A Fuzzy Genetic Classifier for Recognition of Confusing Handwritten Numerals 4, 6, and 9

°Dae Jung Shin*, Seung You Na+, Sun Hee Kim*

*Graduate College, Chonnam National University, Puk-ku, Kwangju 500-757, KOREA
+Dept. of Electronics Eng., Chonnam National University, Puk-ku, Kwangju 500-757, KOREA
Tel: +82-62-520-6394, Fax: +82-62-520-6390, E-mail: syna@chonnam.chonnam.ac.kr

Abstracts A Fuzzy Classifier which deals with very confusing objects is proposed. Naturally this classifier heavily relies on the multi-feature decision-making procedure. For a simple example, this classifier is applied to the recognition of confusing handwritten numerals 4, 6 and 9. The characteristic variables used in this paper are the existence of a loop and the relative location of the starting or ending points(SEP). Thus each sample of handwritten numerals 4, 6 and 9 is classified in one of the 6 groups which are divided according to the sample structure. Each group has its own classifying rules. Also the method of rule-generation using genetic algorithms in each group is proposed.

Keywords Multi-Feature Decision-Making, Fuzzy Classifier, Genetic Algorithms, Rule-Generation

1. INTRODUCTION

Letter recognition, in the pattern recognition, especially number recognition has been given much research for a long time, and there exist many methods of recognition about it. The methods can be classified into 4 categories.[1][2]

Among these are included: (1) the statistical method that classifies by using statistical features of characteristic vectors extracted from numeral image data; (2) the structural method which uses structural features of letters such as inflection point, contour, and skeletons; (3) the method of applying neural network that models human cognition procedure; and (4) the method of template matching that recognizes numerals under recognition and by obtaining similarity and distance.

Among these 4 methods, the structural method is based on the principles of numeral composition and recognizes the numerals by inferring the relationship between each numeral and its basic components. This method has an advantage: it is adequate for recognizing the numeral sets which have complex structures, and handwritten form which has much variance in shape.

However, it is hard to extract the characteristics that are fundamental for recognition algorithm. Also, there has not been enough research on the automatic generation of rules expressing the numeral structure and the grammatical inference method which is based on the rules.

In this paper, we try to solve the problem of the characteristics extraction by using relative location of SEP and Loop as a characteristic vector. Also, it proposes the method of generating the classification rules easily, which classifies the structurally similar data into groups and uses Genetic Algorithms. For the recognition of confusing handwritten numerals 4, 6 and 9, we propose the FGC(Fuzzy Genetic Classifier) which uses human knowledge in designing procedures.[3] The use of fuzzy logic for using human knowledge and for dealing with fuzzy information has been confirmed for the validity in many areas.[4][5][6]

Section 2.1 deals with the characteristic variables which are used in this experiment, section 2.2 deals with the method of classifying the similar classes into groups, and section 2.3 explains the automatic generation of the rules using Genetic Algorithms. Chapter 3. examines the results of recognition of 84 printed numerals and 330 handwritten numerals, and finally a conclusion is given.

2. FEATURE CLASSIFICATION AND GAs

2.1 The extraction of characteristic vectors

In the classification method using the structural characteristics, the exact extraction of features is an important concern for the high rate of the numeral recognition. However it's not easy to obtain the exact characteristics such as an inflection point about the very confusing handwritten numerals. Accordingly, for the binary input data we extracted the most primary units, that is, a starting point, an ending point and the location of a loop by using the contour which is exact and available easily.

According to the characteristics of the obtained features, the handwritten input data are divided into 6 groups, as shown in section 2.2

2.2 Classification of groups

We can find the structural characteristics of the handwritten numeral input data through characteristic vectors which are obtained in section 2.1. By classifying the handwritten numerals into the groups which have the same structural characteristics, the following advantages can be obtained.

First, the data treatment can be easily made by the equal number of feature elements in a group. Second, the automatic classifying rule generation can be handled by dividing the complex objects into subgroups in which numerals are similar each other. According to the obtained characteristics, sample data are divided into 6 groups.

Examples of confusing handwritten numerals are given in Fig. 1 and Fig. 2. A right middle SEP is not detected in (a) of Fig. 1, thus both (a) and (b) are classified in $Group_2$ which has 0 Loop and 3 SEP's.

(a) 4 (b) 9

Fig 1. Confusing Numerals 4 and 9 in $Group_2$

Table 1. Classification of Groups

N_{Loop}	N_{SEP}	Group
0	2	$Group_1$
0	3	$Group_2$
0	4	$Group_3$
1	1	$Group_4$
1	2	$Group_5$
1	3	$Group_6$

N_{Loop} : Number of Loop
N_{SEP} : Number of SEP

2.3 Automatic generation of classification rules using GAs

Genetic Algorithms are the search algorithms which are based on the genetics and 'the survival of the fitness' in the natural system. GAs have excellent capabilities, especially in the areas of optimization and pattern classification.[7][8][9] In this paper, GAs are used for generating the fuzzy classification rules.

$Group_5$ of section 2.2 is taken as an example of automatic generation of rules using GAs. $Group_5$ has one Loop and 2 SEP's in terms of its structural properties.

Among the handwritten numerals which are classified into $group_5$, a very confusing example is shown in Fig. 2.

Let the coordinates of an upper SEP be (X_1,Y_1), a lower SEP be (X_2,Y_2) in $Group_5$. It is natural to deduce a rule by the locations of SEP's relative to the loop in $Group_5$.

(a) 4 (b) 6 (c) 9

Fig. 2. Confusing Numerals 4, 6, and 9 in $Group_5$

With the introduction of this kind of human knowledge, two coordinates of SEP's can be transformed as follows:

$$\{X_1, Y_1, X_2, Y_2\} = \{X_{1Left}, X_{1Center}, X_{1Right}, Y_{1Upper}, Y_{1Middle}, Y_{1Bottom}, X_{2Left}, X_{2Center}, X_{2Right}, Y_{2Upper}, Y_{2Middle}, Y_{2Bottom}\}$$

Therefore, the elements of a string for GAs which are used for rule generation are 12 bits. The membership function for generating the fuzzy rules uses the statistical values of each characteristic variables: minimum, average, and maximum.

The string and fuzzy rules which represent the classification rules obtained by using GAs are shown in Table 2 and Fig. 3, respectively.

Each 3 bits of the string is one unit; center "1" of 3 bit unit means "OR"; and exterior "1" means "AND", Also "0" means that it is not included as a variable for the rules.

In Table 2, the final values of 3 bits classifying "6" are "010" and "100" and the others are same. Therefore, the fuzzy classification rules for "6" can be made such as "000 100 100 110".

The generated fuzzy classification rules are used as an effective one so far as they have no mismatch. And then we accumulate the fuzzy classification rules until a required matching rate is obtained. Unless the desired rate is obtained, the classifier will require even more characteristic vectors. In this case, the designer should add more effective characteristic vectors and obtain an adequate fuzzy classification rule for the additional characteristic vectors.

Table 2. Strings representing Fuzzy Rules

String	Rule	Pattern
000 100 000 001	Rule 1.	"4"
000 100 100 010	Rule 2.	"6"
000 100 100 100	Rule 2.	"6"
100 011 000 001	Rule 3.	"9"

Rule 1.
IF Y1 is Upper, AND Y2 is Lower, THEN it is "4".
Rule 2.
IF Y1 is Upper, AND X2 is Left, AND Y2 is Middle OR Upper, THEN it is "6".
Rule 3.
IF X1 is Left, AND Y1 is Middle OR Lower, AND Y2 is Lower, THEN it is "9".

Fig. 3. Fuzzy Classification Rules of $Group_5$ generated by GAs

3. RESULTS OF RECOGNITION

To test the effectiveness of the proposed method, it is applied to 330 handwritten numerals obtained from 11 people and 84 printed numerals. Table 3 shows the results of recognition for the 3~6 groups when the less significant groups are either excluded or added.

For the printed numerals, all numerals have a loop and the obtained SEP's are less than 5. For the handwritten numerals, the highest matching rate of 97.58% is obtained when the samples are divided into 6 groups in which all have similar feature values. The rate can be improved by adding more features.

The result of the experiment demonstrates the use of independent classification features is more effective method.

Table 3. Recognition results

Number of groups	Printed Numerals	Handwritten Numerals
3	83.33%	51.27%
4	100%	80.61%
5		89.45%
6		97.58%

4. CONCLUSION

In this paper, the problems of feature extraction and the reduction of characteristic vectors are solved by using the relative coordinates of SEP's and a loop as feature values. Also by dividing the whole sample into subgroups where data have similar feature values, the classification rules can be generated with less computation using GAs. In this way, we can extend the sample data to the whole numerals with little additional groups.

The classification method of this paper can be extended and should be applied to the handwritten alphanumeric data to confirm the effectiveness.

References

[1] J. C. Bezdek and S. K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, 1992

[2] R. J. Marks II, *Fuzzy Logic Technology and Applications*, Technical Activities Boards, 1994

[3] M. Garman, *Psycholinguistics*, Cambridge University Press, 1990

[4] A. K. Nath and T. T. Lee, "On the Design of a Classifier with Linguistic Variables as Inputs", Fuzzy Sets Syst., vol. 11, pp. 265~285, 1983

[5] S. K. Pal and B. C. Chakraborty, "Fuzzy Sets Theoretic Measure for Automatic Feature Evaluation", IEEE Trans. Syst., Man, Cybern., vol. SMC-16, no. 5, pp. 754~760, Sept./Oct. 1986

[6] W. Homenda, "Fuzzy Reasoning with Feature Cumulation as Cluster Analysis", FUZZ-IEEE'94, vol.4, pp. 1922~1926, 1994

[7] D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley, 1989

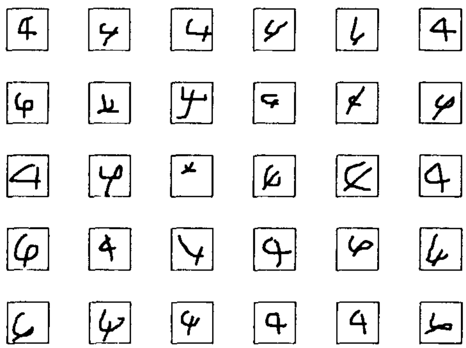
[8] K. D. Jong, "Learning with Genetic Algorithms: An Overview", Machine Learning 3, pp. 121~138, 1988

[9] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, "Acquisition of Fuzzy Classification Knowledge Using Genetic Algorithms", FUZZ-IEEE'94, pp. 1963~1968, 1994

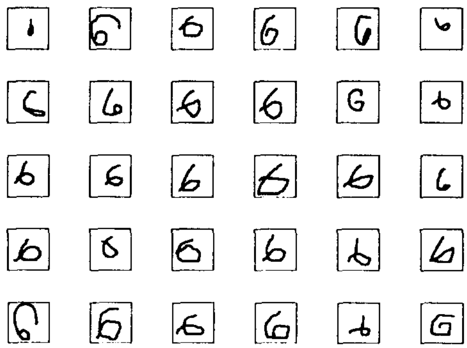
Appendix

Samples of handwritten numerals used for tests

Numeral 4



Numeral 6



Numeral 9