Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 1337–1354
Expert Systems with Applications
www.elsevier.com/locate/eswa

Increasing accuracy of two-class pattern recognition with enhanced fuzzy functions

Asli Celikyilmaz [a,*], I. Burhan Türkşen [a,b], Ramazan Aktaş [b], M. Mete Doganay [c], N. Basak Ceylan [d]

[a] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ont., Canada M5S3G8
[b] TOBB Economics and Technology University, Sogutozu Caddesi No: 43 Ankara, Turkey
[c] Department of Business Administration, Cankaya University, Turkey
[d] Department of Business Administration, Atilim University, Turkey

Abstract

In building an approximate fuzzy classifier system, significant effort is laid on estimation and fine-tuning of fuzzy sets. However, in such systems little thought is given to the way in which membership functions are combined within fuzzy rules. In this paper, a robust method, improved fuzzy classifier functions (IFCF) design is proposed for two-class pattern recognition problems. A supervised hybrid improved fuzzy clustering for classification (IFC-C) algorithm is implemented for structure identification. IFC-C algorithm is based on a dual optimization method, which yields simultaneous estimates of the parameters of *c*-classification functions together with fuzzy *c* partitioning of dataset based on a distance measure. The merit of novel IFCF is that the information on natural grouping of data samples i.e., the membership values, are utilized as additional predictors of each fuzzy classifier function to improve accuracy of system model. Improved fuzzy classifier functions are approximated using statistical and soft computing approaches. A new semi-non-parametric inference mechanism is implemented for reasoning. The experimental results of the new modeling approach indicate that the new IFCF is a promising method for two-class pattern recognition problems.

Keywords: Fuzzy classification; Improved fuzzy clustering; Fuzzy functions; Data mining; Early warning system; Decision support systems

1. Introduction

Fuzzy pattern recognition is often identified with fuzzy clustering or with fuzzy *IF...THEN* rule base systems that are used as classifiers, which yield a class label (crisp or soft) for each given pattern. This paper adopts a similar approach, but replaces the fuzzy rule base systems with novel improved fuzzy classifier functions.

Several fuzzy system modeling approaches have been proposed to understand the system behavior and applied on different problem domains of information processing. In traditional methods of Fuzzy System Models (FSMs) expert knowledge has been used to define the linguistic properties of given system. However, these methods have issues of subjectivity and lack of generalization. In order to overcome some of the latter issues, modeling strategies (Kilic, Sproule, Türkşen, & Naranjo, 2002; Kim, Park, Kim, & Park, 1998; Nakanishi, Türkşen, & Sugeno, 1993; Sugeno & Yasukawa, 1993; Uncu & Türkşen, 2004) that are more objective are developed and expert intervention is minimized. In these approaches, fuzzy sets are learnt from a given dataset through optimization algorithms, e.g., fuzzy clustering, and they are based on either projection of the output clusters onto input spaces or projection of input–output clusters onto input and output spaces separately. Additionally, some self-tuning hybrid algorithms have been proposed such as fuzzy-neural networks (Ishibuchi, Murata, & Türkşen, 1997; Wang & Lee, 2002), or genetic algorithms (Chien, Lin, & Hong,

* Corresponding author. Tel.: +1 416 946 75 81.
E-mail address: asli@mie.utoronto.ca (A. Celikyilmaz).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.039

2002), for optimization of the parameters of the membership functions and fuzzy rules.

FSMs based on fuzzy rule bases (FRB) (if-then rules) have some challenges (Türkşen & Celikyılmaz, 2006) with varying degree of uncertainties; identification of membership functions, most suitable combination operator and operations, e.g., t-norm or t-conorm, conjunction, defuzzification, and implication operators to capture the uncertainty associated with the linguistic "AND", "OR", and "IMP" for the representation of rules, as well as reasoning with them. Hence, to overcome these challenges, in this work we develop an improved fuzzy classifier functions (IFCF) approach to estimate decision boundaries, which does not require construction of if-then rules. The proposed approach, based on the foundation of our previous novel "Fuzzy Functions" approaches for regression problems (Celikyilmaz et al., 2007; Türkşen & Celikyılmaz, 2006; Türkşen, 2007), is an alternative system modeling methodology to Fuzzy Classifiers with FRBs (Castellano, Fanelli, & Mencor, 2004; Chen & Chen, 2006; Ishibuchi et al., 1997; Setnes & Babuska, 1999) etc.

The new IFCF implements an updated form of improved fuzzy clustering (IFC) algorithm, namely, "*improved fuzzy clustering for classification*" (IFC-C), specific to two-class pattern recognition problems, to cluster data into several overlapping clusters, each of which is used to define a separate decision surface. Earlier, IFC (Celikyilmaz et al., 2007) algorithm was proposed to find improved membership values for regression problems. IFC-C elicits membership values with a new membership value calculation equation through combination of fuzzy c-means clustering algorithm and a classification algorithm within one clustering schema and forms a dual structure-clustering algorithm. Herein, the objective of IFC-C algorithm is to identify the hidden patterns, as well as to find "*improved*" membership values, which are good predictors of fuzzy classifier functions of each cluster. One of the merits of our approach, being robust and constructing high performance fuzzy classifier models aside, is that during structure identification, similarity of the objects is enhanced with additional fuzzy identifiers, i.e., the membership values, by utilizing them as additional input variables. With this approach, the classifier employs valuable information, which in turn "*improves*" accuracy of the whole classifier model.

For the last few years, there have been increasing interests in fuzzy logic systems to improve the performance of fuzzy learning algorithms by incorporating tools well-known from statistical learning theory. For instance, new approaches to SVM algorithm in Hong and Hwang (2003), Wang, Wang, and Lai (2005) incorporate membership values as additional weight parameters. In Leski (2006) fuzzy logic is incorporated into SVM in terms of identifying a linguistic kernel matrix. Herein, motivated by the above-mentioned methods, implementation of SVM to approximate fuzzy classification function parameters of each granule (cluster) is the interest of this work.

This paper is organized as follows: in Section 2, we will review two classification algorithms. They will be later used as parts of learning process of proposed fuzzy classifier function approach. In Section 3, we present hybrid structure IFC-C for pattern recognition domains, the extended version of earlier IFC method (Celikyilmaz et al., 2007). In Section 4, we present learning and inference mechanisms that will be used for reasoning. In Section 5, we present simulation results of models using benchmark datasets and a discussion of fuzzy modeling of an early failure warning system for selected Turkish banks, and finally, conclusions will be drawn in Section 6.

2. Classification algorithms

Proposed improved fuzzy classification design of this work uses posterior probabilities of training data vectors for reasoning. Among many techniques, we have chosen to apply a rather non-complex classifier and a sophisticated non-linear classifier to approximate classifiers and to estimate posterior probabilities at granular level. Hence, in this section we will shortly review these well-known classifiers and calculation of posterior probabilities of given instances with them.

2.1. Posterior probabilities from logistic regression (LR)

Given training dataset of nd data points, $\Omega = \{(x_k, y_k), k = 1, \ldots, nd\}$, where x_k represents input data vectors of nv features, $x_k = (x_{k,1}, \ldots, x_{k,nv})^t \in X \subseteq \Re^{nv}$, and class labels $y_k \in \{0,1\} \in Y$, logistic regression function tries to estimate the probability $P(y=1|x)$ of $y=1$ with the following formula:

$$\widehat{P}(y_k = 1 \mid x_k) = 1/(1 + \exp(-(w_0 + w^T x_k))), \tag{1}$$

where w is the parameter vector, w_0 refers to the intercept, and $x_k \in \Re^N$ is nv dimensional input data vector, $P(y=0|x) = 1 - P(y=1|x)$. Parameters, w and w_0 are estimated using maximum likelihood formula.

2.2. Posterior probabilities with support vector machines (SVM) for binary classification problems

SVM (Vapnik, 1998) is originally designed to estimate a classifier function, $f{:}\Re^{nv} \to \{0,1\}$ by maximizing the margin between the nearest samples of opposite classes. Let x' be a test sample. The following decision surface:

$$\begin{aligned} \hat{y}' &= sign(f(x')) = sign(w^T x' + b) \\ &= sign\left(\sum_{k}^{nd} \beta_k y_k K(x_k, x') + b\right) \end{aligned} \tag{2}$$

is used to assign x' a class label depending on the sign of $f(x')$. So, when $f(x') > 0$, x' gets a positive label, and vice versa. $K(.)$ is the kernel function used to map original data vector onto a higher dimensional space, either linearly or non-linearly. In this paper, we analyzed three different kernel functions: linear: $K(x_k, x_j) = x_k^T x_j$, polynomial: