A Practical Neural Network for Handwritten Character Recognition Based on Dynamics-Based Active Learning and Self-Organization of Feedback

Member Naotake Natori (RWCP\textsuperscript{1} Neural Systems Toshiba Laboratory\textsuperscript{2})
Member Kazuo Nishimura (RWCP Neural Systems Toshiba Laboratory)

This paper proposes a novel neural network which realizes practical recognition ability for handwritten characters by modeling human active learning and recognition. The neural network incorporates a human handwriting model for active learning and a self-organizing feedback mechanism for active recognition. Emphasis is placed on the feedback mechanism which imitates human selective attention to a particular portion of a character. The result of the first rough classification by a structured neural network is fed back to the early processing of an input character. This feedback mechanism is self-organized by using neural network technology. The recognition ability of the proposed neural network has been evaluated for actual field data. The evaluation results have shown that the self-organizing feedback mechanism considerably improves the recognition ability and that the combination of the human handwriting model and the feedback mechanism leads to a practical level of recognition ability.

Key words: Neural networks, Active learning, Pattern recognition, Handwritten character recognition, Attention

1. Introduction

Neural networks have been investigated for years because they are regarded as effective tools for a variety of applications including character recognition. However, they still need some improvements for practice. One of the problems of a neural network is the time to learn example data. A popular learning algorithm like error backpropagation \cite{1} often needs formidable learning time. On the other hand, human learning is efficient. Once a human learns a given pattern, he can recognize not only the pattern but also unknown distorted patterns of the same category. To simulate such high efficiency on an artificial neural network, we have proposed a new efficient learning named Dynamics-Based Active Learning (DBAL) in the previous paper \cite{2}. It is a combination of a distorted-pattern prediction and an exceptional data treatment. For the distorted-pattern prediction, we have proposed a new handwriting model named Nishimura’s Hand Dynamics Model (NHDM) \cite{3}. It represents the overall influence of the dynamics of a human’s arm, hand, fingers and the other mechanisms on the trajectory of a character written by him. As the handwriting model creates the information on the possible distortions of given patterns, the neural network acquires the ability to recognize unknown patterns without learning a number of field data. The proposed learning DBAL is implemented on a Structure for Active Learning (SAL) network \cite{4} and enables to imitate human efficient learning to a great extent.

Although DBAL has solved the problem in the learning mode of a neural network, another problem has been left unsolved in the recognition mode. While recognition by a neural network is a simple classification of an input pattern into a category, human recognition is a more complicated information processing. For example, when a human encounters a very distorted pattern, he may focus his attention selectively on a particular portion of the pattern after he classifies the pattern roughly. This selective attention is regarded as a feedback from the first rough classification to the early processing of the input. This paper proposes an effective method to simulate such selective attention. The input pattern has been divided into five portions. Each portion has been processed in parallel and integrated on a SAL network. The output of the SAL network, which means rough classification of the input pattern, is fed back to change the relative importance of each portion.

In the study of voltage diagnosis, the feedback mechanism was realized by a rule-based system \cite{4}. However, in the application to handwritten character recognition, it soon turns out that such a simple system is unsatisfactory and time-consuming to build. In this paper, we propose the self-organizing feedback mechanism based on LVQ neural network technology. After discussing a feedback mechanism realized as a rule-based system, the use of a dedicated neural
network for the feedback has been proposed. In the process of DBAL, the dedicated neural network learns how to focus on a portion of an input pattern adaptively. The recognizing ability of a neural network built by DBAL and equipped with the self-organizing feedback mechanism has been evaluated for handwritten digits. The recognition results have shown that the combination of DBAL and the self-organizing feedback mechanism has provided an effective means to build a practical neural network for handwritten character recognition.

2. Dynamics-Based Active Learning

Human learning can be called active learning because he not only memorizes the given patterns but also creates some information from the given data. As our proposed active learning model for handwritten character recognition is based on hand dynamics, we call it Dynamics-Based Active Learning (DBAL) [2]. To implement DBAL, we have utilized Structure for Active Learning (SAL) network [4] and Nishimura's Hand Dynamics Model (NHDM) [3].

SAL network has been developed originally for applications where only a small number of learning patterns are available [5][6]. It prepares a unit to react to a given learning pattern and, at the same time, it prepares some units to react to its variations. The SAL network has to be accompanied by the methods to predict the variations. The hand writing model NHDM [3] has been employed for this purpose in handwritten character recognition. It has simulated human handwriting very well and covered the greater portion of the distorted handwritten patterns [3]. NHDM is given as the following dynamical equation:

$$m\ddot{x} + r\dot{x} = \varepsilon(x_r - x)$$

where, $x_r$ is an actual trajectory vector, $x$ is a predicted trajectory vector, $m$ is a coefficient of inertia of hand mechanism, $r$ is a coefficient of friction of pen point and $\varepsilon$ is a coefficient of attractive force on $x$ towards $x_r$. Examples of the prediction will be presented later.

As shown in Fig.1, this SAL network consists of an input layer, a receptive layer, an associative layer, an output layer and a feedback mechanism proposed in this paper. We also introduce division of an input pattern into groups to achieve selective attention. The SAL network processes the groups in parallel and then turns its attention to one of those groups during the feedback process. It is our future study to optimize the division. In this study, we propose a heuristic dividing method, and evaluate its effect experimentally. Fig.2 shows our proposed dividing method. An input pattern is divided into five groups, namely upper left, upper right, lower left, lower right and center portions. A coefficient $f_n$ shown in Fig.1 determines the relative contribution of the information from the n-th group to the final output [4]. We call $f_n$ the relative contribution ratio. The SAL network focuses on some group by increasing the corresponding coefficient, and the group contributes more to the final result than the others.
3. Self-Organization of Feedback Mechanism

A human selective attention is regarded as a feedback from the first rough classification of an input pattern to the early processing of the input pattern. The early processing means the adjustment of relative contribution coefficients to increase the influence of the important portions on the final classification. In the first rough classification, the output of the SAL neural network is evaluated. When the output cannot give a clear result, the feedback mechanism is activated. The feedback mechanism changes relative importance of each portion of an input pattern, and the classification is attempted again.

The relative contribution ratio \( f_n \) is given for the \( n \)-th group of the associative layer of SAL network. \( f_n \) determines how much the \( n \)-th group contributes to the final output. The output value \( O_{ni} \) of the \( i \)-th unit of the \( n \)-th group of the associative layer is given by the following equation:

\[
O_{ni} = f_n A(I_{ni1}, I_{ni2}, ...) \]

where, \( I_{ni} \) is the output value of the \( j \)-th unit of the \( i \)-th receptor of the \( n \)-th group of the receptive layer, \( A(\cdot, \cdot, ...) \) is the activation function of the unit of the associative layer. The feedback mechanism controls these \( f_n \). They are identical during the first rough classification. After the feedback mechanism has made decision on attention, they are varied so that the \( f_n \) of the attention portion may increase and the others decrease as shown in Fig.3.

First we have realized the feedback mechanism as a rule-based system as shown in Fig.4. A rule is described as follows:

\[
\text{if (the first classification is ambiguous)} \quad \text{then (turn its attention to a particular portion)}
\]

The contents of the rules have been determined statistically to reflect good attention trials which lead failed results to successful results in recognizing sample patterns. For example, we have made the following rule from a trial, that is 'for an input pattern which belongs to the class '5', the first rough classification is failed ('8'), but the final one is successful ('5') by turning its attention to the 3rd group (lower left portion)'.

\[
\text{if (} M_1 = 8 \text{ & } M_2 = 5 \text{ & } d < \theta \text{)} \quad \text{then (} f_3 := 3.0 \text{, } f_i := 0.5(i = 1, 2, 4, 5) \text{)}
\]

where, \( M_1 \) is the class of the most activated unit of the output layer on the first rough classification, \( M_2 \) is the second, \( \theta \) is a threshold and \( d \) is given by:

\[
d = \text{act}(M_1) - \text{act}(M_2)
\]

where, \( \text{act}(M_j) \) is the activation value of the most activated unit and \( \text{act}(M_2) \) is the second. The values such as 3.0 or 0.5 have been determined experimentally. Here, we assume that a result whose \( d \) is smaller than \( \theta \) is ambiguous. It is such that a human cannot well know whether a distorted pattern belongs to the class \( C_1 \) or \( C_2 \).

Although the performance of the rule-based feedback mechanism was good, it required a lot of time to determine the rules, in fact it was complicated to cover all cases. In the above example, \( M_n(3 \leq n \leq 10) \) might be 5. To cover such cases, we had to determine a number of rules as shown in Fig.5. Moreover, the result A and the result B as shown in Fig.6 may be treated similarly or dissimilarly according as the level of the threshold. It is determined experimentally in this study. It takes a lot of time to acquire the optimum threshold and increases experimental decision.
Neural networks are suited to manage multidimensional information like SAL network output. So we have applied a neural network to the feedback mechanism and solved the problems described above. We call it the feedback network (FB network). Fig. 7 shows the overview of SAL network with FB network. FB network treats SAL network output directly as opposed to determining if-then rules. It doesn't take a lot of time. It doesn't need experimental decision like thresholds. Its input is the first result, namely the first SAL network output, and its output is the number of the attention group which lead failed results to successful results. As DBAL proceeds, FB network learns how to focus on a portion of an input pattern and acquires the excellent selective attention ability.

In this experiment, we utilize Learning Vector Quantization (LVQ) [7] for FB network as shown in Fig. 8. LVQ classifies input patterns into the class of the nearest reference vector [7]. The reference vectors are adaptively determined to minimize the classification error. We used the following learning algorithm to determine them.

If \( C_i = C_r, \) \( C_j = C_r, \)
\[
\begin{align*}
    m_i(t+1) &= m_i(t) + \alpha(t)[z(t) - m_i(t)] \\
    m_j(t+1) &= m_j(t) + \alpha(t)[z(t) - m_j(t)]
\end{align*}
\]

If \( C_i = C_r, \) \( C_j \neq C_r, \)
\[
\begin{align*}
    m_i(t+1) &= m_i(t) + \alpha(t)[z(t) - m_i(t)] \\
    m_j(t+1) &= m_j(t) - \alpha(t)[z(t) - m_j(t)]
\end{align*}
\]

If \( C_i \neq C_r, \) \( C_j = C_r, \)
\[
\begin{align*}
    m_i(t+1) &= m_i(t) - \alpha(t)[z(t) - m_i(t)] \\
    m_j(t+1) &= m_j(t) + \alpha(t)[z(t) - m_j(t)]
\end{align*}
\]

If \( C_i \neq C_r, \) \( C_j \neq C_r, \)
\[
\begin{align*}
    m_i(t+1) &= m_i(t) - \alpha(t)[z(t) - m_i(t)] \\
    m_j(t+1) &= m_j(t) - \alpha(t)[z(t) - m_j(t)]
\end{align*}
\]

Fig. 5. Problem of Scale

Fig. 6. Problem of Threshold Decision

Fig. 7. SAL Network with FB Network

Fig. 8. Self-Organized Feedback Mechanism by LVQ
where, \( x(t) \) is the training vector which belongs to the class \( C_r \), \( m_i(t) \) is the nearest reference vector labelled according to the class \( C_r \), \( m_j(t) \) is the second nearest reference vector labelled according to the class \( C_j \). Fig.9 shows an example of the third condition.

In this experiment, we determine the amount to be equal to the amount of training vectors. Their initial values can be randomly set or determined according to the distribution of training vectors. Here, each initial value of reference vectors is set to be equal to each value of training vectors. Therefore, in the initial state, each nearest reference vector of training vectors is each related reference vector. However, as the learning proceeds, the reference vectors move under the influence of the neighbour training vectors, so the nearest reference vector of a training vector may be replaced with another one.

**4. Evaluation of Recognition Rate**

We have evaluated the recognition ability of the proposed neural network for handwritten digits. In this experiment, we have prepared three typical patterns for each category, and nine variations have been predicted by NHDM for each typical pattern. Fig.10 shows some examples of the variations. The SAL network acquires the recognition ability by creating the units to react to those patterns. First we have evaluated the recognition ability by using 36,000 handwritten digits, and then treated the unrecognized digits as exceptions. The SAL network creates exceptional units to memorize those unrecognized digits. After that, we have evaluated the neural network again for the same data. Besides, we have obtained attention samples from the trials and made FB network learn them. Finally, we have evaluated the neural network with the self-organizing feedback mechanism. Therefore, the experiment consists of following three phases.

1. Recognition by a neural network built by only NHDM prediction
2. Recognition by a neural network built by only DBAL
3. Recognition by a neural network built by DBAL and FB network (proposed method)

In order to evaluate the recognition ability for unknown patterns, we have prepared another data set consisting of 10,000 handwritten digits which are not overlapped with the above 36,000 data set. The 36,000 data corresponds to learning or training data for a conventional statistical method or neural network.

Table 1 shows the recognition rates obtained by a conventional statistical method, the neural network built by only DBAL and the neural network built by DBAL and FB network. Although only DBAL leads to the very good recognition ability compared to that by the conventional statistical method, the addition of the feedback mechanism results in considerable improvement of the recognition ability. This result shows that the combination of DBAL and the self-organizing feedback mechanism is a practical method in terms of efficiency and recognition ability.
5. Conclusion

In this paper, a new method has been proposed to build a practical neural network for handwritten character recognition based on the idea of active learning and active recognition. The active learning has been realized by Dynamics-Based Active Learning, and the active recognition has been realized by a self-organizing feedback mechanism. A dedicated neural network based on Learning Vector Quantization has been used to implement the proposed feedback. When a very distorted pattern is given, the feedback mechanism amplifies the similarity of the pattern to some category by focusing on a particular portion of the pattern.

The proposed neural network has been evaluated in terms of recognition rates for 36,000 and 10,000 handwritten digits. The recognition rates exceed those by a conventional statistical method. They reached 99.0% for 36,000 data (corresponding to learning data) and 95.6% for 10,000 unknown data. The result has shown that our proposed method provides an effective means to build a practical neural network for handwritten character recognition.

(Manuscript received September 22, 1995)

References


Naotake Natori (Member)
He was born in 1967. He received the B.E. and M.E. degrees in electronics and telecommunications from Waseda University, Tokyo, Japan in 1989 and 1991, respectively. He was a researcher at Systems and Software Engineering Laboratory, Toshiba Corporation from 1991. In 1995, he moved to Multimedia Engineering Laboratory, Toshiba Corporation. His research interest includes neural networks and pattern recognition.

Kazuo Nishimura (Member)
He was born in Kanagawa, Japan, in 1951. He received the B.S.E.E., M.S.E.E., and Ph.D. degrees from Tokyo University in 1974, 1976, and 1979, respectively. In 1979, he joined Toshiba Corporation, where he was engaged in studies on power transmission systems. From 1983 to 1985, he was with the Massachusetts Institute of Technology, Cambridge, MA, as a Visiting Scientist. Since 1987, he has been working for the Systems and Software Engineering Laboratory of Toshiba. His research interests are neural networks and their applications. Dr. Nishimura is a member of IEE of Japan, IEEE, and International Neural Network Society.