Interpretable Classifier of Diabetes Disease

Nesma Settouti, Meryem Saidi, and Mohamed Amine Chikh

Abstract—**Interpretability represents the most important driving force behind the implementation of fuzzy-based classifiers for medical application problems. The expert should be able to understand the classifier and to evaluate its results. The main purposes in this work is the application of a new method based on FCM and ANFIS to diagnose the diabetes diseases by using a reduced number of fuzzy rules with relatively small number of linguistic labels, removing the similarity of the membership functions, preserving the meaning of the linguistic labels (interpretability), and in same time improving the classification performances. Experimental results show that the proposed approach FCM-ANFIS can get high accuracy with fewer rules. On the contrary, by using ANFIS more rules are needed to get a lower accuracy. Moreover the features projected partition in ANFIS is ambiguous and cannot preserve the meaning of the linguistic labels. The best number of the rules is a trade-off between the accuracy and the rules number, also with a minimum of clusters (c=2) and just two fuzzy rules, FCM-ANFIS approach has given the best results with CC = 83.85%, Se = 82.05% and Sp = 84.62% comparing to the other cases.**

Index Terms—**Interpretable classification; fuzzy rules; FCM; neuro-fuzzy ANFIS; UCI machine learning database.**

I. Introduction

Diabetes is a chronic disease that occurs when the pancreas does not produce enough insulin, or when the body cannot effectively use the insulin it produces. Hyperglycemia, or raised blood sugar, is a common effect of uncontrolled diabetes and over time leads to serious damage to many of the body's systems, especially the nerves and blood vessels. Diabetes is the most rapidly growing chronic disease of our time. It has become an epidemic that affects a large number of people in the world [1].

Diagnose of diabetes for medical expert is a difficult task. For this reason a much research effort has been put till today in diagnosis of diabetes disease literature. Actually fuzzy logic and neural networks have provided attractive structures to complex systems. Adaptive network-based fuzzy inference system (ANFIS) is a specific approach in neuro-fuzzy modeling which utilizes the neural networks to tune the rule-based fuzzy systems [2]. Successful applications of ANFIS in biomedical engineering have been reported recently in data identification [3], [4], and pattern recognition [5].

The exponential increasing number of fuzzy rules in neuro-fuzzy classifier is one of the difficult problems to be overcome in this paper, such problem is solved by adopting the FCM clustering method [6] for the structure identification in the ANFIS, wish is one of the most widely used neuro-fuzzy models proposed by Jang [2], against diabetes diagnosis problem. The algorithm fuzzy c-means (FCM) is a typical clustering algorithm, which was used in a wide variety of engineering and scientific disciplines such as modeling [7] the decision [8], pattern recognition and classification [9], segmentation [10]. Recently the classification with the approach FCM-ANFIS [11] was applied on database of diabetes collected by the Faculty of Computer Science and Information, University of Technology Malaysia; the results have reached 72.66%.

The remainder of the paper is organized as follows, the Adaptive Neuro-Fuzzy Inference System (ANFIS) is proposed based on the model of Takagi-Sugeno. Pima Indian Database and neuro-fuzzy Classifier learning are presented in Section 3. In section 4, the results are presented and discussed. Finally, section 5 concludes the findings.

Manuscript received March 10, 2012; revised May 5, 2012.

Nesma Settouti is with the Biomedical Engineering Laboratory, Tlemcen University –Algeria (Tel: +213 055 610 476 39; fax: +213 043285686; e-mail address: nesma.settouti@gmail.com).

Meryem Saidi is with the Biomedical Engineering Laboratory, Tlemcen University – Algeria.

Mohamed Amine Chikh is with the Tlemcen University-Algeria. Actually he is the head of CREDOM research team at Biomedical Engineering Laboratory.

II. Theory

A. Fuzzy C-Means Algorithm

The fuzzy c-means (FCM) clustering algorithm was first introduced by Dunn [12] and later extended by Bezdek [6]. It is based on the concept of fuzzy c-partition, introduced by Ruspini [13]. FCM is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 , 1 \le m < \infty \quad (1)$$

where

- m is any real number greater than 1, it was set to 2.00 by Bezdek.
- u_{ij} is the degree of membership of x_i in the cluster j;
- x_i is the ith of d-dimensional measured data;
- c_j is the d-dimension center of the cluster,
- ||*|| is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the c_j cluster centers by: