A New Algorithm for Handwritten Character Recognition[1]

Zhu Xiaoyan, *Member, IEEE*, Shi Yifan
State Key Lab of Intelligent Tech. & Systems,
Dept. of Computer Science & Technology, Tsinghua University, Beijing 100084, China

Abstract:
In this paper, a new algorithm of handwritten character recognition based on feedback theory is proposed. We suggest this new method by adding confidence back-propagation and input modification on the Neural Network model, thus preprocessing and recognition are integrated closely. Convergence of the algorithm is proved. Experiments show that it greatly reduced the system's error rate and was robust to environmental noise.

I INTRODUCTION

OCR systems are widely used in various fields of our life, but there still remained many problems in the OCR techniques, one of which is the well-known contradiction of correction rate with rejection rate. It means that in the current OCR system, if we enhance rejection to avoid error recognition, the overall recognition rate will be correspondingly lower. In order to solve the problem, we should look inside the OCR techniques to find out the limitations.

Current OCR techniques are mainly based on the forward structure of signal flow which process the input signal sequentially through segmentation, pre-processing, normalization, feature extraction and classification, without any feedback[1-4]. But feedback is one of the major patterns of human behavior, and is also one of the major concepts in the modern cybernetic systems. When we recognize a blurred word through our eyes, we just look it over and over with guess and expectation, then we get rid of additional noise and find the correct answer. That is a procedure including feedbacks between each pairs of recognition steps.

Thus we can introduce the concept of feedback into the character recognition system structure where the sample image, dictionary information and environment are considered as the input signals of controller, the feature extractor and classifier as controlled objects and the methods selection module as the controller, through all of which a close loop recognition system is formed. Therefore the output recognition result can be a feedback to the controller to adjust the sample signals and environmental information processing procedure. Because of the supervisor ability and the robustness characteristic to the environment of feedback, the whole recognition system error decreased, and the system's robustness to the initial variables, noise and nonlinear distortion improved. It also provides a way to optimize the system's performance alternatively.

In this paper, we proposed an algorithm to realize part of the feedback strategy by modifying the input image in the feature phase through the feedback of the recognition result recursively. Hence, the classification phase and pre-processing phase are integrated together closely in a way that the samples are adapted to the specific character classifier to get rid of the additional noise from the original pattern, thus increase the recognition rate while decrease the rejection rate. In the algorithm, an artificial neural network (ANN) structure is introduced as the acquiring and processing part of the feedback signal[5,6], which we have worked on for many years[7,8]. Experiments also showed that this algorithm really improved the system's performance and robustness.

This paper is organized in the following manner. In Section II, we describe the connection between the classification result and feedback and the method of acquiring the feedback signal based on the ANN. The visual analysis is also proposed. In Section III, we introduce the recognition criterion for evaluation and then the system architecture. Section IV is the experiment and discussion and section V is the conclusion.

II FEEDBACK METHOD

In the feedback system, how to acquire and process the feedback signal is most important. Here we use ANN as the forward classifier and the output of the ANN classifier as the input signal of feedback processing module. The feedback ANN processes the feedback signal and adjusts the system input signal accordingly, thus completes the close loop system.

2.1 Relationship between classifier result and feedback

In all recognition system, input samples are first mapped from image space to feature space (feature extraction), which is divided into various areas by

[1] *This work is supported by National Nature Science Foundation of China (69982005), and foundation for the university key teacher by the Ministry of Education*

0-7803-6725-1/01/$10.00 ©2001 IEEE

classification rules, then are classified through deciding which areas they are belong to. Feature extraction procedures are nonlinear and are usually different according to the definitions of the feature. Also, there exist many methods of how to divide the feature space, one of which is the ANN who has been successfully applied to the handwritten character recognition.

In our system, traditional MLP is used to obtain the classification result and the according confidence, and the training algorithm of which is the well-known BP algorithm, described as follows:

Suppose the input vector $x \in R^n$, $x=(x_0, x_1, \ldots x_{n-1})^T$, there are n' neurons in the mid-layer as $x' \in R^{n'}$, $x'=(x_0', x_1', \ldots x_{n'-1}')^T$, and there are m output neurons $y \in R^m$, $y=(y_0, y_1, \ldots y_{m-1})^T$。 The weights between the input layer and the mid-layer are w_{ij}, while w_{jk}' are the ones between the mid-layer and the output layer.

The system outputs is described as follows:

$$y_k = f(\sum_{j=0}^{n'-1} w'_{jk} x'_j) \qquad (1)$$

where,

$$x'_j = f(\sum_{j=0}^{n'-1} w_{ij} x_i)$$

$$f(x) = \frac{1}{1+e^{-x}} \qquad (2)$$

The translation function $f(x)$ in the equation (1) is continuous, the classification result can be represented analytically by feature vectors as follows:

$$\vec{y} = g(x, w) \qquad (3)$$

When training the MLP network by BP algorithm, we use $\left|\vec{y} - \vec{T}\right|^2$ as the objective function, where $\vec{T}$ is the teacher vector. Then the Gradient Descent method is applied to find the optimal resolution of w by searching in the w space recursively for the local optimal, where f is fixed, x is constant, and $\partial\left|\vec{y} - \vec{T}\right| / \partial w$ is the gradient.

In formula (1), we see that x and w is exchangeable. So we can consider w as constant, and use the $\partial\left|\vec{y} - \vec{T}\right| / \partial x$ to search the optimal solution of x in the vector space. In this way, we can achieve the effect of pre-processing by modifying the feature vector, where the modification made is decided by the output confidence, i.e. the feedback signal.

The traditional pre-processing was usually made in the image space, which modified the input image where the modification was represented implicitly by the change of feature vectors. But by our ANN feedback method, we can modify the feature vector directly, so long as the modification is reasonable. In fact, the modification is the guess of the blurred stokes. Thus we find out an feedback algorithm of modifying the image in the teaching of the classification result, and re-classifying to get more accurate result.

2.2 Acquiring the feedback signal

In the training phase, the teaching signal $\vec{T}$ is set by human, while in the feedback acquiring phase, $\vec{T}$ is the unknown classification result. So we need to guess the result, and relpace $\vec{T}$ with the guess which is feedback to get the gradient of x_i., then modify x_i to x_{i+1} to complete one iteration.

Suppose the guess result $T' \in R^m$, and the teach signal of one sample $t=(0,..,0,1,0..,0)$, where $t_i=1$, $t_j=0$, $j \neq i$, then the objective error function can be:

$$E = \frac{1}{2}\left|\vec{y} - \vec{t}\right|^2 = \frac{1}{2}\sum_{k=0}^{m-1}(y_k - t_k)^2 \qquad (4)$$

The gradient of the energy of the error function by the input vector x can be:

$$\frac{\partial E}{\partial x_i} = \sum_j \frac{\partial E}{\partial x'_j}\frac{\partial x'_j}{\partial x_i} = \sum_{j=0}^{n'-1}\sum_{k=0}^{m-1}(y_k - t_k)y_k(1-y_k)w'_{jk}\frac{\partial x'_j}{\partial x_i}$$
$$= \sum_{j=0}^{n'-1}\sum_{k=0}^{m-1}(y_k - t_k)y_k(1-y_k)w'_{jk}x_j(1-x_j)w_{ij}$$
$$= \sum_{k=0}^{m-1}((y_k - t_k)y_k(1-y_k)\sum_{j=0}^{n'-1}(w'_{jk}x_j(1-x_j)w_{ij})) \qquad (5)$$

So when iterating s times, we can get the modification representation of the input feature $x(s)$ as follows:

$$x_i(s+1) = x_i(s) - \varepsilon\frac{\partial E}{\partial x_i}(s) \qquad (6),$$

where ε is the step length.

The iteration ends until the error decrease is little than a threshold, *MinGate*.

2.3 Proof of Convergence

The algorithm mentioned above will be converged to a local optimal point in some limited steps. The convergence of the algorithm is proved as follows:

In the gradient descendent algorithm, the objective function, E, has a downward trend. Since E has a boundary, if it decreases by at least *MinGate* in one iteration, the ending condition can be attained in some limited steps, even the decrease is less than *MinGate*. This means that the iteration enters into the local minimum. The only condition that may cause non-convergence is that in one step, $x(s+1)$, is modified to traverse the local minimum, and $E(s+1) > E(s)$. At this time, the iteration cannot be continued.

From Eq. (5), it is indicated that $\partial E/\partial x$ should be bounded if all w, w', y and x are bounded,. Because $|x|$ and $|y|$ are both less than 1, $\partial E/\partial x$ must has a boundary decided only by w and w'. Since they remain constant during the recognition, the boundary will be a constant, T.

From Eq. (6), $|x(s+1) - x(s)| < K\varepsilon$, then $|E(s+1) - E(s)|$

$< KT\varepsilon$. This means that the iteration step can be decided solely by ε, which is a descending sequence. In this way, the traverse of the local minimum can be avoided and attain the ending condition finally, i.e., the algorithm converges. This means that the proof of the convergence has been finished.
In our experiments, it shows that the initial value of ε is different according to various features. This is probably because of the diversity of the feature space distributions. If ε is set larger, the faster the system converges, but it may cause the modified image too far away from the original image. So, we should get ε by observing the up-bound of the different features, which makes the iteration error decreases in the first several steps. For the used features, ε, it often takes 0.01, 0.03 and 0.1, and the actual iteration number is lower than 10 in our experiments.

2.4 The visualized feedback

Modifying the feature directly is different to modifying the image, because the latter is limited by connectivity and smoothness condition. So for some features, the point in the feature space can't be all mapped to a legal image. Here we select the gray scale feature, which has the explicit mapping relationship with the image space, to visualize the feedback result.

Figure1.(a) and (b) are original image. Suppose the teacher vector is $(0,..,0,9)^T$, (a) converges to (c), and (b) converges to (d). We can see clearly that the feedback algorithm can fill up the gap and eliminate the noise effectively.

III RECOGNITION CRITERION

By now, we still have a basic problem of definition of modification cost to be solved, which decides ultimately the recognition results. It should consider the following two factors:
(1) The difference between the converged feature x and the original x_0: It reflects the amount of modification made to the original image, e.g., it is the total number of pixels to be modified for the gray scale feature. So $|x-x_0|$ should be the main part of the modification cost.
(2) Converged output confidence, *conf*: The modified image may still be different from what we expect it should be after the feedback iteration. The larger the difference, the lower the output confidence. So the output confidence can reflect the degree of difficulty for modification through feedback, and it should be one of the factors in the modification cost.

From the above analysis, the evaluation of modification cost function can be defined as follows:

$$Eval(x)=|x-x_0|*|1-conf(x)| \qquad (7)$$

A direct way of guess the teacher signal is to try all the patterns at first, which may be time consuming. So we first limit the guessing candidates guided by the initial forward classification results, which is obtained by the MLP during the first iteration. Then on a large sample set, we can count the error distribution matrix, and take out the most frequent candidate to form the guess array $R=\{R_i\}$, e.g.$R=\{R_0,R_1,R_2,...,R_M\}=\{\{a_{ij}\}\}$,

Finally, we use $\{a_{ij}|a_{ij}\in R_i\}$ as the guessing candidates if the classification result of the first iteration is pattern i. So, the search time is greatly decreased by the above method.

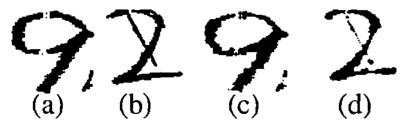

Figure.1 Effect of feedback modification

IV Experiments

4.1 Artificially added noise

In order to test the system's performance on the standard image, 100 samples for each digit were selected from the NIST, thus composing a base test set T_0 of 1000 samples. The set was very clean, without any surplus strokes and additional noise, but the style of the handwriting was diverse

From T_0, six set was made by artificially adding noise of different type and degree, i.e.: T_1, little blurred (the width is 1, and the length is 3); T_2, severely blurred (the width is 1, and the length is 3); T_3, with top and right frames reserved (the frame width is 1); T_4, with top and right frames reserved (the frame width is 3); T_5, with gapped strokes (gap width is 1, length is 6); T_6 with gapped strokes (gap width is 1, length is 6). The result is as follows.

From table 1, the feedback system's performance is greatly improved compared with the original system. For set T_1 and T_6, which have little blurs and gaps, although the original recognition rates are already more than 95%, the feedback system can still improve it. For set T_2, T_3 and T_4, which have severe blurs, gaps and surplus strokes, the original system could not recognize them normally with the recognition rate of 50% and 20%, while our feedback system improves it to a high level of 73%, 89% and even 95%. This means that the feedback architecture can make full use of every sub-system, thus enhance the system performance.

As to the speed, our feedback system is 39.7% lower average than the original system, but it can be made up by the fast development of hardware. One thing is to know that the lower the speed, the more the recognition rate improved, e.g. T_3 and T_4. So, the cost in speech is worthwhile.

Table 1 Results of artificially added noise samples

	System	Error (%)	Reject (%)	Acc (%)	Speed (W /s)
T_1	Original	4.40	0.70	94.90	90.1
	FB	3.40	1.40	95.20	58.8
T_2	Original	33.60	6.70	59.70	76.9
	FB	25.70	5.50	68.80	55.6
T_3	Original	27.70	22.10	50.20	83.3
	FB	8.20	2.00	89.80	29.4
T_4	Original	33.90	39.10	27.00	76.9
	FB	9.70	0.40	89.90	19.7
T_5	Original	2.40	0.20	97.40	83.3
	FB	2.50	0.00	97.50	55.6
T_6	Original	3.70	1.00	95.30	76.9
	FB	2.80	0.50	96.70	55.6

4.2 Natural samples

In this experiment, 5 naturally acquired sample sets Nist, OCR1, 2, 3, 4 are used. Nist is the American standard sample set. OCR1, 2, 3, 4 were collected by ourself, which has a lot of gaps, many prolong strokes, many segmentation errors and has a style quite different from other sets, respectively. The experiment results are shown in table 2

Table 2 Recognition results of natural sample sets

	System	Error (%)	Reject (%)	Acc (%)	Speed (W /s)
NIST	Original	0.38	0.05	99.57	125.6
	FB	0.39	0.04	99.57	121.6
OCR 1	Original	4.60	1.34	94.06	111.6
	FB	4.76	0.57	94.67	96.9
OCR 2	Original	2.09	0.57	97.34	116.3
	FB	1.83	0.17	98.00	103.1
OCR 3	Original	4.93	1.57	93.50	111.2
	FB	3.80	0.67	95.54	91.0
OCR 4	Original	3.23	0.74	96.03	118.4
	FB	3.35	0.54	96.11	102.4

From table 2, we can see that in NIST set, which has good handwriting quality, our system made no deference in the recognition speed with the original system. So, on this good condition, which has many limitations in handwriting, the speed decreased little, whereas the recognition rate was nearly 100%. In OCR1, 2, 3, 4, factors such as gap, prolonged stroke and segmentation, greatly affected the original system, but not our feedback system, which represented great robustness to different types of noise. Further more, it improves a lot in recognition rate with only a 10% decrease in the speed. OCR 4 set has a handwriting style not emerged in any training set. The results show that our system is also better than the original system even in processing brand new patterns.

Through above two experiments, we compared our system with the original system in both designed environments and real applications. The experiments show that the feedback architecture greatly improves the system's performance.

V CONCLUSION

We successfully applied the feedback principle in the training phase of ANN to the classification and modification phase of the digital recognition system based on ANN. The new algorithm didn't lower the recognition speech when the quality of the input images was good, and greatly improved the system's robustness when the samples were blurred by noise. It's a practical and effective approach. Since the algorithm is based on the BP network, which is quite powerful, it can be applied to other methods of pattern recognition.

However, it is only an attempt to utilize the confidence as the feedback signal, there may exist many other ways. Further more, the image generated by the feedback modification may be evaluated in detail by other analysis techniques, new applications are thus expected.

REFERENCES

1. R.M.Brows, T.H.Foy "Handprinted symbol rocognition system", Patt.Recog., Vol.21,No2, p981-118,1988
2. F.Kimura, M.Shiridhar "Handwrittten Numerical Recognition based on multiple algorithms", Patt.Recog. 1991,Vol.24, No10, p969-983
3. Michael D Garris,Charles L Wilson, "neural network-based systems for handprinted OCR applications", IEEE Tran. Image Processing, 1998,Vol.7, No8, p1097-1112
4. F.Kimura, M.Shiridhar "Handwrittten Numerical Recognition based on multiple algorithms", Patt.Recog. 1991,Vol.24, No10, p969-983
5. R.M.Brows, T.H.Foy "Handprinted symbol rocognition system", Patt.Recog., Vol.21,No2, p981-118,1988
6 A.W.Senior,F.F.allsid "Off-line handwiting Recognition by Recurrent Error Propagation Networks", Cambridge Univ Engineering Department.
7 Zhu Xiaoyan " Multiple Neural Networks Model and its Application in Pattern Recognition " , IEEEE International Conference on Neural Information Processing, 1995, pp.996-969, Beijing
8 Wang Song and Zhu Xiaoyan " Multiple Experts Recognition System Based on Neural Network " , Proceedings of International Conference on Pattern Recognition Proceedings, 1996, pp.545-548, Vienna, Austria