate the two-level classification scheme is shown in Figure 3. From this point, the network is referred as QMLP (Quantitative-classification Multi-Layer Perceptron)

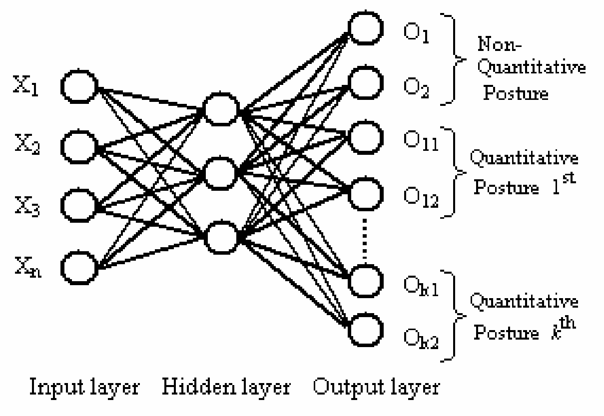

Fig. 3. A multi-layer Perceptron for quantitative posture classification

From Fig. 3, the network outputs can be divided into two categories. The first one, in which there are two output nodes O_1 and O_2 in the figure, corresponds to groups of the non-quantitative posture outputs. The second category is a group of quantitative posture outputs. From the figure, it is noted that for each quantitative posture type, there are two output nodes associated with it. For example, the output nodes $O_{1,1}$ and $O_{1,2}$ belongs to the 1^{st} quantitative posture type (or class). This two-node per class arrangement makes possible the separation between quantitative posture type classification and the quantitative level classification. This point will be clarified later in this section.

3.2 Defining desired responses for supervised training

To train the network in Fig. 3, a desired response for each training input must be defined. Desired responses for the non-quantitative and quantitative postures are defined differently as described below.

- For the input data corresponding to an i^{th} non-quantitative posture class, desired response of the output of the i^{th} non-quantitative posture is set to 0.9. Desired responses of other outputs node are set to 0.1.
- For the input data corresponding to an i^{th} quantitative posture class, two desired responses corresponding to the i^{th} quantitative posture are defined as a point on the circle as described in Fig. 4.

From Fig. 4, d = 0.9 is the circle radius, and θ is the angular value representing the quantitative level. For example, in this paper, the largest quantitative level for any quantity type is associated with $\theta = 85^o$, while the smallest level is associated with $\theta = 5^o$. The levels in between the two extremes are proportionally given angular values between 5 and 85. To simplify labeling the input data with an appropriate quantitative level, however, the whole range of θ may be divided into N angular intervals. Each of those intervals corresponds to a particular quantitative level. Thus, the value of N controls the precision of the recognizer in differentiating postures of similar quantitative levels.

With θ being given according to the quantitative level of the training input data, the corresponding desired responses of the associated output pairs ($O_{i,1}$ and $O_{i,2}$ in the figure) are given by

$$O_{i,1} = d \cdot \cos(\theta) \tag{11}$$

$$O_{i,2} = d \cdot \sin(\theta) \tag{12}$$

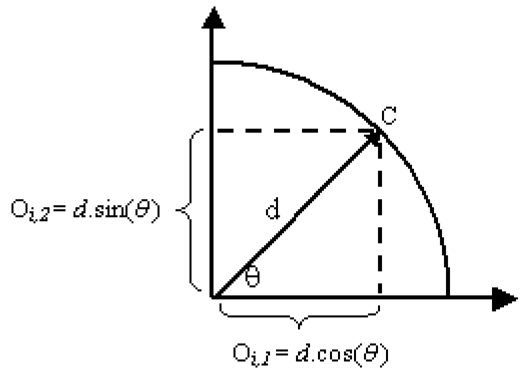

Fig. 4. Mapping between the quantitative level and the desired response pair for quantitative posture data

3.3 Quantity type and quantitative level classification

After training the network using the method for defining the desired responses, as just detailed, the network can be used for hand posture recognition. However, because each quantitative posture class has two output nodes, values of these outputs must be first transformed so that they can be compared against the output nodes of non-quantitative postures. To do so, for arbitrary input data, let's $o_{i,1}$ and $o_{i,2}$ be the output values corresponding to $O_{i,1}$ and $O_{i,2}$ respectively. From these output values, compute the following transformed output values

$$\hat{d}_i = \sqrt{(o_{i,1})^2 + (o_{i,2})^2} \tag{13}$$

$$\hat{\theta}_i = \tan^{-1}(o_{i,2} / o_{i,1}) \tag{14}$$

From these two transformed output values, $\hat{d}_i$ is used to compared against other output values, to obtain the output that yields the highest value. The input data is then recognized as belonging to the class corresponding to the output node with the highest value. In case that the i^{th} quantitative class is the winning class, the value of $\hat{\theta}_i$ is used to obtain the associated quantitative level. By comparing $\hat{\theta}_i$ against the lower and upper bounds (5 and 85 degrees, respectively), a comparative quantitative level of the input data can be identified. Alternatively, if the whole range of θ is divided into N angular intervals during the training stage, the input data is classified as belonging to one of those N quantitative levels.

4. EXPERIMENT

In the experiment, hand posture data was collected using an electronic glove containing 22 sensors. Five signers participated in the experiment. Each of them was asked to perform a total of 25 different static non-quantitative postures, and 3 static quantitative postures. For each quantitative posture, each signer was asked to perform signs corresponding to 5 different quantitative levels. Therefore, each signer performs 40 signs in total. Each sign was repeatedly performed for 460 times.

A standard multi-layer Perceptron method

When a standard multi-layer Perceptron was used, during the training and testing stages, all 40 signs were first classified into Q different classes, where Q was varied among 4 experiments (2, 3, 5 and 9). However, for the case where Q = 9, an original data set containing 5 different quantitative levels was first interpolated to obtain a new data set of 9 quantitative levels. Thus, combined with 25 non-quantitative signs, the total number of posture classes is 52.

In training the network, the desired response of 0.9 was given to the output node corresponding to the correct posture class, while the response of 0.1 was given to the other nodes. Number of hidden nodes in the hidden layer is 40, 60 and 80 nodes.

The QMLP method

In the experiment using QMLP method, for the purpose of network training, an original data set containing 5 different quantitative levels was first interpolated to obtain a new data set of 17 quantitative levels. Each posture data was then classified into Q different classes, where Q was varied among 4 experiments (2, 3, 5 and 9).

In this case, the total number of the output nodes was 31 because each quantitative class requires two output nodes. The number of the hidden nodes was the same as used in the standard multi-layer Perceptron method.

Table 1. Recognition rates of the two methods

Neural Network Type	Accuracy (%)
Standard Method	100
QMLP Method	100

Table 1 compares the recognition rates of the both methods as described. The figures as shown were calculated from the ability of a recognizer in providing correct posture classes for non-quantitative postures, and correct classes of quantity type for quantitative postures. Misinterpreting the quantitative levels is not counted as a recognition error. From the table, both methods give the same result.

Next, accuracy in classifying the quantitative level for quantitative postures is analyzed. The quantitative level classification errors of the two methods are shown in Table 2 - 5. The error rate was computed as a percentage of quantitative level misclassification, weighted by the absolute difference between the correct quantitative level and the one estimated by any of the two recognizers.

Table 2 . Results using quantitative data classified into 2 classes during the training phase.

No. Hidden Node / No. of Recognition Precision Level	Accuracy (%)		
	40	60	80
2	100	100	100

a) Standard MLP Method

No. Hidden Node / No. of Recognition Precision Level	Accuracy (%)		
	40	60	80
2	99.52	99.73	99.60
3	93.64	93.15	93.37
5	91.06	91.20	91.41
9	90.98	91.44	91.36

b) QMLP Method

Table 3. Results using quantitative data classified into three classes during the training phase.

No. Hidden Node / No. of Recognition Precision Level	Accuracy (%)		
	40	60	80
3	100	100	100

a) Standard MLP Method

No. Hidden Node / No. of Recognition Precision Level	Accuracy (%)		
	40	60	80
3	98.46	98.76	98.72
5	97.20	96.96	96.76
9	96.84	96.98	96.92

b) QMLP Method

Table 4. Results using quantitative data classified into five classes during the training phase.

No. Hidden Node / No. of Recognition Precision Level	Accuracy (%)		
	40	60	80
5	99.20	99.70	99.59

a) Standard MLP Method

Table 4. (Con't)

No. of Recognition Precision Level \ No. Hidden Node	Accuracy (%)		
	40	**60**	**80**
5	98.30	98.11	98.18
9	96.33	96.56	96.51

b) QMLP Method

Table 5. Results using quantitative data classified into nine classes during the training phase.

No. of Recognition Precision Level \ No. Hidden Node	Accuracy (%)		
	40	**60**	**80**
9	98.84	99.00	98.96

a) Standard MLP Method

No. of Recognition Precision Level \ No. Hidden Node	Accuracy (%)		
	40	**60**	**80**
9	98.05	98.32	98.09

b) QMLP Method

From the above tables, it was found that the standard MLP performed slightly better than the QMLP, when the recognition precision level matches the quantitative levels used during the training phase. Unlike the standard method, however, the QMLP method can be used to classify quantitative postures using a higher precision level than that used in the training stage. For example, in Table 3, three times increase in the precision level during the recognition stage, as compare with the training stage, result in only slightly decrease in accuracy levels. Nonetheless, in all experiments, the accuracy was found to be reduced when the recognition precision level was increased.

5. CONCLUSION

In this paper, the quantitative-classification multi-layer Perceptron network has been proposed. The architecture has been motivated by the need to deal with quantitative postures (those that convey quantitative level information), since the standard multi-layer Perceptron inefficient when applied to those posture classes. The paper proposes the use of the two-level classification scheme, applied to the standard multi-layer Perceptron. Details on assigning the desired response as well as the classification decision for the proposed method have been given. Experimental results included show that the propose method offers higher quantitative recognition precision, as compared with the number of quantitative levels used in the training phase. This can greatly simply the network training, and yet provides high precision in quantitative level classification.

REFERENCES

[1] Manjula B. Waldron, "Isolated ASL Sign Recognition System for Deaf Persons", IEEE Transactions on Rehabilitation Engineering, Vol. 3, No. 3, September 1995, pp. 261-271.

[2] S. Sidney Fels and Geoffrey E. Hinton, "Glove-Talk : A Neural Network Interface Between a Data - Glove and a Speech Synthesizer", IEEE Transactions on Neural Network Vol. 4, No. 1, January 1993.

[3] Philip A. Harling, "Gesture Input using Neural Network", BSc degree, Department of Computer Science University of York Heslington, UK, March 1993.

[4] Worawit Werapan and Nopporn Chotikakamthorn, "Gesture Segmentation Technique for Thai Sign Language Translation", The Second International Symposium on Communications and Information Technology, Chonburi, Thailand, 23 – 25 October 2002, pp. 483-486.

[5] D.E. Rumelhart et al., "Parallel Distributed Processing : Explorations in the Micro Structure of Cognition Vol. I: Foundations", MIT Press, Cambridge, MA, 1986.

[6] Rumelhart, D.E. Hinton, G.E. and Williams, R.J. "Learning internal representations by error propagation", In Rumelhart, D.E. and McClelland, J.L. Eds, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, pp.318-362, 1986

[7] Parker, K.L. and Thornbrugh, A.L. "Parallelized back – propagation training and its effectiveness", International Joint Conference on Neural Network, Vol. 2, pp.179-182. Washington, DC. 1990.