Fuzzy-Rough Set Based Nearest Neighbor Clustering Classification Algorithm

Xiangyang Wang, Jie Yang, Xiaolong Teng, and Ningsong Peng

Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China
{wangxiangyang, jieyang}@sjtu.edu.cn

Abstract. We propose a new nearest neighbor clustering classification algorithm based on fuzzy-rough set theory (FRNNC). First, we make every training sample fuzzy-roughness and use edit nearest neighbor algorithm to remove training sample points in class boundary or overlapping regions, and then use Mountain Clustering method to select representative cluster center points, then Fuzzy-Rough Nearest neighbor algorithm (FRNN) is applied to classify the test data. The new algorithm is applied to hand gesture image recognition, the results show that it is more effective and performs better than other nearest neighbor methods.

1 Introduction

Nearest Neighbor (NN) algorithm is one of the most important classification methods for pattern recognition. Rough Set [2] takes the imprecise features description of the elements in universe as one of the reasons to wrongly classification. Fuzzy-Rough sets [3,4] combines rough set with fuzzy set to handle datasets with both roughness and fuzziness. In [5], a Fuzzy-Rough Nearest Neighbors (FRNN) algorithm is proposed, which introduces roughness uncertainty into Fuzzy KNN [1]. FRNN takes every training pattern as the neighbor of the test pattern with different fuzzy similarities, and it is more robust than Fuzzy KNN. However, for each test pattern it has to compute the similarity for every training pattern, which leads to more complex.

Our algorithm is proposed to reduce the complexity of FRNN and improve the speed of classification. In [9], based on the properties of rough sets, Lingras P extents the K-means and Fuzzy C-means to represent clusters as interval set. In our algorithm, Mountain Clustering method is used to select representative cluster center points from training data, which is different from Lingras P's method.

2 Fuzzy-Rough Set and Fuzzy-Rough Nearest Neighbor Algorithm

For input data set $X = \{x_1, x_2, ..., x_n\}$ with total C classes, let R an equivalence relation on X, and $[x]_R$ is the equivalence class contains $x \in X$. For any output class $c \subseteq X$, lower and upper approximation is: $\underline{R}(c) = \cup\{[x]_R \mid [x]_R \subseteq c, x \in X\}$ and

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613, pp. 370 – 373, 2005.

$\overline{R}(c) = \cup\{[x]_R \mid [x]_R \cap c \neq \phi, x \in X\}$. Rough set is the set of lower and upper approximation, i.e., $R(c) = <\underline{R}(c), \overline{R}(c)>$. When the equivalence classes in rough set are imprecise, they will be in the form of fuzzy clusters $\{F_1, F_2, ..., F_H\}$ and X is generated by fuzzy weak partition. Here, every fuzzy cluster F_j is a fuzzy set. The output class c can also be fuzzy and in the form of lower approximation $\underline{c}$ and upper approximation $\overline{c}$:

$$\mu_{\underline{c}}(F_j) = \inf_{x \in c} \max\{1 - \mu_{Fj}(x), \mu_c(x)\}, \forall x \tag{1}$$

$$\mu_{\overline{c}}(F_j) = \sup_{x \in c} \min\{\mu_{Fj}(x), \mu_c(x)\}, \forall x \tag{2}$$

$<\underline{c}, \overline{c}>$ is a fuzzy-rough set, $\mu_c(x) \in [0,1]$ is the fuzzy membership of x to class c. The fuzzy-rough membership function of x to class c is defined as:

$$\tau_c(x) = \frac{1}{|X|} \sum_{y \in X} \tilde{\mu}_x(y) \mu_c(y) \tag{3}$$

where $\tilde{\mu}_x(y)$ denotes the fuzzy similarity between x and y. In our algorithm, we adopt the similarity, $\tilde{\mu}_x(y) = \exp(-\frac{\|y-x\|^2}{\beta})$ [8], $\|y-x\|$ is the distance in Euclidian norm, and $\beta = \frac{\sum_{i=1}^{n} \|x_i - \overline{x}\|^2}{n}$ is the normalized term, where $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Fuzzy-rough nearest neighbor algorithm (FRNN) [5] is based on fuzzy-rough set. It computes the fuzzy-rough membership value of test pattern to each class according to equation (3). The test pattern is assigned to the maximal fuzzy-rough membership value related class.

3 Fuzzy-Rough Set Based Nearest Neighbor Clustering Classification Algorithm

The steps of our new algorithm are as follows.

(1) By using FRNN, assign fuzzy-rough membership value to each training pattern.

For every pattern in the training set, use leave-one-out FRNN algorithm to compute the fuzzy-rough membership value to all classes. The value implies that to what degree others support this pattern. The higher the value is the less the fuzzy-roughness exists in the neighborhood.

(2) Use muti -edit-nearest-neighbor algorithm to filter the training set

After step (1), there may exist some samples, whose maximal fuzzy-rough membership value related class is inconsistent to their known class label. Such sample points are regarded as in the class boundary or overlapping region. We use muti-edit-nearest-neighbor algorithm [6] to edit them to improve the classifier to test set.

(3) Select cluster representative points from the edited training set.

Mountain Clustering is an approximate clustering method [7]. It firstly constructs Mountain Function from dataset, and then uses Destruction Function to destroy the Mountain step by step and acquire cluster center points. These points can represent the distribution of each class.

For class c, firstly we select the sample point x_j whose fuzzy-rough membership value to c is the highest (recorded as maxc). We take x_j as the fist cluster center point of class c. Then we use the following equation to update the fuzzy-rough membership value of all training patterns to class c:

$$\tau_c(x_l) = [1 - \tilde{\mu}_{x_j}(x_l)]\tau_c(x_l), l = 1,2,....,n \tag{4}$$

Equation (3) can be seen as the Mountain Function, and equation (4) as Destruction Function. The new point with the highest fuzzy-rough membership value is selected as the next cluster center point. Such a process continues until the ratio of the lately selected maximal fuzzy-rough membership value to maxc is less than a given threshold.

(4) Classify the test set with selected cluster representative points with FRNN.

With the selected cluster representative points in step (3), we use FRNN algorithm to classify the test set. Since the cluster representative points are usually far less than original training set, the classification speed can be improved greatly while keep the same or even better classification accuracy than FRNN.

4 Experiments and Discussions

We apply the proposed algorithm, to hand gesture recognition. All the captured images are normalized as 36×36 pixels in gray-level. The dataset contains total 30 kinds of gestures. Each kind represents a single or double letter, as shown in Fig.1.

In the hand gesture recognition problem, since there exist much difference in the same kind of gesture images, the rough uncertainty exists in the input data. On the other hand, different kinds of gesture images may be very similar, e.g. in Fig.1, the different gestures *h*, *x* and *i* are very similar. The classifier tends to wrongly classify these similar gesture images, which is due to the fuzzy output class. We use fuzzy-rough set based FRNNC algorithm to deal with such data.

There are total 4152 images, about 150 images for each kind. 2/3 of the images (3000 samples, 100samples per class) are randomly selected as training set, and the left 1/3 as test set. The algorithms, NN, KNN, Fuzzy KNN, FRNN and FRNNC are evaluated on such dataset respectively. The experimental results are given in Table 1.

Our algorithm performs better than KNN and Fuzzy KNN. Compared with FRNN, it has less computation complexity and higher classification speed, and its classification accuracy is not bad than that of FRNN. The algorithm has two properties: (1) it considers both the fuzziness and roughness in data; (2) It is especially suitable for numerical data with good clustering character.

In addition, the similarity computation method for samples is the key, which directly affects the classification result. The improvement to it will be the future work.

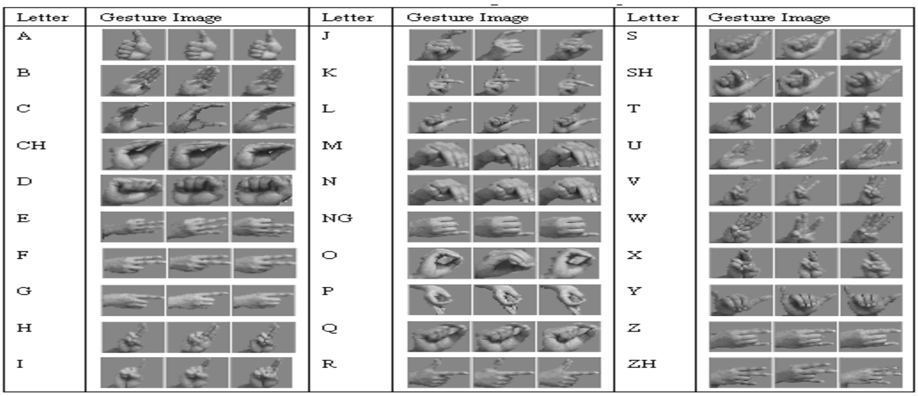

Fig. 1. The Hand Gesture Image and the Represented Letters

Table 1. Experimental results

Algorithms	NN	KNN			Fuzzy KNN			FRNN	FRNNC
Classification accuracy (%)	90.12	K=3	K=5	K=7	K=3	K=5	K=7	96.39	94.96
		90.45	89.45	88.79	90.59	89.92	89.45		

References

1. J.M., Keller, M.R., Gray: A Fuzzy K-nearest neighbor algorithm. IEEE Transactions on System, Man and Cybernetics. 15(4) (1985) 580-585
2. Z., Pawlak: Rough set approach to knowledge-based decision support. European Journal of Operational Research 99 (1997) 48-57
3. D., Dubois, H., Prade: Rough-fuzzy sets and fuzzy-rough sets. International Journal of General Systems. 17(2-3) (1990) 191-209
4. Anna, M.R, Etienne, E.K.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems. 126 (2002) 137-155
5. Sarkar, M.: Fuzzy-Rough nearest neighbors algorithm. In: Proc. Of IEEE Int. Conference on Systems, Man and Cybernetics. Nashville Tennessee USA (2000) 3556-3561
6. C.Z., Ye, J., Yang: Improving performance of decision trees with muti -edit-nearest-neighbor algorithm. Control and Decision. 18(1) (2003) 96-102
7. Yager, R.R, Filev, D.P.: Approximate clustering via the mountain method. IEEE Transactions on Systems, Man and Cybernetics. 24(8) (1994) 1279-1284
8. M.S., Yang, K.L., Wu: A Similarity Based Robust Clustering Method. IEEE Transactions on Pattern Analysis and Machine Intelligence. 26(4) (2004) 434-448
9. Lingras, P., Yan, R.: Interval clustering using fuzzy and rough set theory. In: Processing of NAFIPS '04. Banff Alberta Canada 2 (2004) 780-784