Neurocomputing 000 (2017) 1–9

Contents lists available at ScienceDirect

Neurocomputing

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

Multivariate soft repulsive system identification for constructing rule-based classification systems: Application to trauma clinical data

Abolfazl Doostparast Torshizi[a,*], Linda Petzold[a], Mitchell Cohen[b]

[a] Department of Computer Science, University of California, Santa Barbara, USA
[b] Department of Surgery, University of California, San Francisco, USA

ARTICLE INFO

Article history:
Received 12 December 2016
Revised 5 February 2017
Accepted 12 March 2017
Available online xxx

Communicated by Zhaohong Deng

Keywords:
Fuzzy systems
System identification
Multivariate analysis
Clustering

ABSTRACT

Rule-based classification systems constructed upon linguistic terms in the antecedent and consequent of the rules lack sufficient generalization capabilities. This paper proposes a new multivariate fuzzy system identification algorithm to design binary rule-based classification structures through making use of the repulsive forces between the cluster prototypes of different class labels. This approach is coupled with the potential discrimination power of each dimension in the feature space to increase the generalization potential. To address this issue, first the multivariate variant of a newly proposed soft clustering algorithm along with its mathematical foundations is proposed. Next, the discriminatory power of each individual feature is computed, using the multivariate membership values in the proposed clustering algorithm to achieve the most accurate firing degree in each rule. The main advantage of this method is to handle unbalanced datasets yielding superior true positive measure while keeping the false positive rate low enough to avoid the natural bias toward class labels containing larger number of training samples. To validate the proposed approaches, a series of numerical experiments on publicly available datasets and a real clinical dataset collected by our team were conducted. Simulation results demonstrated achievement of the primary goals of this research.

1. Introduction

Pattern recognition spans a wide range of applications from personalized medicine to engineering. In general, the primary objective of pattern recognition is to learn the dominating structure of the data and then to assign unseen samples into predetermined categories. In the realm of pattern recognition, a large body of classification methods can be found in the literature [1–3]. Some of these methods are quite well-known such as Support Vector Machines (SVM), Artificial Neural Networks (ANNs), AdaBoost [4], and Kernel Fisher Discriminant method (KFD) [5]. One of the major difficulties in using such state-of-the-art algorithms is that they are computationally complex while having poor interpretability. In recent years, many efforts have been made to overcome these issues through making radical changes in the existing algorithms. Some of these methods include: core vector machines [6,7], Lagrangian Support Vector Machines (LSVM) [8], and least square SVM [9]. One of the solutions to overcome such drawbacks has been to use fuzzy classifiers whose interpretability is promising, even though they often yield less accurate classification performance.

In essence, generalization and interpretability play a central role to call a classifier good. The ability of a classifier to correctly categorize unseen samples is referred to as generalization. On the other hand, system abstraction delineating underlying structure of the classifier in an understandable way is called interpretability. There has been an argument in the scientific community to establish a balance between these two concepts. According to statistical learning theory [10], too accurate classifiers on training data often produce poor results on validation data. This phenomenon is called overfitting. Decreasing the level of overfitting necessitates sacrificing accuracy, to achieve a reasonable balance between generalization and interpretability. Fuzzy classification systems provide a promising balance between these two alternatives [11].

Fuzzy classification systems capture the non-linear boundaries between class labels in the frame of IF-THEN rules. A plethora of rule-based classification systems trying to establish a balance between generalization and interpretability can be found [12–17]. Basically, fuzzy rule-based systems map to two distinct categories including: (1) the Mamdani–Assilan (MA) methods where both antecedent and consequent parts of the rules are formed by linguistic variables, and (2) the Takagi–Sugeno–Kang (TSK) methods where the antecedent part of the rule consists of linguistic

* Corresponding author.
E-mail addresses: ad3436@cumc.columbia.edu, a.doostparast@gmail.com (A.D. Torshizi).

http://dx.doi.org/10.1016/j.neucom.2017.03.043

variables but the consequent part is a function of the input variables. These learning philosophies possess distinctive pros and cons in the sense that TSK-based approaches provide good generalization, while MA approaches are more interpretable. In order to form the antecedent part of the rule-base in fuzzy classification algorithms, researchers usually make use of clustering tools [15,18–21]. Such a procedure yields descent classification performance when data vectors are non-overlapping. Nevertheless, they fail to produce acceptable results when dealing with overlapping classes.

The trade-off between generalization and interpretability has also been addressed from another perspective. This issue has been tackled by using multi-objective genetic algorithms and other similar approaches to shape the fuzzy rule-based classification systems. In an effort for tuning fuzzy rules in a fuzzy classifier, Murcia et al [22] have tried the parameter tuning based on adaptive memetic algorithm resulting in successful classification performance compared to genetic algorithm tuning procedure. Application of evolutionary search algorithms has been successfully attempted in the soft computing community for better trade-off between generalization and interpretability. Although the literature in this regard is quite rich, some of the most appealing approaches can be listed including: high-dimensional classification problems using genetic rule selection and lateral tuning [17], handling datasets containing large numbers of instances for classification purposes using multi-objective co-evolutionary genetic training approach [23], hybrid ant-bee colony optimization for parameter tuning in fuzzy systems [24], rule-base generation using genetic algorithms applied to image classification [25,26], steady-state genetic rule-extraction [27], an automatic process for developing hierarchical Takagi–Sugeno (TF) fuzzy classification systems using probabilistic program evolution (PIPE) and evolutionary programming (EP) [28]. Some other fuzzy classification approaches mostly related to TSK and other non-fuzzy learning methods such as SVM can be found in the literature [29–32]. One of the recent studies based on the TSK structure proposes a multiview classification algorithm [32] that employs collaborating learning mechanism to extend the currently existing TSK systems to handle two-view datasets. Employing support vector machines in parallel with utilizing fuzzy clustering to generate the rule base of a TSK system was considered in [31]. In this work, the authors constructed the initial structure of the rule-base using fuzzy clustering and then learned the system parameters using SMV in order to achieve higher generalization ability.

Several research studies have dealt with the clustering of data by using overlapping clusters. In another study [33], unsupervised binary labeling of two datasets based on the sign of the difference between probability densities was introduced. Also, Kullback–Leibler (KL) [34] divergence was calculated using the kernel density estimator and then used for measuring the similarity between uncertain objects. The main disadvantage of all these approaches is that the resultant clusters cannot be described with Gaussian membership functions having different dispersions. To overcome this issue, Leski [11] proposed a new conditional clustering method called Fuzzy (c+p)-means (FCPM) for constructing the antecedent part of each rule in the rule-base of the classification systems. In contrast to the other clustering-based system identification algorithms where data vectors belonging to each class are clustered separately, the FCPM takes into account the repulsive forces between the cluster centers of different class labels. Given this characteristic, IF-THEN rules belonging to each class are influenced by the prototypes of the other classes, leading to an increase in generalization and interpretability of the system when dealing with data having overlapping classes. Consider a binary classification problem. If we have c prototypes in class 1 and p prototypes in class 2, then the optimization model of the FCPM clustering algorithm will be as follows [11]:

$$J(\mathbf{U},\mathbf{T},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^m\|\mathbf{x}_k-\mathbf{v}_i\|^2 + \sum_{j=1}^{p}\sum_{k=1}^{N}(t_{jk})^m\left\|\mathbf{x}_k-\mathbf{z}_j\right\|^2$$
$$S.T. \qquad (1)$$
$$\underset{1\le k\le N}{\forall} \quad \sum_{i=1}^{c}u_{ik}+\sum_{j=1}^{p}t_{jk}=1$$

where $\mathbf{V}=[\mathbf{v}_1,\ \mathbf{v}_2,\dots,\mathbf{v}_c]\in\Re^{n\times c}$ $(\mathbf{Z}=[\mathbf{z}_1,\ \mathbf{z}_2,\dots,\mathbf{z}_p]\in\Re^{n\times p})$ represents the matrix of cluster prototypes of class 1 (2). u_{ik} (t_{jk}) denotes the membership degree of the data point $\mathbf{x}_k$ to cluster prototype $\mathbf{v}_i$ $(\mathbf{z}_j)$.

One of the main disadvantages of most fuzzy clustering algorithms for system identification purposes is that they work in a univariate environment, so that the effects of each individual feature cannot be measured for further analysis. Also of importance is that more informative features capable of having a large impact on the classification output of each sample cannot be measured. In this regards, Pimentel and De Souza [35] proposed the first multivariate fuzzy c-means (FCM) clustering algorithm capable of multivariate fuzzy partitioning of the dataset Ω by minimizing the following objective function:

$$J^M=\sum_{i=1}^{c}\sum_{k=1}^{n}\sum_{j=1}^{p}(u_{ijk})^m d_{ijk} \qquad (2)$$

where u_{ijk} (d_{ijk}) denotes the membership degree (distance) of the data point $\mathbf{x}_k$ on dimension j to the cluster center $\mathbf{v}_i$. d_{ijk} can be computed by (3):

$$d_{ijk}=(x_{jk}-y_{ij})^2 \qquad (3)$$

where y_{ij} represents the value of the prototype c_i on dimension j.

In the next step, in another research [36], the authors provide a weighted version of the multivariate FCM clustering algorithm for interval-valued scientific production data. A modified version of the weighted multivariate FCM clustering algorithm has been proposed [37], where more informative features gain larger weights and play a more important role in the final shape of the detected clusters. We should note that none of these multivariate clustering algorithms take into account the repulsive forces between the overlapping clusters belonging to different class labels. Then, the contributions of this paper are as follows:

1. We develop an improved multivariate version of the FCPM clustering algorithm to take into account the effects of each individual feature on the shape of the final membership functions in the antecedent part of IF-THEN rules, to account for the repulsive forces between the prototypes of overlapping clusters.
2. We introduce importance coefficients of each individual feature based on the multivariate membership values of each data point belonging to each class label.
3. We perform extensive computational simulations on publicly available data and real-world clinical data being collected by our team.

The remainder of this paper is organized as follows: The developed clustering method is discussed in the next section. Then, the proposed system identification approach is presented along with the experimental simulations and discussions. Finally, concluding remarks are given.

2. Multivariate fuzzy (c+p)-means clustering algorithm (MFCPM)

Fuzzy clustering algorithms attempt to partition the data into several clusters, where members of the same cluster share similarities while being dissimilar enough from the members of the other clusters. The most well-known fuzzy clustering algorithm is FCM,

Please cite this article as: A.D. Torshizi et al., Multivariate soft repulsive system identification for constructing rule-based classification systems: Application to trauma clinical data, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.03.043

first introduced by Bezdek [38]. FCM assigns each data point to a predefined number of clusters, where each point belongs to different clusters with a different degree of membership. FCM has been extensively studied and numerous variants of it can be observed in the literature. FCM is usually used in the process of identifying architecture of fuzzy classification rule-based systems to extract the formation of the membership functions in the antecedent part of the IF-THEN rules. Basically, in order to find the membership functions of each class label, the training samples belonging to each class are first clustered and the rules are extracted on their basis. This process continues for other data vectors belonging to other classes. The main flaw of this process is that it ignores the interactions among the cluster prototypes of different class labels. In this situation, if the data points from different classes overlap, then the cluster prototypes of the different classes may end up to be unreasonably close, making some of the extracted rules practically redundant.

Based on the above observations, we propose a multivariate version of FCPM capable of attaining similar cluster prototypes while computing the membership of each data vector for individual dimensions of the feature space. These values are then used for tuning the parameters of the rule-base of the classification system.

Consider the following constrained optimization problem:

$$J(\mathbf{U},\mathbf{T},\mathbf{V}) = \sum_{i=1}^{C}\sum_{j=1}^{p}\sum_{k=1}^{N}(u_{ijk})^m d_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}\sum_{k=1}^{N}(t_{ljk})^m d_{ljk}$$
$$S.T. \tag{4}$$
$$\forall_{1\le k\le N} \quad \sum_{i=1}^{C}\sum_{j=1}^{p}u_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}t_{ljk} = 1$$

where C, H, p, and N represent the number of clusters in class 1, the number of clusters in class 2, the number of features (dimension), and the total number of data points for training the system, respectively. In addition, $m$denotes the fuzzifier value to determine the softness of the cluster boundaries, and $u_{ijk}$and t_{ljk} represent the multivariate membership degree of the point $\mathbf{x}_k$ on the dimension j to the cluster centers $\mathbf{v}_i$ and $\mathbf{z}_l$, respectively. Also d_{ijk} and d_{ljk} represent the distance between the data point $\mathbf{x}_k$ and prototypes of the clusters i and l belonging to classes 1 and 2 in the dimension$j$, respectively. According to the above constraint, the total membership degree of each data point to the prototypes of each class on each dimension must equal 1.

Proposition 1. The multivariate membership degree of the data point $\mathbf{x}_k$ to the ith cluster of the class 1 and the lth cluster of the class 2 on dimension j is as follows:

$$u_{ijk} = \frac{\left(\frac{1}{d_{ijk}}\right)^{\frac{1}{m-1}}}{\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}}} \tag{5}$$

$$t_{ljk} = \frac{\left(\frac{1}{d_{ljk}}\right)^{\frac{1}{m-1}}}{\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}}} \tag{6}$$

Proof. First, the optimization model (4) is unconstrained using the Lagrangian multiplier method, yielding

$$\min\ L\left[J(U,T,V)\right] = \sum_{i=1}^{C}\sum_{j=1}^{p}\sum_{k=1}^{N}\left(u_{ijk}\right)^m d_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}\sum_{k=1}^{N}\left(t_{ljk}\right)^m d_{ljk} - \lambda_k\left(\sum_{i=1}^{C}\sum_{j=1}^{p}u_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}t_{ljk} - 1\right) \tag{7}$$

Taking the partial derivative of $L[J(U, T, V)]$ with respect to the unknown membership variables and setting it to zero, we obtain

$$\forall_{1\le k\le N} \quad \frac{\partial L}{\partial u_{ijk}} = 0,$$
$$\Rightarrow\ m\left(u_{ijk}\right)^{m-1}d_{ijk} - \lambda = 0\ \Rightarrow\ u_{ijk} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left(\frac{1}{d_{ijk}}\right)^{\frac{1}{m-1}} \tag{8}$$

$$\forall_{1\le k\le N} \quad \frac{\partial L}{\partial t_{ljk}} = 0,$$
$$\Rightarrow\ m\left(t_{ljk}\right)^{m-1}d_{ljk} - \lambda = 0 \Rightarrow\ t_{ljk} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\left(\frac{1}{d_{ljk}}\right)^{\frac{1}{m-1}}. \tag{9}$$

Taking the partial derivative of the Lagrangian relaxed objective function, i.e., (7) with respect to the Lagrange multiplier, we have

$$\forall_{1\le k\le N} \quad \frac{\partial L}{\partial \lambda_k} = 0, \Rightarrow\ \sum_{a=1}^{C}\sum_{b=1}^{p}u_{ijk} + \sum_{d=1}^{H}\sum_{b=1}^{p}t_{ljk} - 1 = 0. \tag{10}$$

Substituting (8) and (9) in (10) gives

$$\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}}\sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}} = 1$$
$$\Rightarrow\ \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}}}. \tag{11}$$

Substituting (11) in (8) and (9), we obtain the exact values of the membership degrees for each data point to each cluster belonging to different classes on each dimension.

$$u_{ijk} = \frac{\left(\frac{1}{d_{ijk}}\right)^{\frac{1}{m-1}}}{\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}}} \tag{12}$$

$$t_{ljk} = \frac{\left(\frac{1}{d_{ljk}}\right)^{\frac{1}{m-1}}}{\sum_{a=1}^{C}\sum_{b=1}^{p}\left(\frac{1}{d_{abk}}\right)^{\frac{1}{m-1}} + \sum_{d=1}^{H}\sum_{b=1}^{p}\left(\frac{1}{d_{dbk}}\right)^{\frac{1}{m-1}}}. \tag{13}$$

□

Another major conclusion from (7) is how to compute the cluster prototypes of each cluster on each specific dimension. Here, we can show that the cluster center of the cluster i on the dimension j can be obtained as follows:

Proposition 2. Suppose the cluster center i of the class label 1 on dimension j to be y_{ij}. It can be computed directly by the multivariate membership values using the following formulation

$$y_{ij} = \frac{\sum_{k=1}^{N}\left(u_{ijk}\right)^m x_{jk}}{\sum_{k=1}^{N}\left(u_{ijk}\right)^m}. \tag{14}$$

Proof 2. Let the unconstrained formulation of the optimization model (4) be as follows:

$$\min\ L\left[J(U,T,V)\right] = \sum_{i=1}^{C}\sum_{j=1}^{p}\sum_{k=1}^{N}\left(u_{ijk}\right)^m d_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}\sum_{k=1}^{N}\left(t_{ljk}\right)^m d_{ljk} - \lambda_k\left(\sum_{i=1}^{C}\sum_{j=1}^{p}u_{ijk} + \sum_{l=1}^{H}\sum_{j=1}^{p}t_{ljk} - 1\right). \tag{15}$$

Suppose d_{ijk} and d_{ljk} are defined as squared Euclidean distances

$$d_{ijk} = \left(x_{jk} - y_{ij}\right)^2. \tag{16}$$

Substituting (16) in (15) and taking the partial derivative of (15) with respect to y_{ij} leads to

$$\frac{\partial L}{\partial y_{ij}} = 0 \Rightarrow -2\sum_{k=1}^{N}(u_{ijk})^m(x_{jk}-y_{ij}) = 0$$

$$\Rightarrow \; y_{ij} = \frac{\sum_{k=1}^{N}(u_{ijk})^m x_{jk}}{\sum_{k=1}^{N}(u_{ijk})^m}. \tag{17}$$

□.

Proposition 3. Suppose the cluster center l of the class label 2 on dimension j to be z_{lj}. Then it can be computed directly by the multivariate membership values, from

$$z_{lj} = \frac{\sum_{k=1}^{N}\left(t_{ljk}\right)^m x_{jk}}{\sum_{k=1}^{N}\left(t_{ljk}\right)^m}. \tag{18}$$

Proof. By considering the following distance measure, the proof is similar to the proof of the Proposition 2.

$$d_{ljk} = \left(x_{jk} - z_{lj}\right)^2. \tag{19}$$

□.

Based on the propositions and formulations provided in this section, we present the MFCPM algorithm.

```
Algorithm 1:
```

Step 1: Fix the number of clusters for both class labels C and H. Set the fuzzifier coefficient $m \in (1, \infty)$ and set iteration the index to$Iter = 1$. Provide the initial membership partitioning matrices U and T.

Step 2: Initialize the cluster prototypes for each class label using (14) and (18). The prototypes of class labels 1 and 2 can be represented by $\mathbf{V}^{(1)}_{Iter} = [\mathbf{v}^{(1)}_1, \mathbf{v}^{(1)}_2, \ldots, \mathbf{v}^{(1)}_C]$ and $\mathbf{V}^{(2)}_{Iter} = [\mathbf{v}^{(2)}_1, \mathbf{v}^{(2)}_2, \ldots, \mathbf{v}^{(2)}_H]$, respectively. The superscripts (1) and (2) denote the class labels 1 and 2, respectively.

Step 3: Update the multivariate partition matrices using (12) and (13), considering the cluster prototypes calculated in Step 2.

Step 4: Update the cluster prototypes of both cluster labels $\mathbf{V}^{(1)}_{Iter+1}$ and $\mathbf{V}^{(2)}_{Iter+1}$.

Step 5: Check the termination criterion. If $\|U^{Iter} - U^{Iter+1}\| > \varepsilon$ let $Iter = Iter + 1$and go to Step 3; otherwise stop.

Remarks. The termination criterion in this algorithm depends on the Frobenius norm of two successive multivariate partition matrices. In this paper we have set the termination threshold $\varepsilon = 10^{-4}$. In this improved algorithm, the fuzzifier coefficient $m = 2$ has been considered.

In the next section, we provide the details on how the MFCPM algorithm can be used to construct the IF-THEN part of the rules in the rule-base, and the firing process of each rule will be described.

3. Application of the MFCPM clustering algorithm in designing rule-based classifiers

This section is two-fold. In the first subsection, the general framework of constructing the antecedent part of the rule-base using the proposed multivariate clustering algorithm is discussed. Next, the multivariate weighting procedure for estimating the importance weight of each feature in each rule is described.

4. Constructing the antecedent part in the rule-base of the classification system

In this section, the rule-based identification procedure is illustrated. Note that the proposed algorithm is merely designed for binary classification problems. Let the training set be $T = \{T^{(1)}, T^{(2)}\}$, where $T^{(1)}$ and $T^{(2)}$ represent the training vectors belonging to the class labels $\theta_1 = +1$ and $\theta_2 = -1$, respectively. The total number of training samples is considered to be $|T| = N$ where $|\,|$ denotes the cardinality symbol. Also $|T^{(1)}| = N_1$and $|T^{(2)}| = N_2$.

In order to construct the elements of the IF-THEN part, we can use the proposed multivariate clustering algorithm. Samples from each class label are clustered considering the prototypes of the other class labels to take into account the repulsive forces between their respective cluster centers. It is possible to consider different numbers of cluster prototypes for different classes, but this will lead to a kind of bias towards the classes having larger numbers of cluster prototypes. Therefore, for each class the number of clusters will equal c.

After clustering the samples from each class, membership functions for each rule on each feature will be determined in the form of a Gaussian membership function. Each Gaussian membership function consists of a center $v^{(j)}_{if}$ and a dispersion value $s^{(j)}_{if}$. $v^{(j)}_{if}$ ($s^{(j)}_{if}$) represents the cluster center (dispersion) value of the cluster $i = \{1, \ldots, c\}$ in the class $j = \{+1, -1\}$on the feature f. The multivariate membership values computed during the clustering process are directly used to obtain the dispersion of each feature in each rule.

Let $u^{(j)}_{ifk}$ be the multivariate membership value of the data point $\mathbf{x}_k$ in the rule (cluster) i in the class j on dimension f. The multivariate dispersion of each Gaussian membership function in the rule-base can be calculated as follows:

$$\underset{\substack{1\le i\le c\\ j\in\{1,2\}\\ 1\le f\le p}}{\forall} \quad \left[s^{(j)}_{if}\right]^2 = \frac{\sum_{k\in T^{(j)}} u^{(j)}_{ifk}\left[x_{fk} - v^{(j)}_{if}\right]}{\sum_{k\in T^{(j)}} u^{(j)}_{ifk}}. \tag{20}$$

In order to compute the firing strength of each rule, we make use of the product t-norm. Let $A^{(j)}_{if}$ be the Gaussian fuzzy membership value on the axis f of the rule i belonging to the class label j. Then the firing strength of the rule $A^{(j)}_i$ when inserting a new unseen sample to the system can be obtained by

$$A^{(j)}_i = \prod_{f=1}^{p} A^{(j)}_{if}. \tag{21}$$

Since each dimension of the feature space is represented by a Gaussian membership function, the firing strength of each rule will be a multiplication of several Gaussian functions over the entire feature space. As a result, it can be directly computed using

$$\underset{\substack{1\le i\le c\\ j\in\{1,2\}}}{\forall} \quad A^{(j)}_i(\mathbf{x}) = \exp\left[-0.5\sum_{f=1}^{p}\left(\frac{x_f - v^{(j)}_{if}}{\rho s^{(j)}_{if}}\right)\right], \tag{22}$$

where ρ plays the role of a scale parameter for estimating the dispersion values. Although estimation of the dispersion value for normally distributed patterns is unbiased, the distribution of the patterns are usually unknown. To tackle such an issue, we can use the scale parameter for unbiasing the learning process. In the following, the feature weighting procedure will be discussed and at the end, the multivariate system identification procedure will be represented.

5. Constructing the antecedent part in the rule-base of the classification system

As mentioned before, one of the goals of this study is to incorporate the individual weight components of each dimension in the feature space to gain a more meaningful rule-base structure aiming at increasing the interpretability of the system while keeping or improving the generalization potentials of the developed system. In order to compute these weighting elements, we have made use of the computed multivariate membership degrees to estimate the effectiveness of each feature.

Let f be the multivariate membership degree of the data point $\mathbf{x}_k$ on dimension j in the rule (cluster) i. The importance weight is given by

$$w_j^{(1)} = \frac{\sum_{i=1}^{C}\sum_{k=1}^{N} u_{ijk}}{\sum_{i=1}^{C}\sum_{k=1}^{N} u_{ijk} + \sum_{l=1}^{H}\sum_{k=1}^{N} t_{ljk}}. \quad (23)$$

where $w_j^{(1)}$ denotes the weight factor of dimension j in the rules belonging to class 1. C and H represent the number of clusters in the classes 1 and 2, respectively. u_{ijk} (t_{ljk}) denotes the membership degree of the pattern $\mathbf{x}_k$ to the cluster i (l) on the dimension $j$in class 1 (2). The same procedure can be applied to class 2.

$$w_j^{(2)} = \frac{\sum_{l=1}^{H}\sum_{k=1}^{N} t_{ljk}}{\sum_{i=1}^{C}\sum_{k=1}^{N} u_{ijk} + \sum_{l=1}^{H}\sum_{k=1}^{N} t_{ljk}} \quad (24)$$

In the following we will present two algorithms. The first one is the based on the MFCPM clustering and the second one is similar but incorporates the weight factors introduced in (23) and (24). When considering importance weights for each dimension of the feature space, the formulation to compute the firing degree is as follows:

$$A_i^{(j)} = \prod_{f=1}^{p} w_f^{(j)} A_{if}^{(j)} \quad (25)$$

Algorithm 2.

Step 1: Set the parameters C and N using the method proposed in the reference [39].
Step 2: Compute $\mathbf{V}_1^{(1)}$ and $\mathbf{V}_1^{(2)}$ in each class separately using a univariate clustering algorithm.
Step 3: Set iteration index $z = 1$.
Step 4: Update the cluster prototypes $\mathbf{V}_{z+1}^{(1)}$ by `Algorithm 1` using $\mathbf{V}_z^{(1)}$ and $\mathbf{V}_z^{(2)}$.
Step 5: Update the cluster prototypes $\mathbf{V}_{z+1}^{(2)}$ by `Algorithm 1` using $\mathbf{V}_z^{(1)}$ and $\mathbf{V}_z^{(2)}$.
Step 6: Update the multivariate membership partitioning matrices for both classes using `Algorithm 1`, based on $\mathbf{V}_z^{(1)}$ and $\mathbf{V}_z^{(2)}$.
Step 7: Check the termination criterion, $|\mathbf{V}_z^{(1)} - \mathbf{V}_{z+1}^{(1)}| \le \xi$, $|\mathbf{V}_z^{(2)} - \mathbf{V}_{z+1}^{(2)}| \le \xi$.
Step 8: If both conditions in Step 7 hold, then stop; otherwise set $z = z + 1$ and go back to Step 4.
Step 9: Calculate the center of the Gaussian membership functions in the rule antecedent using the final optimal cluster prototypes and their dispersion degrees using (22).
Step 10: Compute the importance degree of each dimension using (23) and (24) then compute the firing degree of each rule on a new sample using (25).

In the Algorithm 2, the threshold used in the termination criterion step is set to $\xi = 10^{-3}$. Also, $\mathbf{V}_1^{(1)}$ and $\mathbf{V}_1^{(2)}$ represent cluster prototypes corresponding to the class 1 (2) in the first iteration. It should be mentioned that computing the prototypes of the cluster in each class is directly related to the prototypes of the other class. As a result, when the prototypes in one class become stable, the prototypes of the other class will also become stable, thus we can apply the termination criterion on one class and it will automatically apply for the other class.

6. Computational complexity

In Algorithm 1, the total number of iterations for computing the prototypes of the both classes is N that can be dome in $O(N)$ time. The partition matrices U and T can be recomputed in $O(CP)$ time and $O(HP)$ time where CP=HP leading to $O(2CP)$~$O(CP)$ time. The entire process can take Z times so this algorithm can be executed in $O(2(N+CP))$~$O(N+CP)$ time.

In Algorithm 2, cluster prototypes at the beginning of each iteration should be computed in $O(N+CP)$ time. Then, using Algorithm 1, partition matrices should be obtained in $O(CP)$ time. Given the total number of iterations to be Z times, Algorithm 2 can be run in $O(Z(N+2CP))$ time.

Table 1
Summary of the datasets.

NAME	# Features	# Classes	Size
Wine	13	2	130
Parkinson	22	2	195
Parkinson telemetric	22	2	200
Breast cancer	30	2	568
Clinical data	72	2	1413

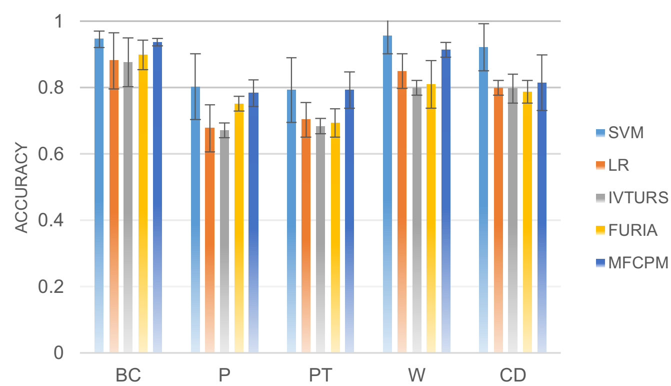

Fig. 1. Cross validation accuracy across different methods.

7. Numerical experiments

This section is devoted to demonstrating the computational performance of the proposed algorithms. To do so, we have used two groups of data sets. The first group is composed of four publicly available datasets downloaded from the University of California Irvine (UCI) machine learning repository [40]. These datasets include Breast Cancer (BC), Parkinson (P), Parkinson Telemetric (PT), and Wine quality (W). Some of the statistics of these datasets are provided in Table 1. We have also implemented the proposed method on a clinical dataset representing vital signs of patients having traumatic injuries. This data (called CD) has been collected by the UCSF/San Francisco General Hospital and Trauma center, and contains 1413 patients. We have used the static part of the dataset in our experiments, which includes demographic data, toxicology screening results, substance use history and hour zero psychological and biological measurements (such as coagulation and inflammatory biomarkers). Hour zero measurements include some of the basic measures such as temperature, blood pressure, respiratory rate, heart rate, platelet count, blood factors such as factor V and VII and protein C level. The output of the data for each sample is the survival of the patient after 28 days, which makes the problem a binary classification case. Note that 30% of the data has been used for validation and the remaining is used in the cross validation process. We have adopted 10-fold cross validation, with a total number of iterations of 50. For the sake of simplicity and comprehension, numerical results have been provided in the form of charts presented in Figs. 1–4.

In order to analyze the effectiveness of the proposed algorithm, we have compared this method with some of the existing state-of-the-art classification algorithms, both in the form of rule-based systems to demonstrate the generalization capabilities of our method, and other algorithms such as Support Vector Machines (SVM) and Logistic Regression (LR). The fuzzy rule-based classification algorithms being used in this study include: FURIA [41], IVTURS [42], and FCPM [11]. On each dataset, two series of

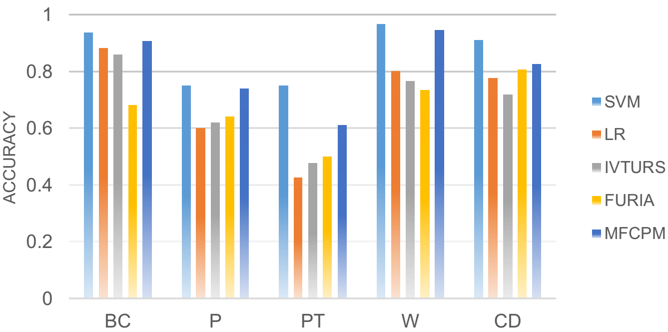

Fig. 2. Validation accuracy across different methods.

experiments were performed including 10-fold cross validation and generalization tests using validation datasets. In addition to the accuracy of each algorithm, we have provided the values of other metrics such as True Positive Rate (TPR, a.k.a Sensitivity), True Negative Rate (TNR, a.k.a Specifity), Positive Predictive Value (PPV), and Negative Predictive Value (NPV). The PPV and NPV measures across all the datasets on both validation and cross-validation experiments are reported in Table. 2.

One of the major flaws of fuzzy rule-based classification systems has always been their poor generalizability which has made them less applicable in real-world problems. A primary goal of this research is to tackle this issue by using the repulsive forces between the prototypes of each cluster in a multivariate fashion taking into account the effects of each individual feature on the classification performance. According to Fig. 1, MFCPM outperforms LR and the three powerful rule-based classification systems, while representing less variation on each iteration of the cross-validation process. It is notable to mention that predictive power of MFCPM has significantly improved compared to FCPM where the entire features are treated equally irrespective of their discriminatory power. It can be observed that in two sets of BC and PT, the MFCPM yields better accuracy while achieving close metrics to the SVM on the remaining datasets. Similar observations can be made in Fig. 2 representing the validation accuracy of each method across five datasets. In dataset P, the accuracy of the MFCPM is equal to that of the SVM, where it gives the second best accuracy level on the other datasets. This is a positive response from the proposed method, illustrating enhanced generalization ability compared to FURIA and IVTURS as two powerful fuzzy rule-based methods. Regarding the validation accuracy numbers, MFCPM and SVM represent equal accuracy level on the Parkinson data, and achieve similar accuracy metrics on the other datasets. The other feature of MFCPM lies in its ability to avoid overfitting when comparing validation and cross-validation performance, since the respective values on both experiments are less different for the proposed method compared to the other methods. Similar behavior can be seen in SVM but the other methods fail to represent such behavior. In general, logistic regression and FURIA methods yield the poorest performance in terms of accuracy.

TPR and FPR measures are good tools to evaluate how a method reacts when encountering non-equal number of positive and negative samples. The higher (lower) the TPR (FPR), the more robust the method in distinguishing class labels. Taking a look at Fig. 3, we note that SVM outperforms the other algorithms on the validation sets, though the difference between MFCPM and SVM in the cross-validation sets shows less difference in terms of TPR. Comparing the three rule-based classification algorithms, MFCPM outperforms IVTURS and FURIA in the cross-validation experiments. It also exhibits better performance for the BC, PT, and W datasets, while yielding performance similar to FURIA for the P and CD datasets.

Table 2
PPV & NPV measures.

		BC		P		PT		W		CD	
		PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
SVM	CV	0.967 ± 0.0456	0.972 ± 0.030	0.942 ± 0.078	0.6427 ± 0.2325	0.938 ± 0.077	0.594 ± 0.21	0.938 ± 0.1204	0.965 ± 0.064	0.902 ± 0.09	0.651 ± 0.04
	Test	0.9091	0.9712	0.9655	0.4762	0.9565	0.5294	0.9375	1	1	0.625
Fuzzy	CV	0.851 ± 00.255	0.869 ± 0.080	0.947 ± 0.0740	0.479 ± 0.0201	0.953 ± 0.0852	0.551 ± 0.0441	0.834 ± 0.074	0.923 ± 0.052	0.885 ± 0.087	0.40 ± 0.017
	Test	0.884	0.91	0.9286	0.41	0.934	0.411	0.934	0.934	0.925	0.43
LR	CV	0.887 ± 0.048	0.848 ± 0.074	0.907 ± 0.052	0.4366 ± 0.023	0.927 ± 0.037	0.444 ± 0.08	0.791 ± 0.022	0.884 ± 0.085	0.875 ± 0.067	0.33 ± 0.051
	Test	0.817	0.874	0.852	0.3043	0.733	0.24	0.8	0.8	0.9038	0.3333
IVTURS	CV	0.877 ± 0.025	0.833 ± 0.066	0.8947 ± 0.036	0.428 ± 0.081	0.895 ± 0.066	0.420 ± 0.08	0.75 ± 0.083	0.84615 ± 0.035	0.873 ± 0.052	0.324 ± 0.03
	Test	0.8196	0.8818	0.85714	0.318182	0.8	0.28	0.7857	0.75	0.865	0.2
FURIA	CV	0.887 ± 0.065	0.863 ± 0.036	0.8947 ± 0.06	0.508 ± 0.071	0.885 ± 0.036	0.472 ± 0.05	0.78 ± 0.033	0.861 ± 0.041	0.872 ± 0.053	0.344 ± 0.02
	Test	0.84375	0.6657	0.83871	0.315789	0.778	0.273	0.7058	0.769	0.907	0.384
FCPM	CV	0.871 ± 00.255	0.847 ± 0.080	0.907 ± 0.074	0.459 ± 0.0201	0.903 ± 0.085	0.551 ± 0.0441	0.784 ± 0.074	0.892 ± 0.052	0.865 ± 0.087	0.35 ± 0.017
	Test	0.854	0.882	0.8986	0.41	0.883	0.311	0.834	0.834	0.905	0.373

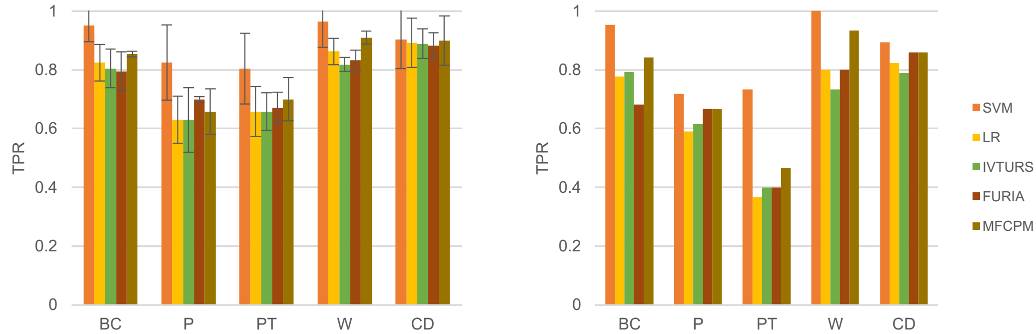

Fig. 3. Sensitivity measures (TPR).

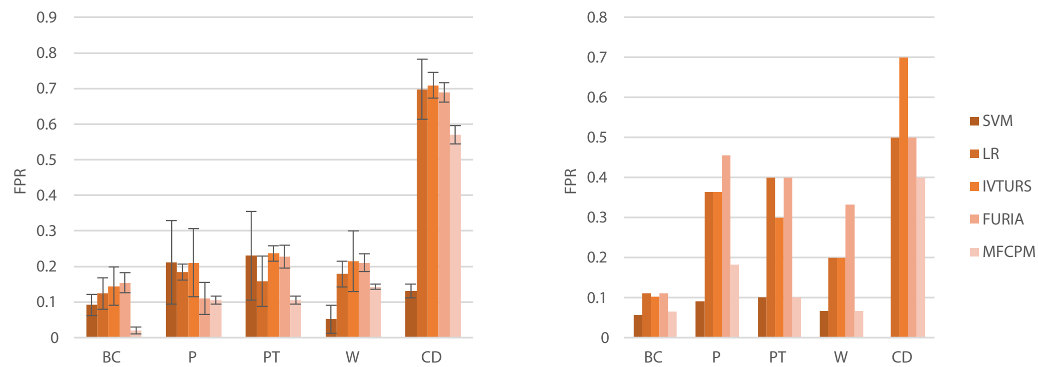

Fig. 4. False positive rate (FPR).

A similar situation can be observed in Fig. 4, where MFCPM exhibits the lowest FPR among all other methods for the BC, P, and PT datasets, while ranking second for W and CD. The FPR numbers for the CD are significantly lower for SVM, compared to the other four methods. Nonetheless, in the validation data, the difference between the FPR values of the SVM and MFCPM methods is rather small. Overall these two methods exhibit similar metrics on all the datasets.

One of the primary contributions of this research is to close the gap between fuzzy rule-based classification algorithms and the state-of-the-art non fuzzy methods such as SVM. This goal has been achieved, according to the PPV and NPV values presented in Table 2. Similar to the other experimental results presented so far, MFCPM outperforms LR, FURIA, and IVTURS and yields accuracy measures similar to the SVM results. MFCPM is able to handle topological conditions of the samples in each class label so that the largest number of correctly classified positive and negative samples in both cross-validation and validation experiments can be achieved.

The number of positive and negative samples in all datasets are not equal. For example, in the CD the ratio of positive labels to negative labels is roughly 65%. This observation denotes that the classifier tend to classify most of the unseen samples as positive, yielding large TPR and FPR. According to Fig. 3, the TPR values of the SVM and MFCPM algorithms are among the highest, for both the training and test experiments. Like most classification algorithms encountering such data, we expect to achieve a large FPR for MFCPM method on the unseen samples. Nevertheless, The FPR of the MFCPM on CD data in the test experiment is the lowest among all the other methods (Fig. 4). This can be translated into the repulsive forces between the class labels when constructing the rules along with looking at the most important features instead of assigning equal importance to the entire feature space. According to Fig. 3, similar observations can be made in the BC, P, PT and W datasets, where the ratio of positive training samples to negative vectors is 55%–60%. In all of these datasets, MFCPM outperforms its rule-based counterparts and yields similar metrics to SVM. The main strength of MFCPM can be verified in Fig. 4, where its FPR values on the PT and W data on the test experiments are the best while yielding the second best FPRs on the BC and P data. Similar conclusions can be made according the information presented in Table 2.

Table 3
Pairwise McNemar's statistical test results.

	BC	P	PT	W	CD
MFCPM-LR	0.051	0.0012	0.00087	0.00088	0.038
MFCPM-IVTURS	0.052	0.0035	0.00090	0.00077	0.015
MFCPM-FURIA	0.0014	0.0041	0.0013	0.00065	0.063
MFCPM-SVM	0.072	0.083	0.031	0.066	0.022
MFCPM-FCPM	0.0134	0.021	0.0022	0.0097	0.056

For further investigation of the differences between classification performance of the other classifiers compared to our proposed algorithm, we have conducted McNemar's statistical test to see whether or not the misclassification rates between different classifiers are statistically significant. For this purpose, we have conducted pairwise statistical experiments on the outputs of the MFCPM versus the other methods. The confidence level of 0.05 was taken in all these tests. The statistical p-values for the pairwise test along the entire datasets on the test sets are represented in Table 3.

The highlighted cells indicate tests which are statistically significant. Putting these numbers beside Fig. 2 graph, we can observe that MFCPM demonstrates superior performance compared to the other method particularly fuzzy-based algorithms. MFPCM yields significant improvements compared with FCPM while presenting four significant test out of five. The only dataset where MFCPM and FCPM output similar metrics is CD. This is also the case with LR, IVTURS and FURIA where MFCPM significantly improves the

Table 4
The number of generated fuzzy rules.

	W	P	PT	B	CD
MFCPM	6	7	6	8	10
IVTURS	8	12	12	15	17
FURIA	11	14	13	20	28
FCPM	7	8	6	9	12

classification accuracy. On the other hand, MFCPM and SVM indicate pretty similar accuracy over BC, P, and W while SVM outperforms MFCPM in PT and CD datasets. These test confirm how the fuzzy logic based classification system proposed in this research has decreased the gap between fuzzy methods such as FURIA and IVTURS versus non-fuzzy algorithms such as SVM in terms of accuracy on various datasets. The total number of generated rules in the rule-base for the compared fuzzy methods provided in Table 4.

It can be observed in most cases, MFCP can capture the structure of the data with the lowest number of rules implicating less complex system structure while FCPM shows close numbers to the MFCPM. For example, in PT, both methods share the same number of rules while in W, P, and B FCPM generates one rule more than MFCPM and 2 rules more in the CD. On the other hand, the total number of generated rules in IVTURS and FURIA are significantly larger than MFCPM which goes back to their structure identification process that contains sophisticated global-search for tuning the rules which leads to an increase in the number of generated rules. We also checked the number of identified support vectors found by the SVM algorithm to be 17, 30, 28, 79, and 152 in the W, P, PT, B, and CD datasets, respectively. In terms of the running time, we have ranked the benchmarked algorithm and have sorted as follows: (1) FCPM, (2) MFCPM, (2) SVM, (4) IVTURS, (5) FURIA. This in accordance with the number of rules being generated by the fuzzy models being compared and demonstrates that MFCPM outperforms almost all other methods while providing reasonable accuracy level.

8. Conclusion and future study

This paper presents a multivariate representation of the FCPM clustering algorithm along with some other approaches regarding determination of the importance weight of each feature in the rule-base of fuzzy classification systems. This work is one of the very first efforts to take into account the repulsive forces between the prototypes of overlapping clusters in order to enhance the generalization capabilities of rule-based classification systems. Besides using publicly available datasets, an effort was made to collect medical data from patients having traumatic injuries. This valuable dataset was used to demonstrate the superior performance of the proposed method. We showed that the proposed approaches are able to close the gap between rule-based classification systems and other powerful classification algorithms, in terms of accuracy and generalization. In this study, we have only concentrated on conventional Euclidean distances due to its widespread use in the scientific community. This may cause in some problems whose data structure is more complicated or follows specific geometric properties that cannot captured by Euclidean distance. Using graph-theoretic approaches and manifold distances can be a great alternative for future works in this area and to analyze effectiveness of such methods compared to the current work. Another future proposal can be the application of higher order fuzzy sets and deriving their respective formulations to enhance sensitivity and specifity of the current approach.

Author contributions statement

A.D.T. developed the algorithms, implemented the method, conducted the experiments and prepared the manuscript. L.P. was the senior supervisor of this study, editing the manuscript from technical and writing perspectives. M.C. was responsible for collecting the clinical data.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgment

The authors gratefully acknowledge the financial support by the Institute for Collaborative Biotechnologies through Coagulopathy grant W911NF-1O-2-0114 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

[1] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, 2001.
[2] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.
[3] A. Webb, Statistical Pattern Recognition, Arnold, London, U.K., 1999.
[4] R. Ratsch, T. Onoda, K.R. Muller, Soft margins for AdaBoost, Mach. Learn. 42 (3) (2001) 287–320.
[5] S. Mika, G. Ratsch, J. Weston, S. Scholkopf, K.R. Muller, Fisher discriminant analysis with kernels, Proc. Neural Netw. Signal Process. (1999) 41–48.
[6] I.W. Tsang, J.T. Kwok, J.M. Zurada, Generalized core vector machines, IEEE Trans. Neural Netw. 17 (5) (2006) 1126–1140.
[7] I.W. Tsang, A. Kocsor, J.T. Kwok, Large-Scale maximum margin discriminant analysis using core vector machines, IEEE Trans. Neur. Netw. 19 (4) (2008) 610–623.
[8] O.L. Mangasarian, D.R. Musicant, Lagrangian support vector machines, J. Mach. Learn. Res. 1 (1) (2001) 161–177.
[9] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.
[10] V. Vapnik, Statistical Learning Theory, Wiley, 1998.
[11] J.M. Leski, Fuzzy (c+p)-means clustering and its application to a fuzzy rule-based classifier: toward good generalization and good interpretability, IEEE Trans. Fuzzy Syst. 23 (4) (2015) 802–812.
[12] E. Lughofer, O. Buchtala, Reliable all-pairs evolving fuzzy classifiers, IEEE Trans. Fuzzy Syst. 21 (4) (2013) 625–641.
[13] A. Doostparast Torshizi, L. Petzold, M. Cohen, Direct higher order fuzzy rule-based classification system: application in mortality prediction, in: Proceedings of the International Conference on Bioinformatics and Biomedicine, Washington D. C., 2015, pp. 846–852.
[14] P.P. Angelov, X. Zhou, Evolving fuzzy-rule-based classifiers from data streams, IEEE Trans. Fuzzy Syst. 16 (6) (2008) 1462–1475.
[15] X. Yang, G. Zhang, J. Lu, J. Ma, A kernel fuzzy c-means clustering based fuzzy support vector machine algorithm for classification problems with outliers or noises, IEEE Trans. Fuzzy Syst. 19 (1) (2011) 105–115.
[16] J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[17] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[18] W. Pedrycz, H. Izakian, Cluster-Centric Fuzzy Modeling, IEEE Trans. Fuzzy Syst. 22 (6) (2014) 1585–1597.
[19] A. Doostparast Torshizi, M.H. Fazel Zarandi, Hierarchical collapsing method for direct defuzzification of general type-2 fuzzy sets, Inf. Sci 277 (2014) 842–861.
[20] A. Doostparast Torshizi, M.H. Fazel Zarandi, G. Doostparast, K. Eghbali, A hybrid fuzzy-ontology based intelligent system to determine level of severity and treatment recommendation for Benign Prostatic Hyperplasia, *Comput. Meth. Programs Biomed.* 113 (20114) 301–313.
[21] M.H. Fazel Zarandi, A. Doostparast Torshizi, I.B. Turksen, B. Rezaei, A new indirect approach to the type-2 fuzzy systems modeling and design, Inf. Sci 232 (2013) 346–365.
[22] C. Murcia, G. Bonilla, M. Melgarejo, Fuzzy classifiers tuning through an adaptive memetic algorithm, IEEE Trans. Fuzzy Syst. 12 (2) (2014) 197–204.
[23] M. Antonelli, P. Ducange, F. Marcelloni, Genetic training instance selection in multiobjective evolutionary fuzzy systems: a coevolutionary approach, IEEE Trans. Fuzzy Syst. 20 (2) (2011) 276–290.

[24] P. Ganeshkumar, C. Rani, D. Devraj, T.A. Victoire, Hybrid ant bee algorithm for fuzzy expert system based sample classification, IEEE/ACM Trans. Comput. Biol. Bioinform. 11 (2) (2014) 347–360.
[25] O. Gordo, E. Martinez, C. Gonzalo, A. Arquero, Classification of satellite images by means of fuzzy rules generated by a genetic algorithm, IEEE Lat. Am. Trans. 9 (1) (2011) 743–748.
[26] D.G. Stavrakoudis, G.N. Galidaki, I.Z. Gitas, J.B. Theocharis, A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery, IEEE Trans. Geosci. Rem. Sens. 50 (1) (2011) 130–148.
[27] E. Mansoori, M.J. Zolghadri, S.D. Katebi, SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzzy Syst. 16 (4) (2008) 1061–1071.
[28] Y. Chen, B. Yang, A. Abraham, L. Peng, Automatic design of hierarchical Takagi–Sugeno Type fuzzy systems using evolutionary algorithms, IEEE Trans. Fuzzy Syst. 15 (3) (2007) 385–397.
[29] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, J. Lu, Fuzzy regression transfer learning in Takagi–Sugeno fuzzy models, IEEE Trans. Fuzzy Syst. (2016) 99.
[30] F. Juang, C.D. Hsieh, TS-fuzzy system-based support vector regression, Fuzzy Sets Syst 160 (17) (2009) 2486–2504.
[31] C.F. Juang, S.H. Chiu, S.J. Shiu, Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation, IEEE Trans. Syst. Man Cybern. A Syst. Hum. 37 (6) (2007) 1077–1087.
[32] Y. Jiang, Z. Deng, F.L. Chung, S. Wang, Realizing two-view TSK fuzzy classification system by using collaborative learning, IEEE Trans. Syst. Man Cybern.: Syst. 47 (1) (2017) 145–160.
[33] M.C. Du Plessis, G. Niu, M. Sugiyama, Clustering unclustered data: unsupervised binary labeling of two datasets having different class balances, in: Proceedings of the Conference on Technologies and Applications of Artificial Intelligence, 2013, pp. 1–6.
[34] B. Jiang, J. Pei, Y.F. Tao, X.M. Lin, Clustering uncertain data based on probability distribution similarity, IEEE Trans. Knowl. Data Eng. 25 (4) (2013) 751–763.
[35] B.A. Pimentel, R.M. De Souza, A multivariate fuzzy c-means method, Appl. Soft Comput. 13 (2013) 1592–1607.
[36] B.A. Pimentel, R.M. De Souza, A weighted multivariate fuzzy C-means method in interval-valued scientific production data, Expert Syst. Appl. 41 (2014) 3223–3236.
[37] B.A. Pimentel, B.A. De Souza, Multivariate fuzzy C-means algorithms with weighting, Neurocomputing 174 (B) (2016) 946–965.
[38] J.C. Bezdek, Pattern Recognition With Fuzzy Objective Function Algorithms, Plenum, New York, 1982.
[39] A. Doostparast Torshizi, M.H. Fazel Zarandi, A new cluster validity measure based on general type-2 fuzzy sets: application in gene expression data clustering, Knowl. Based Syst. 64 (2014) 81–93.
[40] http://archive.ics.uci.edu/ml/. Accessed Sep. 2016.
[41] J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[42] J. Huhn, E. Hullermeier, FURIA: An algorithm for unordered fuzzy rule induction, Data Min. Knowl. Discov. 19 (3) (2009) 293–319.

Abolfazl Doostparast Torshizi is an Associate Research Scientist at Institute for Genomic Medicine (IGM), Columbia University. Prior to joining Columbia University, he was a postdoc researcher in the department of Computer Science at University of California, Santa Barbara. He holds a Ph.D. in Industrial and Systems Engineering. His research interests include: biomedical informatics, omics-data analysis and modeling, and machine learning.

Linda Petzold is a professor of Computer Science at University of California, Santa Barbara. She holds a Ph.D. in Computer Science minor in Mathematics from University of Illinois at Urbana Champagne. Her research interests include: medical data analytics, Systems Biology, and Ordinary Differential Equations (ODEs).

Mitchell Jay Cohen is the director of surgery at Denver Health medical center and a professor in the department of surgery at the University of Colorado School of Medicine. He is a professor of surgery in residence at the University of California-San Francisco and San Francisco General Hospital, where he also is associate trauma medical director and director of trauma research. Cohen received his medical degree from the Mount Sinai School of Medicine in New York in 1997 and did his surgical residency at Rush University in Chicago and a research fellowship at Loyola University Medical Center.

Please cite this article as: A.D. Torshizi et al., Multivariate soft repulsive system identification for constructing rule-based classification systems: Application to trauma clinical data, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.03.043