4D−1

Expression Recognition Using Back-propagation Networks

バックプロパゲーションにおける表情認識について

J. Ding M. Shimamura* H. Kobayashi T. Nakamura

Tohoku University *Furugawa Midorigaoka Hospital

1. Introduction

As a nonverbal language, facial expressions have been studied by psychologists for a long time. It is more universal than verbal languages, in view of reflecting more sensitive information from the emotion of human beings at the same time. In this paper, we introduce a method using back-propagation neural networks to facial expression recognition. As an important technique for pattern recognition, back-propagation networks have many successful applications. In our scheme, the number of hidden nodes are determined by means of logically analyzing internal representationsof features in a face. Moreover, we show the simulation results of our methods and discuss the ability of theneural networks in recognizing facial expressions.

2. Background Material

2.1. Facial Expression Research

According to Ekman[1], human facial expressions have so far been found to be associated with six emotional states, which include angry, fear, happiness, sadness, surprise and disgust. A face can be divided into three regions-brows (including forehead), eyes and mouth. Besides, these three regions can present different expressions simultaneously and signify emotional blends (e.g., fear-anger) or other subtle emotional differences. In order to develop a comprehensive system that could distinguish all possible visually distinguishable facial movements, Ekman and Friesen[2] developed the Facial Action Coding System. All facial movements can be described using single action units(AUs) or their combinations.

2.2. Back-propagation Neural Networks

The back-propagation network is a widely used option among the current types of available neural network systems [3]. In essence, it uses a gradient descent method to minimize a cost function equal to the mean square difference between the desired and the actual outputs.Desired output of all nodes is typically 0 or close to 0 unless that node corresponds to the class the current input is from in which case it is 1.0 or close to 1.0. The network is trained by initially selecting small random weights and internal thresholds and then presenting all training data repeatedly. Weights are adjusted after every trial using information specifying the correct class until weights converge and the cost function is reduced to an acceptable value. The weights in networks are often initializedby small random values and there are no special rules for determining the number of hidden nodes. However, determining this number efficiently is the first problem in applications.

3. Expression Recognition Using Back-propagation

Corresponding to the three regions in a face, three networks should be set up. They are a brows network, a eyes network and a mouth network. In each network, output nodes correspond to the classes of expressions. The inputs of the three networks are movements of characteristic points which can be obtained from 2-D image processing. As shown in Figure 1, we define the characteristic points in the three regions which are enough to distinguish necessary and sufficient AUs for expression recognition.

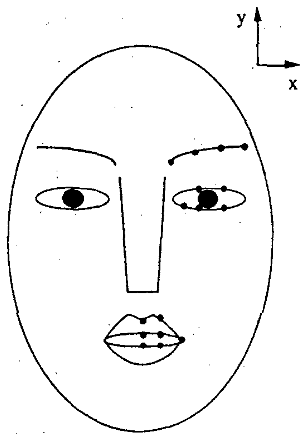

Figure 1. Charactistic Points of a Face

So, there are 8 inputs for the brows network (4 in X direction and 4 in Y direction); 5 inputs for the eyes network (all in Y direction) and 10 inputs for the mouth network(3 in X direction anc 7 in Y direction).

The relationship between expressions and the relation-

Expression Recognition Using Back-propagation Networks
J. Ding, M.Shimamura, H.Kobayashi, T.Nakamura
Tohoku University, Furukawa Midorigaoka Hospital

ship between AUs and movements of characteristic points can be used to set up our networks. We define hidden nodes corresponding to necessary AUs or combination of AUs so that the number of the hidden nodes is determined by the number of necessary AUs. Therefore, the eyes network with 4 hidden nodes,the brow network with 4 hidden nodes and the mouth network with 7 hidden nodes are determined.

According to the above corresponding relationship, we initialize the weights logically in the networks. So, before the training data is presented, hyper-space has been divided into hyper-cubes in order. Notice that it always takes a regular back-propagation a long time to obtain the same situation. However, the final accurate decision boundary lines, which depend on weights and thresholds, are defined by the training data using the regular back-propagation algorithm.

4. Simulation Results

According to the rules of facial expressions we set up a simulation model to generate random movement values of the characteristic points as training data for networks. Using our simulator, we examined the networks by classification rate, error calculation and the ability of noise tolerance. The error is defined as

$$\text{Error} = \sum_{i} [\ Od(i) - Oc(i)\ ]$$

where Od(i) is the desired value and Oc(i) is the calculated value of the ith output node.

Compared with the case in which whole weights are initialized by random values, the classification rates are increased very quickly and the error sums of the six expression patterns are reduced stably in case of initializing weights logically(See Figure 2,3). However, the ability of the noise tolerance is not so good(See Figure 4). Initializing weighs can accurate convergence, but can not guarantee the strongest ability of the noise tolerance. The neural networks can find more excellent weights set automatically but it means a longer learning time.

5. Conclusion

In this paper, we have analyzed facial expressions and proposed the method to define hidden nodes and set initial weights for the back-propagation networks of expression recognition. In the future, we will try to build the practical recognition system and evaluate our networks again. In addition, applying bi-directional networks to this research is also a very interesting topic.

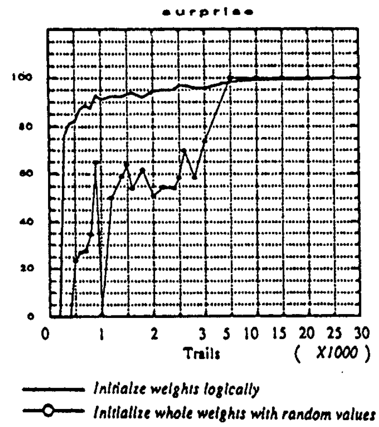

Figure 2. Compariton of the Classification Rates of the NNs Using Different Mothods to Initialize Weights (%)

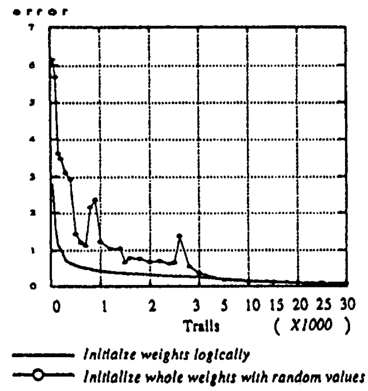

Figure 3 Compariton of the Error Sums of the NNs Using Different Motheds to Initialize Weights (%)

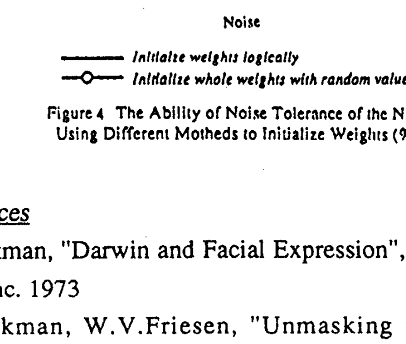

Figure 4 The Ability of Noise Tolerance of the NNs Using Different Motheds to Initialize Weights (%)

References

[1] P.Ekman, "Darwin and Facial Expression", Academic Press, Inc. 1973

[2] P.Ekman, W.V.Friesen, "Unmasking the Face", Prentice-Hall,Inc. 1975

[3]D.E.Rumelhart, "Parallel Distributed Processing",Vol.1, The MIT Press, 1986