Expert Systems with Applications 41 (2014) 6786–6795

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A design of granular fuzzy classifier

Seek-Beom Roh [a], Witold Pedrycz [b,c], Tae-Chon Ahn [a,*]

[a] Department of Electronics Convergence Engineering, Wonkwang University, 344-2, Shinyong-Dong, Iksan, Jeonbuk 570-749, South Korea
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta T6G 2G7, Canada
[c] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Available online 9 May 2014

Keywords:
Granular fuzzy classifier
Pattern classifier
Information granules
Weighting scheme

ABSTRACT

In this paper, we propose a new design methodology of granular fuzzy classifiers based on a concept of information granularity and information granules. The classifier uses the mechanism of information granulation with the aid of which the entire input space is split into a collection of subspaces. When designing the proposed fuzzy classifier, these information granules are constructed in a way they are made reflective of the geometry of patterns belonging to individual classes. Although the elements involved in the generated information granules (clusters) seem to be homogeneous with respect to the distribution of patterns in the input (feature) space, they still could exhibit a significant level of heterogeneity when it comes to the class distribution within the individual clusters. To build an efficient classifier, we improve the class homogeneity of the originally constructed information granules (by adjusting the prototypes of the clusters) and use a weighting scheme as an aggregation mechanism.

1. Introduction

Classification is a method of supervised learning producing a mapping from a feature space onto classes encountered in the classification problem. Classification problems are encountered in various domains including medicine (Sun, Zhang, & Zhang, 2007), economics (Zhang, Lu, & Li, 2010), and fault detection (Guo, Jack, & Nandi, 2005), etc. In order to improve classification performance, a large number of methods have been developed. There are various important categories of classification techniques including statistical techniques, neural networks, and rule based classification techniques (Pajares, Guijarro, & Ribeiro, 2010).

In statistical techniques, we encounter various techniques such as weighted voting scheme (Jahromi & Taheri, 2008), naive Bayes approach (Tao, Li, Zhu, & Li, 2012), least square and logistic regression (Pendharkar, 2012), and nearest neighbor classification (Chosh, 2012). Most "conventional" statistical classification techniques are based on the Bayesian decision theory where the class label of a given pattern is decided based on the posteriori probability. This aspect of the statistical classification approach results in a certain drawback: if assumptions are not met, the efficiency of these classification techniques could be negatively impacted (Wu, Lin, & Lee, 2011).

There are numerous research activities in neural classification. In light of the existing developments, neural networks form a promising alternative to various conventional classification methods (Wu et al., 2011). Neural networks have been applied to the various fields such as pattern recognition, modeling, and prediction.

Although various types of neural networks classifiers have shown very good classification performance, there are several difficulties when using neural network classifiers.

There are a large number of parameters (connections) to be estimated in neural networks classifiers (Oh, Kim, Pedrycz, & Park, 2011). Neural networks are "black boxes" that lack interpretability (Wu et al., 2011).

Radial basis function neural networks (RBF NNs) came as a sound design alternative when it comes to the reduction of the number of parameters to be adjusted. RBF NNs exhibit some advantages including global optimal approximation and classification capabilities, and a rapid convergence of the learning process (Wu et al., 2011). Although RBF NNs exhibit powerful capabilities with respect to their classification performance and learning speed, they do not offer interpretability aspects. On the other hand, it is well known that fuzzy logic can handle uncertainty and vagueness (Ganji & Abadeh, 2011). Subsequently the use of "if-then" rules improves the interpretability of the results and provides a better insight into the structure of the classifier (Alcalá-Fdez, Alcalá, & Herrera, 2011; Chacon-Murguia, Nevarez-Santana, & Sandoval-Rodriguez, 2009; Ishibuchi & Yamamoto, 2005; Juang & Chen, 2012) and decision making process (Sh, Eberhart, & Chen, 1999).

* Corresponding author. Tel.: +82 63 850 6344; fax: +82 63 853 2196.
E-mail addresses: nado@wku.ac.kr (S.-B. Roh), wpedrycz@ualberta.ca (W. Pedrycz), tcahn@wku.ac.kr (T.-C. Ahn).

http://dx.doi.org/10.1016/j.eswa.2014.04.040
0957-4174/

These observations lead to the emergence of hybrid architectures bringing together learning capabilities of neural networks and interpretability of fuzzy systems giving rise to neurofuzzy architectures (Nandedkar & Biswas, 2009).

In this study, we propose a new design methodology for fuzzy classifiers. The proposed design method is based on information granulation originally introduced by Zadeh (1997). More specifically, information granulation is a process, which decomposes a universe of discourse into some regions of high homogeneity (Liu, Xiong, & Fang, 2011).

Lin studied granular computing and neighborhood systems, mainly focusing on a granular computing model which included the binary relation, the granular structure, the granule's representation, and the applications in granular computing (Lin, 1988,2000,2005a,2005b). Yao introduced rough sets to granular computing, and discussed data mining methods, rule extraction methods and machine learning methods based on granular computing in Yao (2001). Bargiela and Pedrycz (2003) established the fundamentals of Granular Computing. The most recent advancements along with a comprehensive treatment of the subject area are presented in the literature (Pedrycz, 2013).

We develop information granules on a basis of given numeric data (patterns) by exploiting two approaches such as fuzzy clustering and a supervised optimization algorithm. The Fuzzy C-Means (FCM) clustering algorithm is used to form information granules in an unsupervised mode. In a supervised mode, Particle Swarm Optimization (PSO) searches for an optimal distribution of prototypes over the feature space.

After forming the information granules, we determine a distribution of patterns belonging to each class and allocated to a given cluster. This provides information about the heterogeneity of the clusters, which in sequel is used to determine class assignment of a pattern to be classified.

The paper is structured as follows. First, in Section 2, we introduce the new granular classifier. In Section 3, we present experimental results. Conclusions are covered in Section 4.

2. Design of a granular of fuzzy classifier

Information granulation is defined as a process, which partitions a universe into several regions (Song, Yang, Soh, & Wang, 2010). Information granules are built through information granulation for the given patterns expressed in some feature space. Information granulation can be done from different viewpoints.

Information granules being reflective of the geometry of patterns can be realized by running a certain clustering algorithm. In our study, we are concerned with the Fuzzy C-Means (FCM). As noted earlier, the granulation process is realized in unsupervised mode (via clustering) and subsequently the clustering results are improved by adjusting the position of the prototypes already formed by the FCM method.

In this study, we adhere to the standard notation. A finite set of data (patterns) is denoted by $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, where $\mathbf{x}_i \in \Re^n$.

2.1. Information granulation realized in unsupervised mode and its refinement in supervised mode

The given patterns are clustered by the FCM method into "c" clusters. The method generates "c" prototypes $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$ (centers of the clusters) and a partition matrix whose rows are membership functions of successive fuzzy sets.

2.1.1. The supervised mode for information granulation

Clustering algorithm is the assignment of a set of observations into clusters so that data located in the same cluster are similar in a certain sense. The FCM clustering algorithm can be succinctly described as follows. The FCM clustering method creates a collection of information granules in the form of fuzzy sets (Song et al., 2010).

To elaborate on the essence of the method, let us consider a set of patterns $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, $\mathbf{x}_k \in \Re^m$ (where m stands for the dimensionality of the input space).

The objective function used in FCM clustering is defined as follows

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^q \cdot \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad \text{s.t.} \sum_{i=1}^{c} u_{ik} = 1 \tag{1}$$

where u_{ik} is the membership degree of the kth pattern to the ith cluster, $\mathbf{v}_i$ is the center of the ith cluster and c is the number of clusters.

The optimization problem is expressed in the form

$$\min_{\mathbf{U},\mathbf{v}} J \quad \text{subject to} \sum_{i=1}^{c} u_{ik} = 1 \tag{2}$$

The clustering procedure minimizes the objective function (2) through the two update formulas iteratively. The update formulas such as (3) and (4) successively modify the partition matrix and the location of prototypes.

$$R_i(\mathbf{x}_k) = u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|}\right)^{2/(q-1)}} \tag{3}$$

where $R_i(\mathbf{x}_k)$ denotes the membership function of the fuzzy set R_i.

The center of the ith cluster is determined to minimize the objective function (1) as follows.

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N}(u_{ik})^q \cdot \mathbf{x}_k}{\sum_{k=1}^{N}(u_{ik})^q} \tag{4}$$

As an illustration, Fig. 1 shows a location of the prototypes generated by the FCM for a synthetic two-dimensional data.

Fig. 2 shows the activation levels (membership functions) describing fuzzy clusters.

For the ith cluster, we determine the data belonging to this cluster according to the following expression

$$\Omega_i = \{\mathbf{x}_k | R_i(\mathbf{x}_k) = \max_j R_j(\mathbf{x}_k)\}, \quad k = 1, 2, \ldots, N \tag{5}$$

where N is the number of patterns.

The local areas described by (5) are depicted in Fig. 3.

2.1.2. The supervised mode in the refinement of information granulation

In the unsupervised mode, the prototypes of clusters are determined by the clustering method where we investigate the

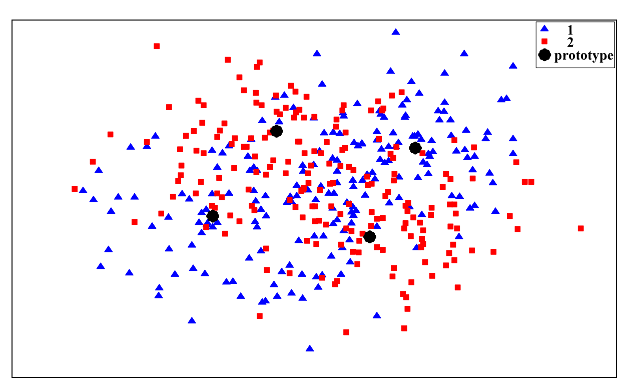

Fig. 1. Example 2-dimensional patterns and their prototypes.