

Empirical Fuzzy Sets

Plamen P. Angelov^{1,2} and Xiaowei Gu^{1*}

¹School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA, UK

²Honorary Professor, Technical University of Sofia, boulevard Kliment Ohridski 8, Sofia, 1000, Bulgaria

e-mail: {[p.angelov](mailto:p.angelov@lancaster.ac.uk), [x.gu3](mailto:x.gu3@lancaster.ac.uk)}@lancaster.ac.uk

Abstract

In this paper, we introduce a new form of describing fuzzy sets (FSs) and a new form of fuzzy rule-based (FRB) systems, namely, empirical fuzzy sets (ϵ FSs) and empirical fuzzy rule-based (ϵ FRB) systems. Traditionally, the membership functions (MFs), which are the key mathematical representation of FSs, are designed subjectively or extracted from the data by clustering projections or defined subjectively. ϵ FSs, on the contrary, are described by the empirically derived membership functions (ϵ MFs). The new proposal made in this paper is based on the recently introduced Empirical Data Analytics (EDA) computational framework and is closely linked with the density of the data. This allows to keep and improve the link between the objective data and the subjective labels, linguistic terms and classes definition. Furthermore, ϵ FSs can deal with heterogeneous data combining categorical with continuous and/or discrete data in a natural way. ϵ FRB systems can be extracted from data including data streams and can have dynamically evolving structure. However, they can also be used as a tool to represent expert knowledge. The main difference with the traditional FSs and FRB systems is that the expert does not need to define the MF per variable; instead, possibly multimodal, densities will be extracted automatically from the data and used as ϵ MFs in a vector form for all numerical variables. This is done in a seamless way whereby the human involvement is only required to label the classes and linguistic terms. Moreover, even this intervention is optional. Thus, the proposed new approach to define and design the FSs and FRB systems puts the human “in the driving seat”. Instead of asking experts to define features and MFs correspondingly, to parameterize them, to define algorithm parameters, to choose types of MFs or to label each individual item, it only requires (optionally) to select prototypes from data and (again, optionally) to label them. Numerical examples as well as a naïve empirical fuzzy (ϵ F) classifier are presented with an illustrative purpose. Due to the very fundamental nature of the proposal it can have a very wide area of applications resulting in a

series of new algorithms such as ϵ F classifiers, ϵ F predictors, ϵ F controllers, etc. This is left for the future research.

Keywords—membership functions, AnYa type fuzzy rule-based systems, empirical data analytics, naïve empirical fuzzy rule-based classifier, non-parametric.

1. Introduction

Fuzzy sets (FSs) theory and the fuzzy rule-based (FRB) systems have been defined over 50 years ago in the seminal paper by Professor Lotfi Zadeh [1] and now matured [2]. Since mid-1970s (Mamdani or Zadeh-Mamdani) [3] and since mid-1980s (Takagi-Sugeno) [4] FRB systems started to be developed and are now widely applied. Although, there are other types of fuzzy systems (relational [5], etc.), one particular type that was introduced recently by Angelov and Yager [6] called AnYa deserves a special attention. Both Mamdani and Takagi-Sugeno type of FRB share the exact same antecedent (IF) part and only (although significantly) differ by the consequent (THEN) part. AnYa type FRB, however, has a quite different antecedent (IF) part.

The main issue in the design of the FSs and FRB systems is how to define the MFs by which they are defined in first place. The traditional way of designing FSs, the subjective approach, has its own very strong rationale in the two-way process of: *i*) formalizing expert knowledge and representing it in a mathematical form through the membership functions (MFs), and *ii*) representing and extracting from data human-intelligible and understandable, transparent linguistic information in the form of IF ...THEN... rules. In 1990s, in addition to the traditional subjective way of designing FSs, the so-called *data driven* design method started to be popular and was developed. Nonetheless, it is practically very difficult and controversial to define MFs both from experts and from data. This is also related to the more general issue of assumptions made and handcrafting that machine learning (including statistical methods) are facing and is now hotly researched.

In this paper, a new approach is proposed to the way in which the FSs and FRB systems are being defined. We call such sets empirical fuzzy sets (ϵ FSs) and such FRB systems - empirical fuzzy rule-based (ϵ FRB) systems. The new approach is grounded at the recently introduced more general concept and a computational framework of Empirical Data Analytics (EDA) [7], [8]. In the next sections, we will recall the basics of FSs, FRB systems as well as EDA, and on this basis, we will further introduce the ϵ FSs and ϵ FRB systems. We will demonstrate how ϵ FSs and ϵ FRB systems allow preserving the subjective specifics that FSs and FRB are strong with. At the same time, it will be shown how ϵ FSs and ϵ FRB systems can benefit from the vast amount of data that may be available. ϵ FSs and ϵ FRB systems will still allow extracting expert knowledge by questionnaires or other forms, but will make this much easier for the experts and not ambiguous (the experts will not be asked to define

membership values or parameters, but only (optionally) the labels/names of the linguistic terms, classes (if any)). For example, if we choose a car, we can simply say which one we like (or possibly how much), but we do not need to define each feature (price, max speed, etc.) or specify why. Moreover, with the proposed ϵ FSs and ϵ FRB systems, one can tackle heterogeneous data and combine categorical (e.g. gender, occupation, number of doors) with continuous and/or discrete variables like price, max speed, size, etc. Further, in this paper, we will demonstrate how, on the basis of ϵ FSs and ϵ FRB systems, one can build empirical fuzzy classifiers (ϵ F Classifiers), predictors (ϵ F Predictors), controllers (ϵ F Controllers), etc. Moreover, these can be evolving, not just fixed structure. This will allow studying the dynamic changes in human preferences as well as building more efficient recommendation systems where the only necessary input from the users is the preference (“likes” or “retweets” or “clicks”).

The newly proposed approach of definition and design of FSs and FRB systems is very suitable for the current era of so-called *big data*. Indeed, there is an exponential growth in the scale and complexity of the data sets and streams being generated by sensors, people, society, industry, etc. This is being increasingly seen as an untapped resource, which offers new opportunities for extracting aggregated information to inform decision-making in policy and commerce. It also stretches the existing techniques because they were developed in the era when the data was not in so large scale and was assumed to be simple and available offline, not streaming and dynamically evolving in a possibly non-stationary manner. It is practically difficult to design a traditional fuzzy (or statistical, for that matter) model if we have a huge amount of unlabelled images, big data representing customer choices or preferences, etc. In contrast, the proposed ϵ FSs and ϵ FRB systems offer an efficient and data-centred (thus, empirical) tool that is clear and intuitive, yet not *ad hoc*, and can facilitate and empower the human experts and users instead of overloading or overwhelming them. The flowcharts of the traditional approaches and the proposed approaches for FRB system identification are presented in Fig. 1 for comparison.

The remainder of this paper is organised as follows. Section 2 recalls the concepts of the FSs and FRB systems and the Empirical Data Analytic (EDA) frameworks. The proposed approach is described in section 3, two versions (objective and subjective) for FRB system identification are introduced in this section as well. A new type of FRB classifier is introduced in section 4, numerical examples and discussion are also presented. Finally, section 5 concludes the paper.

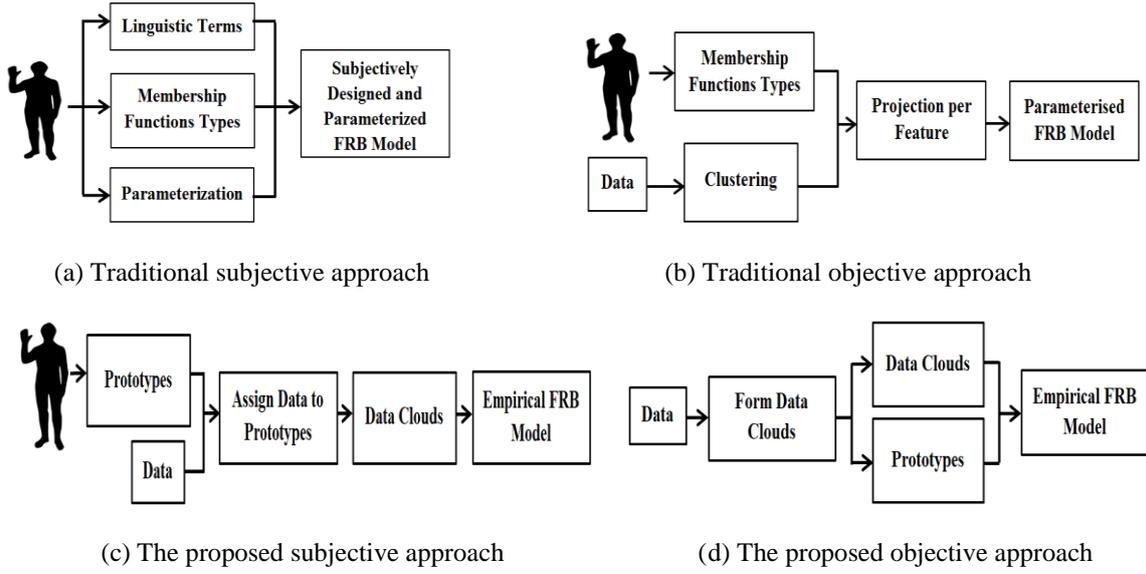


Fig.1. The flowcharts of the traditional approaches and the proposed approaches for system identification

2. Theoretical Basis

In this section, we will recall the theoretical basis needed for the proposed approach.

2.1. Fuzzy Sets and Fuzzy Rule-Based Systems

In this subsection, we will compare the Mamdani type [3], Takagi-Sugeno type [4] and AnYa type FRB systems [6]. To begin with, let us start with an illustrative example.

If we want to build a Mamdani type or Takagi-Sugeno type FRB system to divide hundreds of domestic dogs into three groups (“Small”, “Medium” and “Large”) based on their size in terms of length and weight, the following parameters are needed to be defined in order to build the antecedent (IF) parts of the fuzzy rules (FRs) [9]:

- i) the types of MFs, i.e. triangular type, Gaussian type, bell type, etc.
- ii) linguistic terms for each FR;
- iii) the area of influence for each FR, i.e. hyper-rectangle, -sphere, -ellipsoid (this is closely linked to the types of distance metric used);
- iv) the prototypes for the FSs;
- v) the parameters for the MFs.

To classify the dogs into three groups, we firstly build three FRs expressed linguistically as follows:

$Rule_1: IF (Length \text{ is } Short) AND (Weight \text{ is } Light) THEN (Size \text{ is } Small)$

$Rule_2: IF (Length \text{ is } Medium) AND (Weight \text{ is } Medium) THEN (Size \text{ is } Medium)$

Rule₃: IF (Length is Long) AND (Weight is High) THEN (Size is Large)

Based on the data measured from the 600 domestic dogs (the distribution of the data samples is shown in Fig. 2, the data is synthetic), for the linguistic variable “Length”, we might interpret “Short” as “around 20 cm”, “Medium” as “around 37 cm” and “Long” as “around 54 cm”. For the linguistic variable “Weight”, “Light” could be interpreted as “around 15 kg”, “Medium” as “around 32 kg” and “High” as “around 48 cm”. After we select the type of MFs (for example, triangular and Gaussian) and decide other parameters, finally we obtain the FRB systems as depicted in Fig.3.

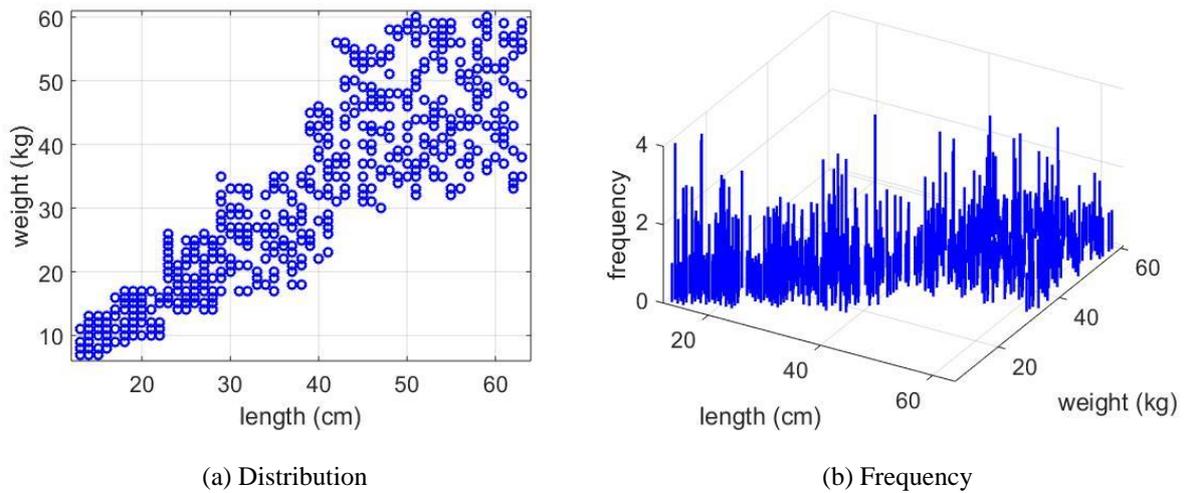


Fig.2. Distribution of the sizes of 600 domestic dogs

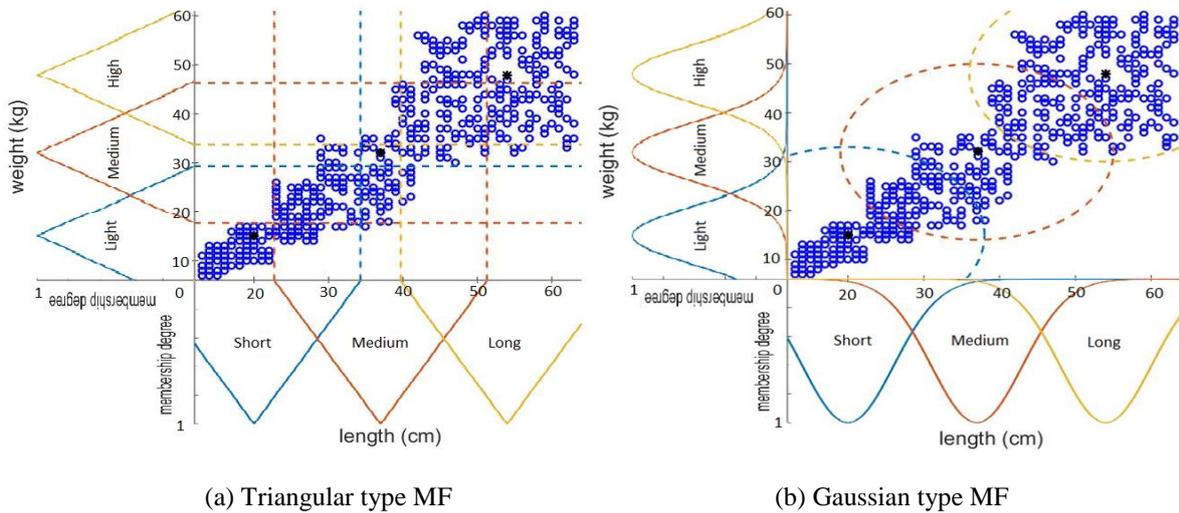


Fig.3. Examples of Mamdani type and Takagi-Sugeno type FRB systems (the black asterisks are the prototypes)

From the above example, one can see the following issues during the process:

- i) Defining a MF requires many *ad hoc* decisions;
- ii) MFs often differ significantly from the real data distribution

In addition, the so-called “curse of dimensionality” may result from handcrafting traditional FRB systems for high dimensional problems because of the exponential growth of the number of FSs required.

Alternatively, to design an AnYa type FRB system [6] with the same prototypes as being used in the previous example, one can easily form three data clouds and, based on them, derive three AnYa type FRs as follows. The visualization of the three data clouds, which are also the areas of influence of the three FRs, is provided in Fig.4.

$Rule_1: IF ([Length, Weight] \sim [20cm, 15kg]) THEN (Size is Small)$

$Rule_2: IF ([Length, Weight] \sim [37cm, 32kg]) THEN (Size is Medium)$

$Rule_3: IF ([Length, Weight] \sim [54cm, 48kg]) THEN (Size is Large)$

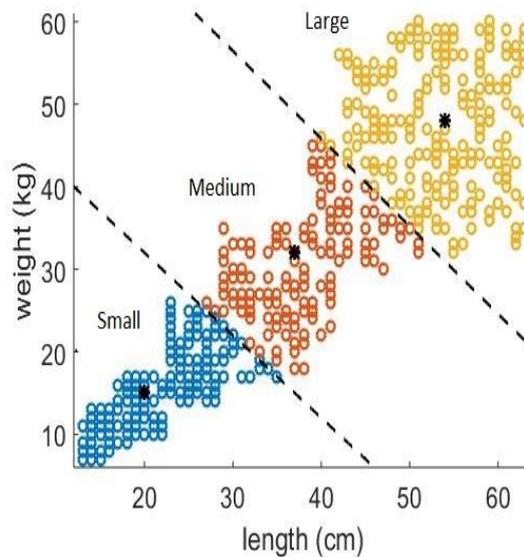


Fig.4. AnYa type FRB system

As one can see, the AnYa type FRB system [6] simplifies the process of designing MFs and FRs. They are uniquely defined by vectors representing the focal points of the non-parametric, shape-free data clouds consisting of data samples associated with the nearest focal points resembling Voronoi tessellation [10]. The data clouds are then used as the antecedent (IF) part of each AnYa type FR. This significantly reduces the efforts of human experts and, at the same time, largely enhances the objectiveness of the FRB system.

To build the Mamdani type and Takagi-Sugeno type FRB systems, one needs to define a number of parameters as described at the beginning of this subsection, but for AnYa type, the only decision that human experts need to make is to find the focal points, which require much less efforts. This advantage makes the AnYa type FRB more competitive in high dimensional problems. In fact, the focal points can be identified in a fully autonomous, data-driven way using the technique for forming data clouds as described in [11]. The focal points identified via

this approach require no user- and problem- specific parameters, namely no human involvement, and can objectively represent the local modes (peaks) of the data distribution

From the comparison between the AnYa type and the traditional type FRB systems one can see that, although traditional MFs and FRB systems contain too many *ad hoc* choices and often require significant expertise, they have the advantage of the high interpretability. The simplicity of the AnYa FRB systems significantly reduces the needs of human expertise and thus, enhances the objectiveness, but at the same time, the simplicity reduces the interpretability and leads to the loss of information. Therefore, in this paper, we will introduce a new type of FSs and FRB systems named, empirical fuzzy set (ϵ FS) and empirical fuzzy rule based system (ϵ FRB) to combine the advantages of the traditional type FRB with the recently introduced AnYa type FRB.

2.2. Empirical data analytics framework

Empirical data analytics (EDA) framework [7], [8] is a recently introduced methodology for data analysis free from pre-defined parameters and assumptions. The main EDA quantities used in this paper include [7], [8]:

i) *Unimodal density*;

ii) *Multimodal density*.

First of all, let us assume a collection of data samples of a data set/stream denoted by $\{\mathbf{x}\}_K = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$ ($\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]^T$, $i = 1, 2, \dots, K$), where K is the number of the observed data samples; the subscript i indicates the time instance at which the i^{th} data sample was observed. More generally, we assume that some data samples in $\{\mathbf{x}\}_K$ repeated more than once, namely, $\exists \mathbf{x}_i = \mathbf{x}_j, i \neq j$. As a result, the set of the sorted unique data samples is defined as $\{\mathbf{u}\}_L = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L\}$ and the frequencies of occurrence are defined as $\{f_1, f_2, \dots, f_L\}$, where $\sum_{i=1}^L f_i = K$ and $L (L \leq K)$ is the number of unique data samples.

i) *Unimodal density*

Unimodal density indicates the main mode (peak) of the data distribution and plays an important role in the data analysis. The *unimodal density* of a particular data sample, denoted as \mathbf{x}_i is defined as follows [7], [8]:

$$D(\mathbf{x}_i) = \frac{\sum_{k=1}^K \sum_{j=1}^K d^2(\mathbf{x}_k, \mathbf{x}_j)}{2K \sum_{j=1}^K d^2(\mathbf{x}_i, \mathbf{x}_j)} \quad (1)$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between \mathbf{x}_i and \mathbf{x}_j , which can be any type of distance metric.

For Euclidean distance, the *unimodal density* takes a form of **Cauchy function** [8]:

$$D(\mathbf{x}_i) = \frac{1}{1 + \frac{\|\mathbf{x}_i - \boldsymbol{\mu}\|^2}{X - \|\boldsymbol{\mu}\|^2}} \quad (2)$$

where $\sum_{j=1}^K \|\mathbf{x}_i - \mathbf{x}_j\|^2 = K \left(\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 + X - \|\boldsymbol{\mu}\|^2 \right)$ (3)

$$\sum_{k=1}^K \sum_{j=1}^K \|\mathbf{x}_k - \mathbf{x}_j\|^2 = 2K^2 \left(X - \|\boldsymbol{\mu}\|^2 \right) \quad (4)$$

and $\boldsymbol{\mu} = \frac{1}{K} \sum_{i=1}^K \mathbf{x}_i$ is the global mean of $\{\mathbf{x}\}_K$ and $X = \frac{1}{K} \sum_{i=1}^K \|\mathbf{x}_i\|^2$ is the corresponding average scalar product.

ii) Multimodal density

The *multimodal density* [7], [8] of a unique data sample, u_i is defined as a weighted *unimodal density* by the corresponding frequencies of occurrence, expressed as:

$$D^{MM}(\mathbf{u}_i) = f_i D(\mathbf{u}_i) = f_i \frac{\sum_{k=1}^K \sum_{j=1}^K d^2(\mathbf{x}_k, \mathbf{x}_j)}{2K \sum_{j=1}^K d^2(\mathbf{x}_i, \mathbf{x}_j)} \quad (5)$$

Similarly, $D^{MM}(\mathbf{u}_i) = \frac{f_i}{1 + \frac{\|\mathbf{u}_i - \boldsymbol{\mu}\|^2}{X - \|\boldsymbol{\mu}\|^2}}$ if the Euclidean type of distance is used.

Multimodal density has the ability of disclosing the local modes of the data distribution directly from the data without using iterative searching algorithms [7], [8].

It has to be stressed that the *unimodal density* and *multimodal density* are obtained from and only valid for the observed data samples. We also have to stress that *unimodal density* and *multimodal density* are not limited to the Euclidean type of distance; other types of distances can be considered as well, but in our paper, we would use the Euclidean distance in the visual examples for simplicity.

3. The Proposed Approach

In this section, we will introduce the proposed concepts of the empirical fuzzy sets (ϵ FSs) and empirical fuzzy rule-based (ϵ FRB) systems in detailed. Two system identification approaches (*i*) objective and *ii*) subjective) will be described as well.

3.1. ϵ FSs and ϵ FRB Systems

The new concepts of the ϵ FSs and ϵ FRB systems are grounded at the recently introduced general computational framework of Empirical Data Analytics (EDA) [7], [8]. From the comparison in section 2.1 one can see that traditional FSs represented by MFs require large amount of expert knowledge and efforts to be built. While AnYa type FRB systems have the strong advantage of simplicity and objectiveness.

Firstly, let us consider a m -dimensional vector of categorical variables, $\mathbf{c} = [c^1, \dots, c^m]$; c^j is the j^{th} categorical variable of \mathbf{c} ($j = 1, \dots, m$, m is a non-negative integer); $\boldsymbol{\chi} = [\mathbf{c}, \mathbf{x}]$ is a particular data sample within the data set/stream; \mathbf{x} is the continuous and/or discrete part of $\boldsymbol{\chi}$ and \mathbf{c} is the categorical part; $\boldsymbol{\Gamma}_i = [\Gamma_i^1, \Gamma_i^2, \dots, \Gamma_i^m]$ is the vector of categorical variables of the i^{th} prototype; Γ_i^j is the corresponding j^{th} categorical variable; the categorical variables can be gender, occupation, brand, etc.; the set of possible values of the j^{th} categorical variable is denoted by $category^j$; c^j, Γ_i^j can only take on one value from $category^j$. Based on the AnYa type of FRB, we introduce the ϵ FR in a general form as follows.

$$Rule_i : IF (\mathbf{c} = \boldsymbol{\Gamma}_i) \quad THEN IF (\mathbf{x} \sim \mathbf{prototype}_i) \quad THEN (Class_i) \quad (6)$$

The output of the categorical (IF) part in the proposed ϵ FR is a Boolean (“true” or “false” only) expressed as:

$$B_i(\mathbf{c}) = \begin{cases} 1 & \mathbf{c} = \boldsymbol{\Gamma}_i \\ 0 & \mathbf{c} \neq \boldsymbol{\Gamma}_i \end{cases} \quad (7)$$

At least one prototype is required for each category in order to build the ϵ FR. For data that contains multiple categorical variables, i.e. c^1, \dots, c^m , at least $A = \prod_{j=1}^m a^j$ prototypes are needed, where a^j is the cardinality of the set $category^j$ ($j = 1, 2, \dots, m$).

We further define the empirical membership function (ϵ MF) of the ϵ FR for the continuous and/or discrete part, \mathbf{x} in the form of *unimodal density*, which is derived automatically from the data cloud around the prototype:

$$\epsilon MF_i(\mathbf{x}) = D_i(\mathbf{x}) = \frac{\sum_{k=1}^{S_i} \sum_{j=1}^{S_i} d^2(\mathbf{x}_{i,k}, \mathbf{x}_{i,j})}{2S_i \sum_{j=1}^{S_i} d^2(\mathbf{x}, \mathbf{x}_{i,j})} \quad (8)$$

where S_i is the support (number of members) of the i^{th} data cloud; $\mathbf{x}_{i,k}$ denotes the k^{th} member within this data

cloud, $k = 1, 2, \dots, S_i$. If Euclidean type of distance is used, then $\varepsilon MF_i(\mathbf{x}) = \frac{1}{1 + \frac{\|\mathbf{x} - \mathbf{p}_i\|^2}{X_i - \|\mathbf{p}_i\|^2}}$; here

$\mathbf{p}_i = [p_{i,1}, p_{i,2}, \dots, p_{i,d}]$ is the prototype (centre) of the i^{th} data cloud; X_i is the corresponding average scalar product.

The degree of membership for \mathcal{X} is defined as the product of the output of the categorical antecedent (IF) part and the output of the continuous and/or discrete antecedent (IF) part:

$$\lambda_i(\mathcal{X}) = B_i(\mathbf{c}) \cdot \varepsilon MF_i(\mathbf{x}) \quad (9)$$

If there is no categorical variable in the data, the antecedent part of εFR is reduced to the vector form used by AnYa type FR (but with εMF):

$$Rule_i : IF (\mathbf{x} \sim \mathbf{prototype}_i) THEN (Class_i) \quad (10)$$

in which the εMF is defined by equation (8).

In contrast to the Mamdani type and Takagi-Sugeno type FRB systems as presented in Fig. 3, the εMF is naturally in the form of a Cauchy function if Euclidean distance is used (see equations (2) and (5)). However, instead of manually selecting the Gaussian type MF or the triangular type MF and parameterising them, the εMF is derived from the data automatically based on the *unimodal density* of the data. Moreover, the proposed approach does not need to partition the data space with the manually defined shapes. Data samples will be attracted by the prototypes and form a number of shape-free data clouds around the prototypes automatically resembling Voronoi Tessellation [10], see Figs. 4 and 5.

Unlike the MFs used in the Mamdani type and Takagi-Sugeno type FRB systems that are defined per feature, the εMF s are extracted from data in a vector form. Nonetheless, one still can draw $(n+1)$ dimensional εMF s based on the particular n ($1 \leq n \leq d$) attributes of the data resembling the $(n+1)$ dimensional probabilistic distributions [7], [8] (see Fig. 6(b)). The difference of the proposed εFS s from probability distributions is that εFS s have peaks (maxima) at which $\mu = 1$ and they can be linguistically interpreted as FSs, e.g. “Low”, “Medium”, “High”, etc., per variable based on projections or as “close to *prototype_i*” in AnYa.

The proposed εFRB systems do not need to assume that the εMF s are in the form of continuous functions like the two tradition types of FRB systems in Fig. 3. The εMF s derived from data are in the discrete domain by their nature [7], [8]. However, if a certain variable is in the continuous domain based on common knowledge or *prior*

assumption, its ϵ MF can be in continuous domain as well. The transition from the discrete domain to the continuous one is only determined by the type of variables.

For example, based on the measured data, we can only derive the discrete ϵ MFs from the three data clouds shown in Fig. 4. However, considering that weights and lengths are from a continuous domain based on common knowledge, the continuous ϵ MFs can also be derived. The ϵ FRB systems with discrete and continuous ϵ MFs per feature are presented in Fig. 5(a) and (b) respectively.

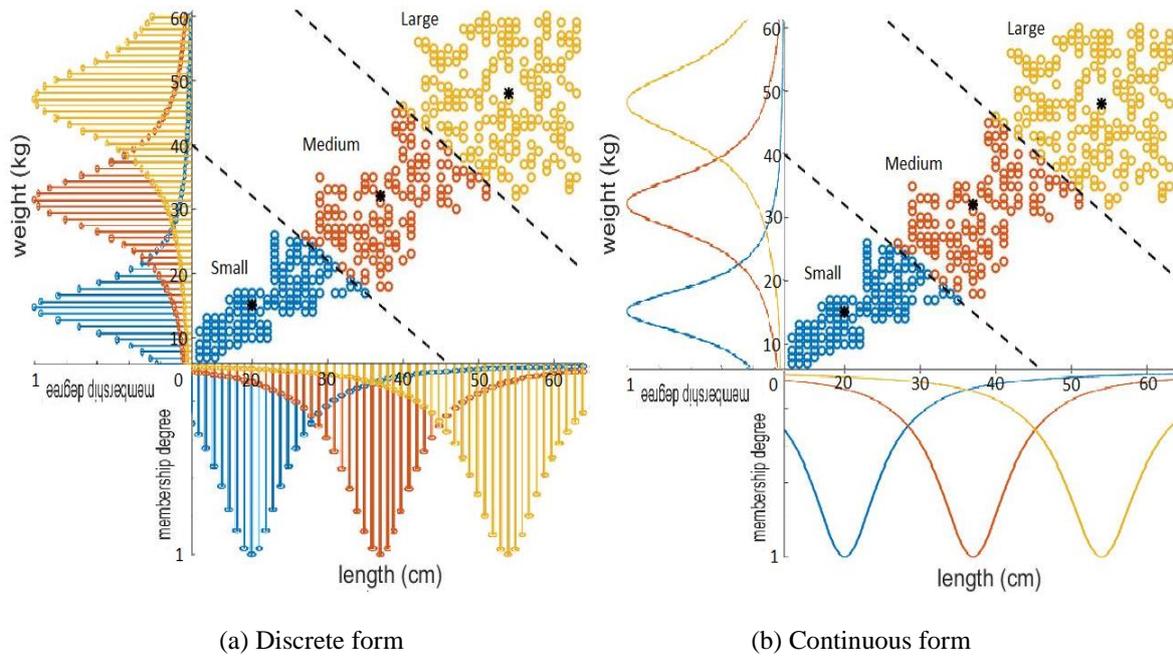


Fig.5. The examples of discrete and continuous ϵ FRB system

Additionally, the $(n+1)$ dimensional ϵ MFs can be derived from either discrete variables or continuous variables or a combination of discrete and continuous variables. The $(n+1)$ dimensional discrete and continuous ϵ MFs of the same example given in Fig. 5 are shown in Fig. 6 (a) and (b) ($n=2$). If we consider the variable “Length” to be discrete and the variable “Weight” to be continuous, the $3D$ ϵ MF derived from both continuous and discrete variables is depicted in Fig. 6(c).

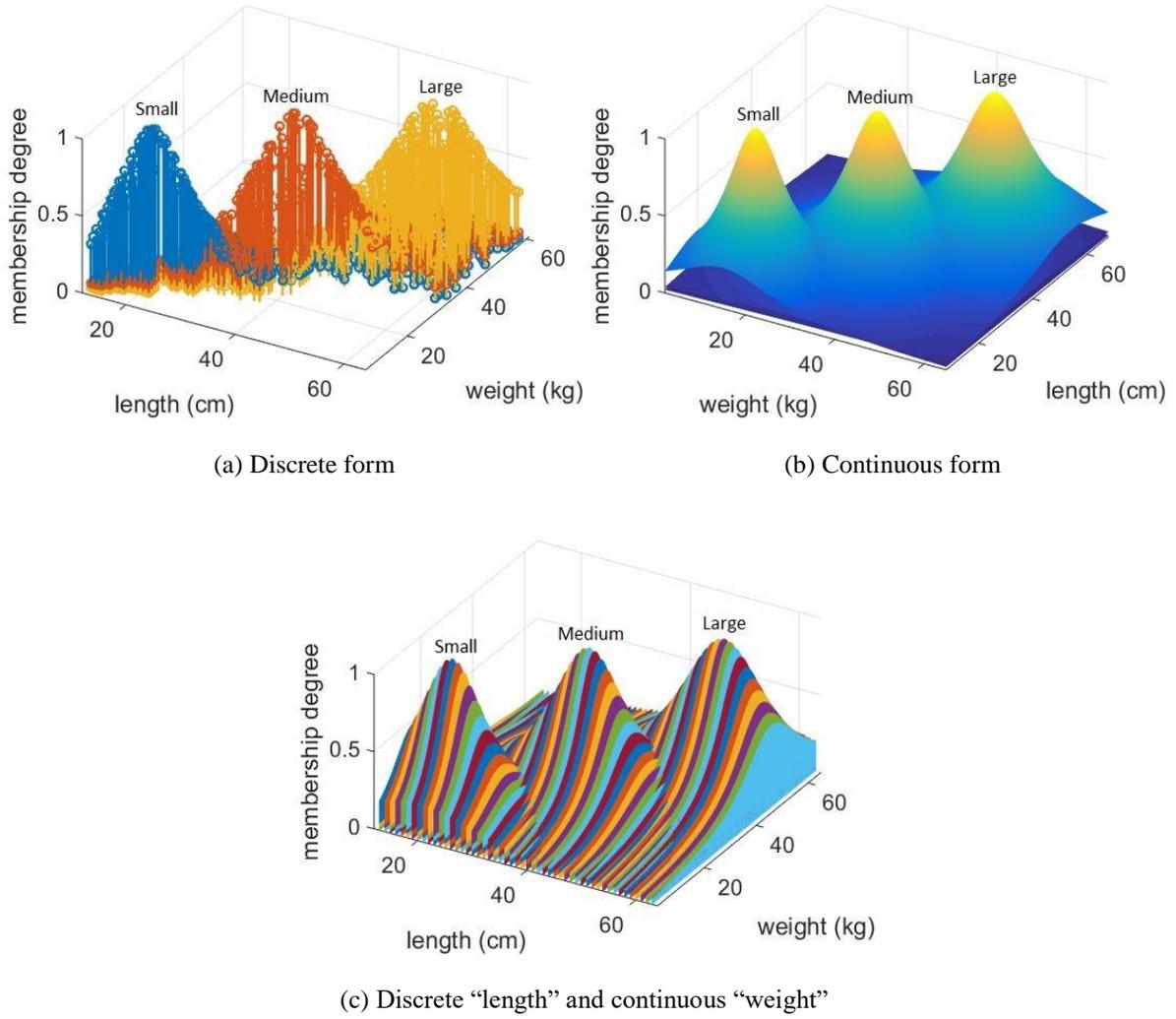


Fig.6. Visualization of 3D ϵ MFs

Similar to the AnYa type FRB system, the proposed ϵ FRB system is much more convenient and computationally simpler in high dimensional problems and it is unique in its ability to deal with problems containing categorical variables. This is thanks to the fact that only the prototypes are needed to be identified for the ϵ FRB system (identified either by users or by the data-driven approach [11]), and the system will derive ϵ MFs from data clouds formed around the prototypes automatically.

For a dataset containing A categories and d dimensional continuous and/or discrete variables, the experts need to define minimum AT^d MFs in order to design a highly descriptive Mamdani type or Takagi-Sugeno type FRB system, where T is the number of linguistic terms. However, to build an ϵ FRB system with the same degree of descriptiveness, the experts needs to select T prototypes for each category (AT prototypes in total), which is a huge improvement. For the above example, $A = 3$; $T = 3$; $d = 2$, but in realistic problems, d may be much

higher (dozens, hundreds or more), A and T can also be larger. Therefore, the improvement is in orders of magnitude.

In the following two subsections, we will describe two approaches to identify the prototypes for the ϵ FRB systems. The first one is using the newly introduced approach for forming data clouds [11], which is a nonparametric, entirely data-driven and objective method; the second one is based on human expertise.

3.2. Objective ϵ FRB System Identification

In this subsection, we will describe the objective approach within the EDA framework for identifying the prototypes for the ϵ FRB systems. The main procedure can be performed using the method for automatic formation of data clouds:

Step1: The *multimodal densities* D^{MM} of all the data samples $\{\mathbf{x}\}_K$ are calculated using equation (5).

For the specific example considered above, the *multimodal densities* D^{MM} of the size of the 600 domestic dogs are depicted in Fig. 7 (a).

Step2: Find the unique data sample \mathbf{u}_1^* with the maximum *multimodal density* $D^{MM}(\mathbf{u}_1^*)$.

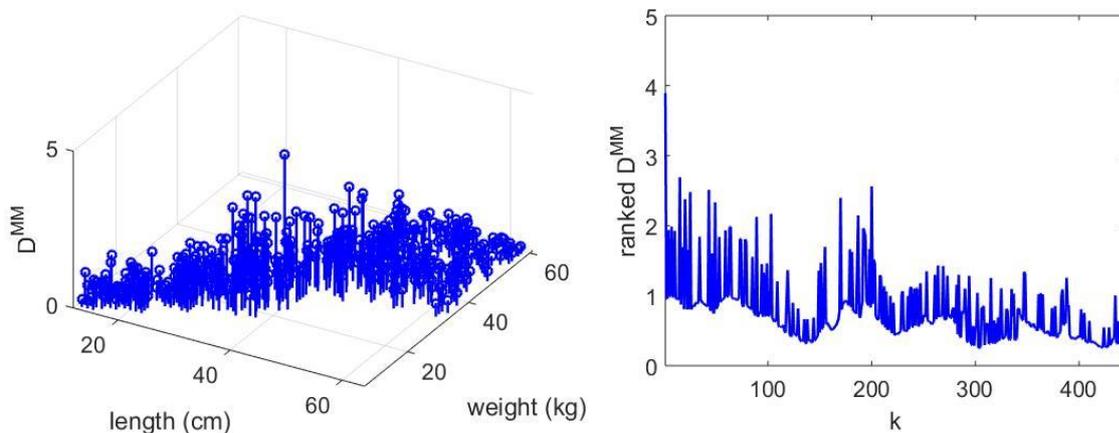
Step 3: Remove \mathbf{u}_1^* from $\{\mathbf{u}\}_L$ and put it into $\{\mathbf{u}\}_L^*$, then set \mathbf{u}_1^* as \mathbf{u}^R .

Step 4: Find the nearest unique data sample \mathbf{u}_2^* to \mathbf{u}^R , remove \mathbf{u}_2^* from $\{\mathbf{u}\}_L$ and send \mathbf{u}_2^* to $\{\mathbf{u}\}_L^*$.

Step 5: Use \mathbf{u}_2^* as the new \mathbf{u}^R and repeat Step 4 until $\{\mathbf{u}\}_L$ become empty.

Step 6: Rank the D^{MM} of $\{\mathbf{u}\}_L^*$ according to their indexes from 1 to L . The ranked D^{MM} are depicted in Fig.

7(b).



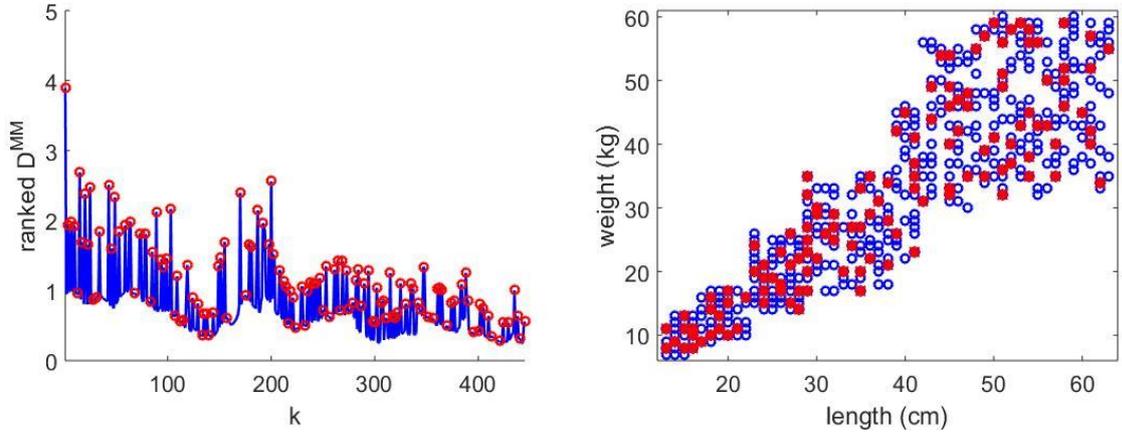
(a) Original D^{MM}

(b) Ranked D^{MM}

Fig.7. The *multimodal densities* D^{MM} for the illustrative example

Step 7: Find the local maxima of the ranked D^{MM} and use the corresponding unique data samples as prototypes,

$\{\mathbf{p}\}$. The local maxima of the D^{MM} are depicted in Fig.8.



(a) Local maxima identified from the ranked D^{MM} (b) Data samples with the local maximum D^{MM}

Fig.8. The local maxima of the ranked D^{MM} for the illustrative example

Step 8: Form data clouds from $\{\mathbf{x}\}_k$ with $\{\mathbf{p}\}$ using equation (11):

$$cloud\ label_i = \arg \min_{y \in \{\mathbf{p}\}} (\|\mathbf{x}_i - y\|); \mathbf{x}_i \in \{\mathbf{x}\}_k \quad (11)$$

Step 9: Obtain the centres $\{\mathbf{p}\}_0$ from the data cloud.

Step 10: Calculate the multimodal densities D^{MM} of $\{\mathbf{p}\}_0$ using equation (5).

Step 11: Find out the centres satisfying the following condition and denote them as $\{\mathbf{p}\}_1$:

$$IF \left(D^{MM}(\mathbf{p}_i) = \max \left(\left\{ \left\{ D^{MM}(\mathbf{q}) \mid \mathbf{q} \in \{\mathbf{p}\}_i^N \right\}, D^{MM}(\mathbf{p}_i) \right\} \right) \right) \\ THEN \left(\mathbf{p}_i \text{ is a member of } \{\mathbf{p}\}_1 \right) \quad (12)$$

where $\mathbf{p}_i \in \{\mathbf{p}\}_0$; $\{\mathbf{p}\}_i^N$ is the collection of data clouds whose centres are neighbouring to \mathbf{p}_i :

$$IF \left(\|\mathbf{p}_j - \mathbf{p}_i\| \leq R \right) THEN \left(\mathbf{p}_j \in \{\mathbf{p}\}_i^N \right) \quad (13)$$

here $\mathbf{p}_j \in \{\mathbf{p}\}_0$; $R = \sigma \left(1 - \frac{\sigma}{\eta} \right)$; η is the average Euclidean distance between any pair of centres; σ is the corresponding standard deviation of the distances between the centres.

Step 12: Set $\{\mathbf{p}\}_1$ as $\{\mathbf{p}\}$.

Step 13: Repeat Step 8 – Step 12 until $\{\mathbf{p}\}$ is not changing any more.

Step 14: Form data clouds from $\{\mathbf{x}\}_K$ using $\{\mathbf{p}\}$.

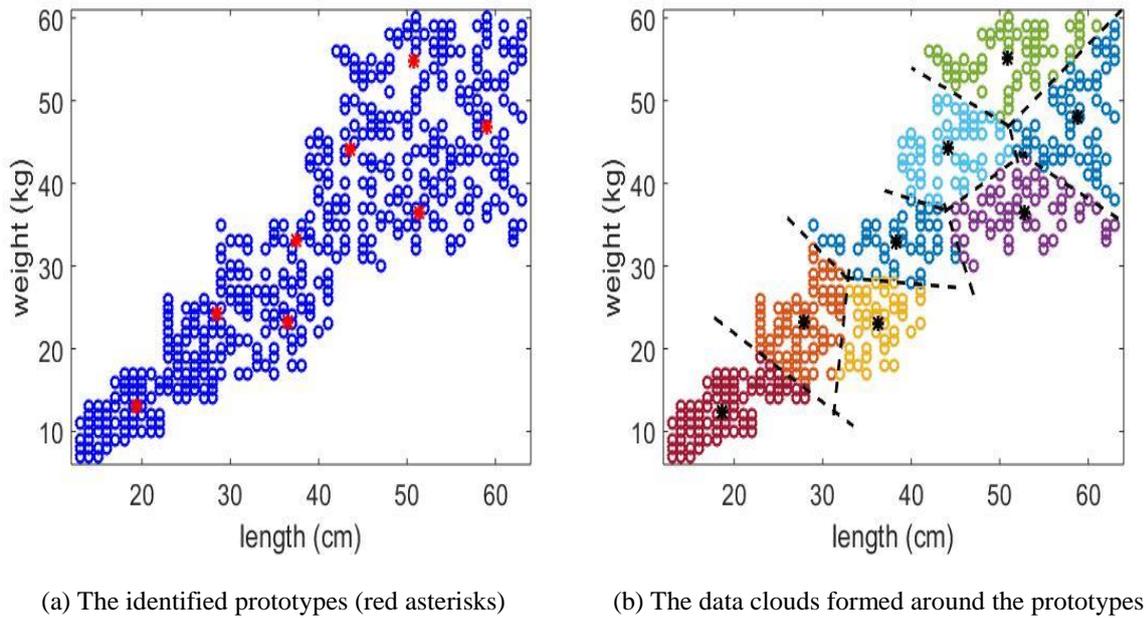


Fig. 9. The final results

If now we consider the example used several times earlier, there are eight prototypes identified from the data and eight data clouds are formed around them. Based on the eight data clouds, eight ϵ MFs are built, the ϵ MFs are also depicted in Fig. 10 in a 3D form.

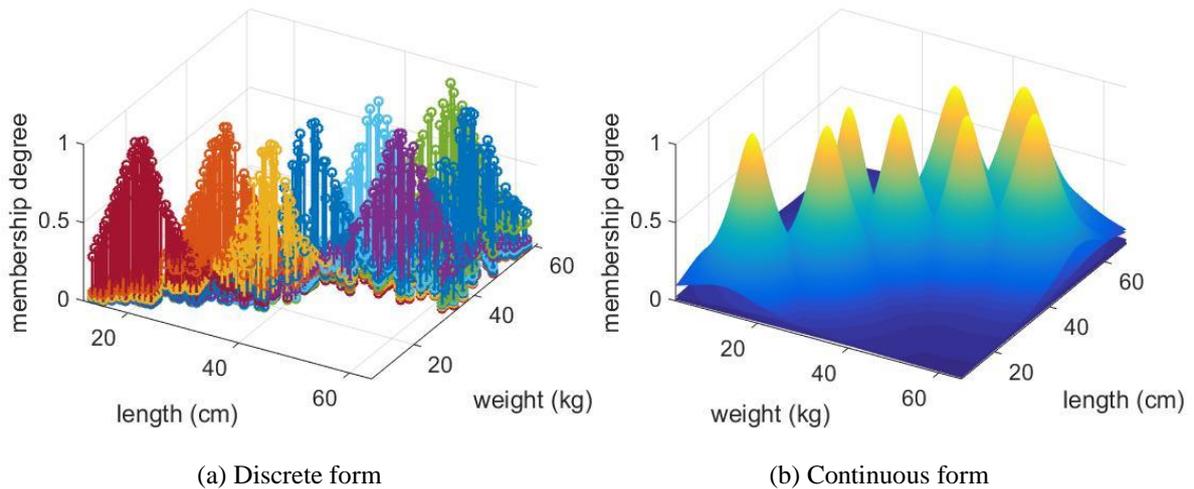


Fig. 10. Visualization of 3D continuous ϵ MFs

As it has been stated in section 3.1, when the data contains categorical variables, the ϵ FRB system requires at least one prototype for each category. Therefore, the data will be split per category and processed separately. This is very different from the traditional approaches, which ignore the real differences between categorical variables. This is, however, very convenient for parallelization.

The common practice for the traditional machine learning approaches to process categorical variables is to map them to different integer numbers. For example, one may use digit “1” to represent job category “worker”, “2” to represent job category “teacher”, “3” to represent “policeman”, etc. Alternatively, one can use the *1-of-C* encoding method [12] to map the categorical variables into a series of orthogonal binary variables like using “001” to represent job category “worker”, “010” to represent “teacher”, “100” to represent “policeman”, etc. However, no matter what kind of mapping is used, the encoding process always minimises the true differences between the data from different categories. This minimization is more obvious in high dimensional problems. In many cases, data from different categories are inconsistent and, in fact, incomparable. The best way for handling different categories is to process them separately and thus, avoid the interferences between each other. Therefore, in the proposed approach, if the data contains A categories, the data is divided into A groups based on their categories and used to form data clouds separately [11]. To be more specific, let us use the real climate dataset (temperature and wind speed) measured in Manchester, UK for the period 2010-2015 [13] for illustration. This dataset contains 479 samples obtained during the winter and 459 samples during the summer. As the dataset contains data samples from two categories (“winter” and “summer”), we firstly separate the two categories and then, form the data clouds using the technique described in [11] to find the prototypes from each category. As we can see from Fig.11, there are 21 prototypes identified from the 479 “winter” data samples and 24 prototypes identified from the 459 “summer” data samples.

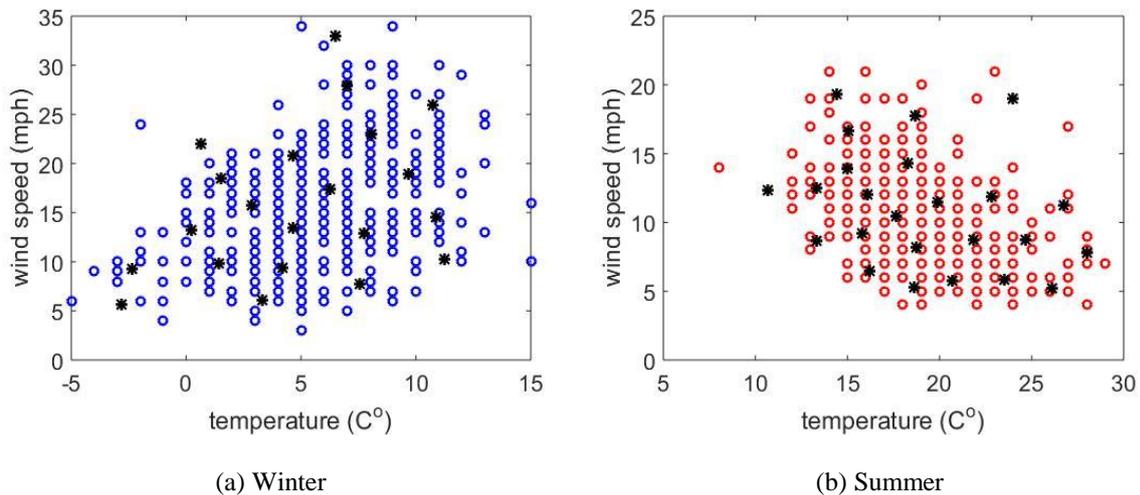


Fig.11. Prototypes identified from the data samples from the two categories (the black asterisks are the prototypes)

Using the identified prototypes, 45 data clouds in total for both categories resembling Voronoi tessellation [11] are automatically formed around these prototypes. After deriving ϵ MFs from these data clouds, the structure of

the ϵ FRB system is identified based on the ϵ FRs built upon the data clouds. The 3D visualization of the ϵ MFs derived from data is depicted in Fig. 12. We also tabulate the ϵ FRs in Table I for a better illustration.

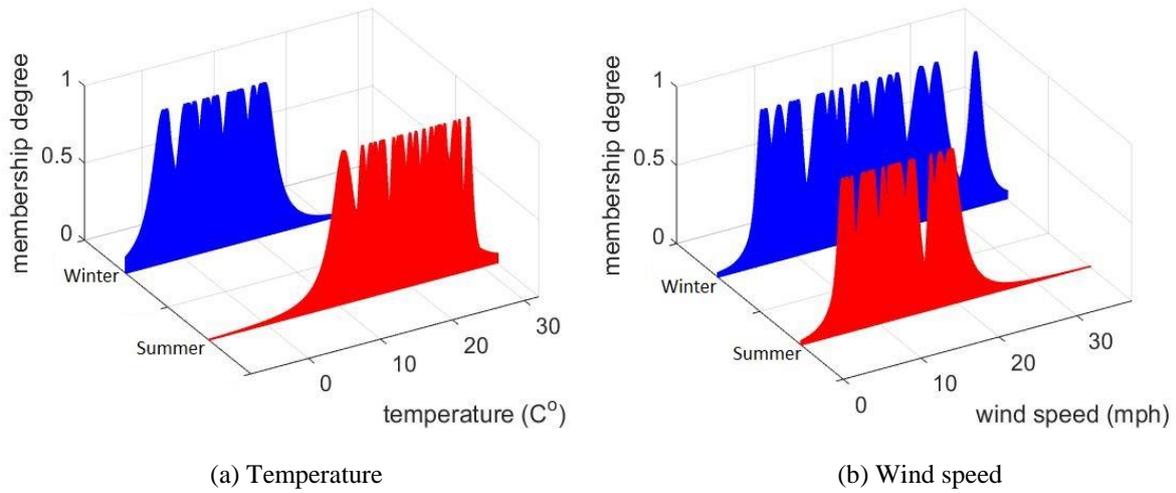


Fig.12. 3D visualization of the ϵ MFs

Table I. ϵ FRs derived automatically from the real climate dataset

Class	ϵ FR
Winter	$IF (x \sim [6.2453 \ 17.4340])$
	$OR (x \sim [0.6667 \ 22.0000])$
	\vdots
	$OR (x \sim [3.3333 \ 6.0833])$
	$THEN (Winter)$
Summer	$IF (x \sim [19.8776 \ 11.5102])$
	$OR (x \sim [28.0000 \ 7.8333])$
	\vdots
	$OR (x \sim [16.1136 \ 12.0682])$
	$THEN (Summer)$

3.3. Subjective ϵ FRB System Identification

As it was stated in section 1, ϵ FRB systems allow the subjective specifics that FSs and FRB systems are strong with to be easier incorporated and formalized. Instead of handcrafting a Mamdani type [3] or a Takagi-Sugeno type [4] FRB systems with significant efforts, experts will only need to select a number of most typical samples as prototypes to build ϵ FSs around them. The ϵ MFs will then be generated automatically from the data. Optionally, the human experts can also help to define the labels/names of the linguistic terms, classes (if any). In this subsection, we will use real climate dataset [13] to show how to build an ϵ FRB system based on human expertise.

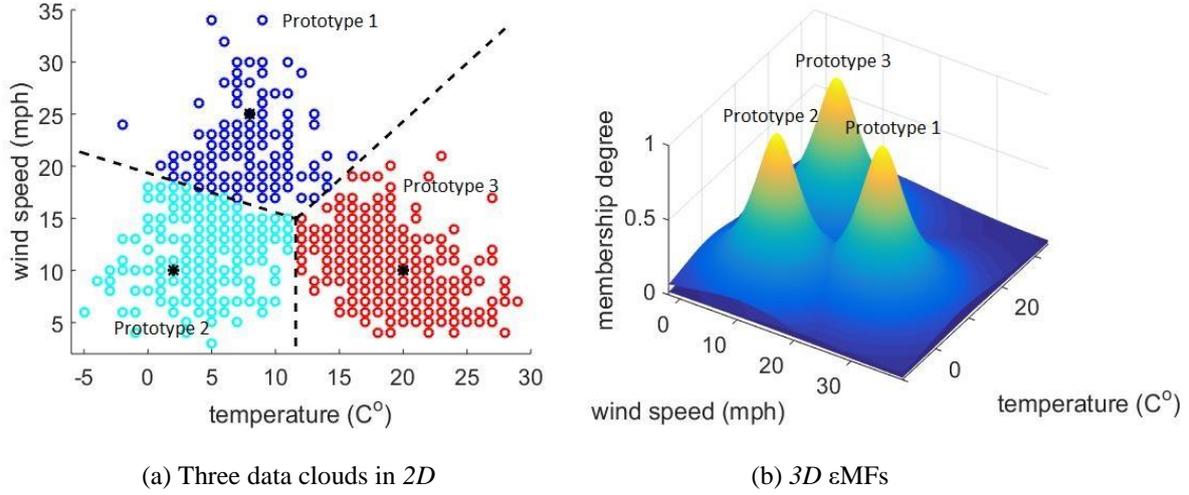


Fig.13. The ϵ FRB system formed with the subjective approach

As it was mentioned in section 3.2 the real climate dataset contains two categories “winter” and “summer”. In order to build a highly descriptive ϵ FRB system, for each category, one needs to select minimum one prototype in order to form at least one data cloud. For example, if we select two typical data samples $prototype_1 = [8C^\circ, 25mph]$ and $prototype_2 = [2C^\circ, 10mph]$ measured in winter as the prototypes of the “winter” category, and one typical data sample $prototype_3 = [20C^\circ, 11mph]$ measured in summer as the prototype of the “summer” category. In this way, three data clouds are formed around the selected prototypes by the data samples associated with each one of these prototypes. They form Voronoi tessellation [11] and ϵ FSs around these prototypes. The ϵ FRB system with the three prototypes is visualized in a 3D form in Fig. 13.

As one can see, compared with defining the linguistic terms, prototypes, MFs, etc., one by one, the ϵ FRB system only requires the prototypes to be defined, which is much simpler and easier. Instead of building mathematical models and handcrafting the whole FRB system piece by piece, the human experts/users only need to select few typical data samples as prototypes, and then the proposed method can autonomously build the ϵ FRB system based on these prototypes. In this case, the prototypes have a clear meaning that: $prototype_1$ -cool and windy day; $prototype_2$ -cold and quiet day; $prototype_3$ -warm and quiet day. The simplification in terms of human involvement of the proposed approach can play a very important role in the collaboration between computer scientists and experts from different areas.

Alternatively, after the ϵ FRB system has been identified via the objective approach as described in section 3.2, the experts can get involved to further improve the descriptiveness of the system, which also saves a lot of efforts and time compared with handcrafting as in the traditional FRB system.

The convenience of the proposed approach may significantly influence the recommendation systems used by retailers. Let us use an example of buying a house. Of course, there are many visible and hidden factors to be considered before buying a new house, i.e. price, the distances to the city centre, schools and main roads, the environment, the safety conditions, the neighbourhood, house floor area, etc. To simplify this problem, we only consider four visible factors/features, *i*) price, *ii*) house floor area, *iii*) distance to the city centre and *iv*) distance to the schools.

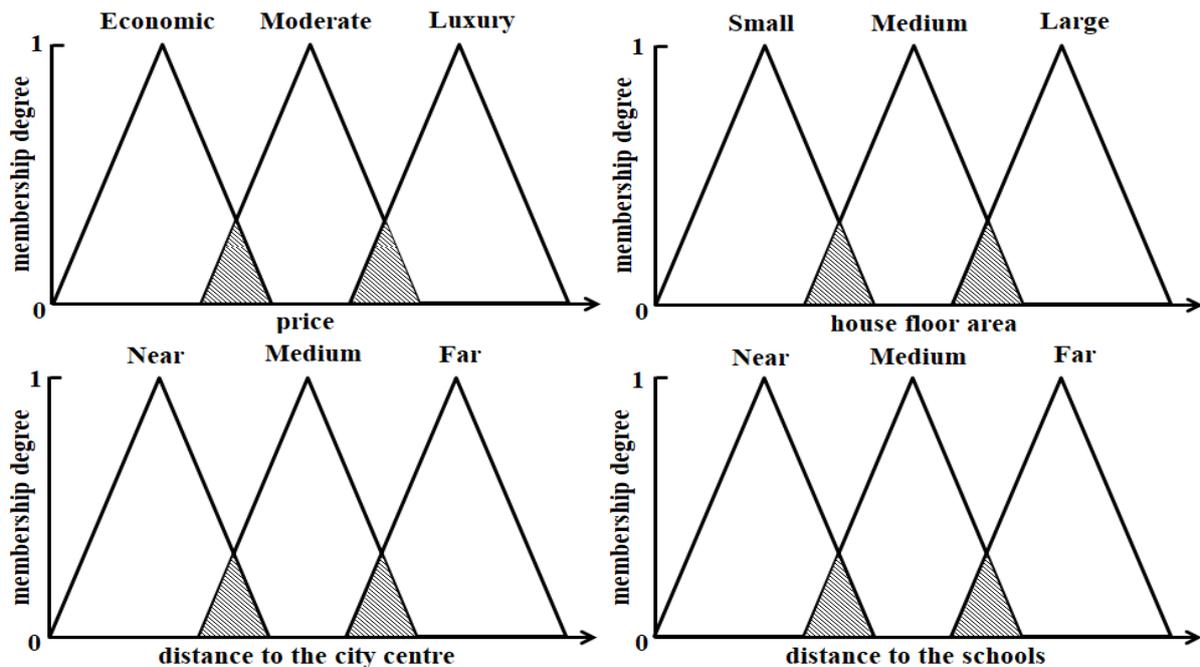


Fig.14. The triangular type MFs of the traditional FRB system

If the estate agency wants to build a recommendation system using traditional FRB systems, the estate agents need to build a number of FSs to categorize the houses based on different features, i.e. prices which may be “economic”, “moderate” and “luxury”, distance to city centre which may be “near”, “medium” and “far”, etc. (presented in Fig. 14). Building these FSs requires a lot of efforts and is subjective (different for each user). Different estate agents as well as different customers may have different perceptions of the four features. For example, an elderly customer may think that a house 1 km away from the city centre is at a *medium* distance. However, a young customer may think that such a house requiring only 10 minutes’ walk to the city centre can be perceived as being *close/near*. Alternatively, a single customer may think that a 70 m² house is “big”, but a couple with five children may think that this house is too “small”. Moreover, the preference may not be smoothly monotonic, it is not clear how many linguistic terms to use each time (three or more or less). In short, the handcrafted FRB systems are difficult to design and use. This may be the main reason that they are still not widely accepted.

In contrast, when the ϵ FRB system is used instead, the estate agent only needs to ask customers to select one or more houses they are most satisfied with. These houses can be any real houses in this city regardless whether they are for sale or not. These may also be imaginary, ideal houses as well. A family with five children may select one house that has large area and is very close to the schools, not far from the city centre and not expensive. A retired elderly couple may select a medium size house with luxury decoration and far away from the city centre and schools. A young couple may select a small economic house close to the city centre.

Then, the selected houses can be used as prototypes to form the data clouds based on the four normalized variables from all the available for sale houses in the database. The ϵ MFs derived from the data clouds formed around the prototypes are visualized in Fig. 15.

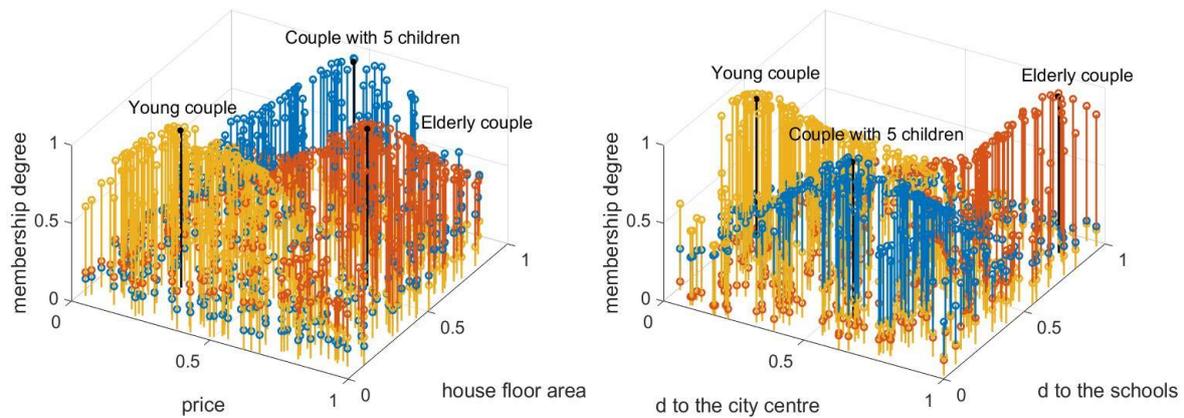


Fig. 15. Visualization of the ϵ MFs based on four normalized attributes of houses
(black asterisks represent the prototypes).

Based on the degrees of similarity of each available for sale house to the prototypes, the estate agent can easily make a list of recommended houses for each couple. All of this is achieved by asking the couples a simple question: “Can you, please, tell me the most satisfactory house in the city you have seen?” Similarly, each customer could also rank order few preferred houses, e.g. *i*) best; *ii*) good; *iii*) definitely no, etc.

As we can see from the above example, there is no need for any deliberately defined parameters or unnecessary efforts, the ϵ FRB recommendation system only needs the users to give some examples of whatever they think are the best as the prototypes, which can be real or virtual. Then the system will automatically form the data clouds based on the prototypes and calculate the degrees of membership for all the available products. The recommendation list can then be generated based on the degrees of membership. This process is user-specific, but at the same time, is also objective and data-driven.

4. Empirical Fuzzy Classifier

In this section, we will describe a new type of classifier based on the ε FRs, named empirical fuzzy (ε F) classifier and conduct numerical experiments to demonstrate the performance of the proposed classifier. The proposed ε F classifier is very close to the concept of Naïve Bayes classifiers which perform classification based on the dominant per class likelihood expressed by a pre-defined (usually, Gaussian) *pdf*. It performs the classification based on the degree of membership of the ε FRs (equation (9)) following the well-known “winner takes all” principle. However, other principles i.e. “few winners take all”, “fuzzily weighted”, “average” can also be considered depending on the specific problem.

As ε FRs can be derived by both, the objective and subjective approaches, without losing generality, we use the objective approach as being described in section 3.2 for the ε FRB system identification. Assuming that there have been N ε FRs automatically derived from data using the technique for forming data clouds [11]. When applied to a new, unlabelled, unseen sample \mathbf{x} , its label is given as:

$$class\ label = \arg \max_{i=1,2,\dots,N} (\varepsilon MF_i(\mathbf{x})) \quad (14)$$

That is, the class label of \mathbf{x} is decided by the label of the prototype that has higher membership degree.

To evaluate the performance of the proposed approach, three numerical examples based on benchmark datasets are conducted in this paper. The following algorithms were used in the comparison:

- i) SVM classifier with Gaussian kernels [14], [15];
- ii) SVM classifier with 4th Order Polynomial kernel [15];
- iii) Naïve Bayes classifier [16];
- iv) Decision tree classifier [17];
- v) eClass0 classifier [18].

The comparison is based on the following criteria:

- i) Confusion matrix of the classification results;
- ii) Average accuracy after 10 times Monte Carlo experiments;
- iii) Average training time after 10 times Monte Carlo experiments (in seconds).

4.1. Wine Dataset [19]

The wine dataset came from a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis contains 178 wine samples of three classes (class 1 consists of 59 samples, class2 consists of 71 samples and class 3 consists of 48 samples) with 13 attributes: 1) Alcohol; 2) Malic acid;

3) Ash; 4) Alkalinity of ash; 5) Magnesium; 6) Total phenols; 7) Flavonoids; 8) Neoflavanoid phenols; 9) Proanthocyanins; 10) Colour intensity; 11) Hue; 12) OD280/OD315of diluted wines; 13) Proline. Due to the high dimensionality, all the data samples are normalized by their norms in advance:

$$x \leftarrow \frac{x}{\|x\|} \quad (15)$$

We use the first 60% of the data samples of each class as the training set and use the rest of the dataset as the validation set. The classification results are tabulated in Table. II. For a better illustration, the ϵ MFs of the first 4 attributes derived from the training samples are visualized in a 3D form per type per feature in Fig. 16.

Table. II Classification results on the wine dataset

Method	True Class	Predicted			Accuracy	Method	True Class	Predicted			Accuracy
		1	2	3				1	2	3	
ϵF^a	1	20	4	0	0.7324	Naïve Bayes	1	21	3	0	0.9437
	2	1	22	5			2	0	27	1	
	3	3	6	10			3	0	0	19	
SVM-G ^b	1	0	24	0	0.3944	Decision tree	1	24	0	0	0.9718
	2	0	28	0			2	2	26	0	
	3	0	19	0			3	0	0	19	
SVM-P ^c	1	18	6	0	0.6479	eClass0	1	21	1	2	0.6479
	2	0	28	0			2	1	12	15	
	3	0	19	0			3	2	4	13	

^a the proposed ϵF classifier ^b the SVM classifier with Gaussian kernel; ^c the SVM classifier with polynomial kernel

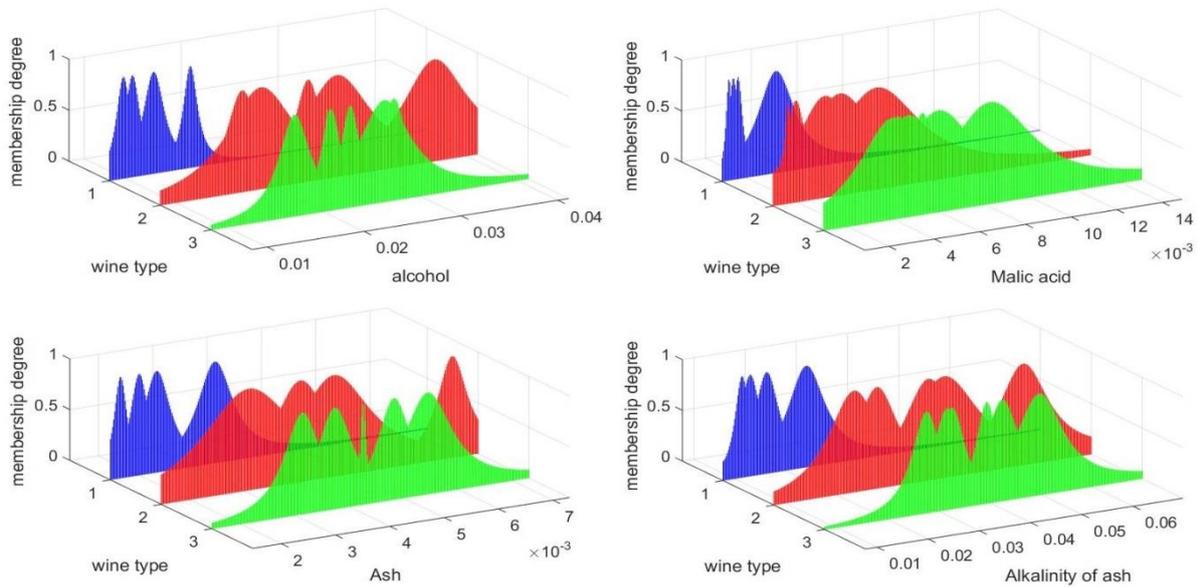


Fig. 16. The 3D visualization of the ϵ MFs for the first 4 attributes of the wine dataset

Then, we conduct 10 Monte Carlo experiments by training the classifiers with randomly selected 60% of the data samples of each class and using the rest for validation. The average overall accuracies of the classification results obtained by the six classifiers are depicted in Fig. 17 and the corresponding average amounts of time consumption (in seconds) for training are presented in Fig. 18.

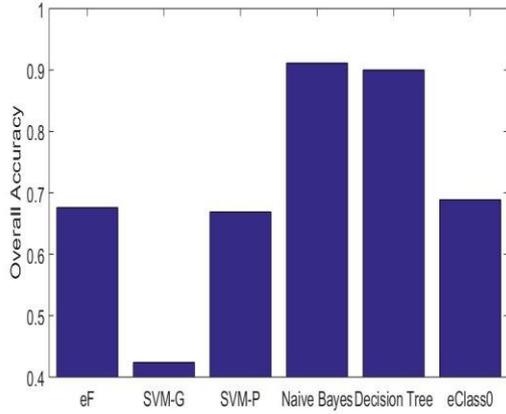


Fig.17. Overall accuracy

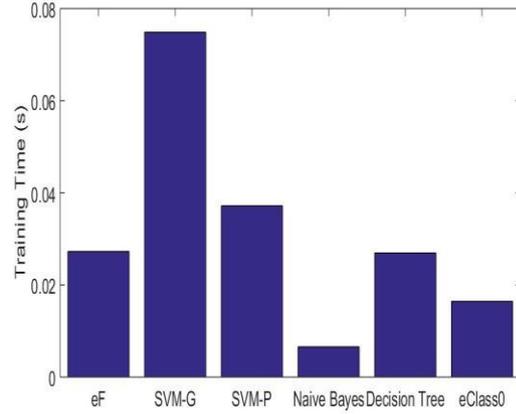


Fig.18. Average training time (in seconds)

4.2. Banknote Authentication Dataset [20]

Banknote authentication dataset was extracted from images that were taken from genuine and forged banknote-like specimens. Wavelet Transform tool was used to extract features from the images [21]. This dataset contains 1372 samples and each sample has four attributes:

- 1) variance of the wavelet transformed image;
- 2) skewness of the wavelet transformed image;
- 3) curtosis of the wavelet transformed image;
- 4) entropy of the image.

and one label: class (0 or 1). 762 of the data samples are in class 0 and 610 samples are in class 1.

We use the first 60% of the data samples of each class (152 samples from class 0 and 122 samples from class 1) as our training set and use the rest of the dataset as the validation set. The classification results obtained by the six classifiers are presented in Table III. The ϵ FRs derived from the training data are presented as examples in Table IV for a better interpretability.

10 Monte Carlo experiments are also conducted on this dataset for further comparison. The average overall accuracies of the six classifiers are depicted in Fig. 19 and the corresponding average time consumption (in seconds) required for the training is presented in Fig. 20.

Table. III Classification results on the banknote authentication dataset

Method	True Class	Predicted		Accuracy	Method	True Class	Predicted		Accuracy
		0	1				0	1	
ϵ F	0	302	3	0.9945	Naive Bayes	0	278	21	0.8725
	1	0	244			1	43	201	
SVM-G	0	305	0	1.0000	Decision tree	0	302	3	0.9891
	1	0	244			1	2	241	
SVM-P	0	239	66	0.8561	eClass0	0	261	44	0.9144

	1	13	231			1	3	241	
--	---	----	-----	--	--	---	---	-----	--

Table. IV The ϵ FRs derived from the banknote authentication dataset

Class	ϵ FR
0	$IF (x \sim [3.1229 \ -4.5316 \ 4.4032 \ -0.1095])$
	$OR (x \sim [3.1019 \ 2.9954 \ 0.7296 \ 0.3001])$
	$OR (x \sim [0.7350 \ 5.8760 \ 1.5374 \ -0.2825])$
	⋮
	$OR (x \sim [1.7877 \ 7.2335 \ -0.3137 \ -2.3080])$
	$THEN (Class \ 0)$
1	$IF (x \sim [-1.2555 \ -0.8157 \ 1.1531 \ 0.5117])$
	$OR (x \sim [-1.4167 \ -6.0412 \ 5.5654 \ 0.5516])$
	$OR (x \sim [-1.9863 \ -8.9029 \ 8.8260 \ -0.3372])$
	⋮
	$OR (x \sim [-3.7914 \ -6.9420 \ 10.9737 \ -0.4656])$
	$THEN (Class \ 1)$

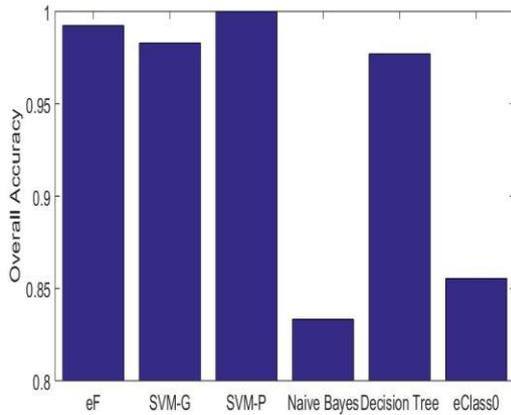


Fig.19. Overall accuracy

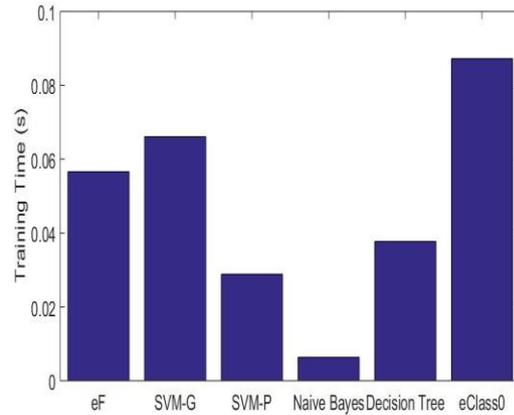


Fig.20. Average training time (in seconds)

4.3. Tic-Tac-Toe Endgame Dataset [22]

The Tic-Tac-Toe Endgame dataset contains a complete set of possible board configurations at the end of tic-tac-toe games. The target concept of this dataset is “winning for x”. This dataset contains 958 data samples with nine attributes and one class label [22]:

- 1) top-left-square: {x, o, b}
- 2) top-middle-square: {x, o, b}
- 3) top-right-square: {x, o, b}
- 4) middle-left-square: {x, o, b}

- 5) middle-middle-square: {x, o, b}
- 6) middle-right-square: {x, o, b}
- 7) bottom-left-square: {x, o, b}
- 8) bottom-middle-square: {x, o, b}
- 9) bottom-right-square: {x, o, b}
- 10) Class: {positive, negative}

In this experiment, we encode “x” as “1”, “o” as “5” and “b” as “3”. Obviously, all the variables are in the discrete domain. The data samples are normalized by their norms before classification using equation (15).

The dataset is divided into two parts. We use the first 60% samples of each class for training, and the rest of them for validation. The classification results of the proposed classifier and the five comparative classifiers are tabulated in Table V.

We also conduct 10 Monte Carlo experiments by randomly selecting 60% of the data samples of each class for training the classifiers and using the rest for validating the classifiers. The average overall accuracy and the time consumption (in seconds) of the training process of the six classifiers are depicted in Fig. 21 and Fig. 22, respectively.

Table. V. Classification results on the Tic-Tac-Toe endgame dataset

Method	True Class	Predicted		Accuracy
		positive	negative	
εF	positive	176	74	0.7937
	negative	5	128	
SVM-G	positive	176	74	0.6475
	negative	61	72	
SVM-P	positive	136	114	0.6658
	negative	14	119	
Naïve Bayes	positive	179	71	0.6292
	negative	71	62	
Decision tree	positive	250	0	0.7337
	negative	102	31	
eClass0	positive	183	67	0.6136
	negative	81	52	

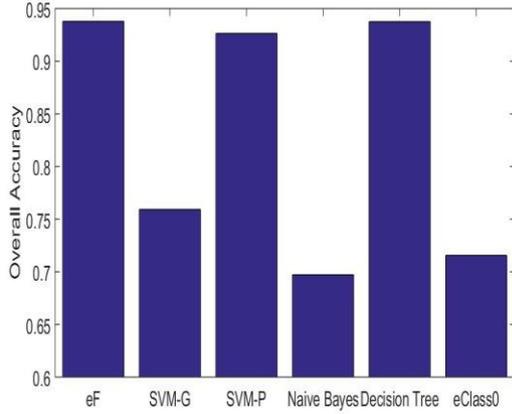


Fig.21. Overall accuracy

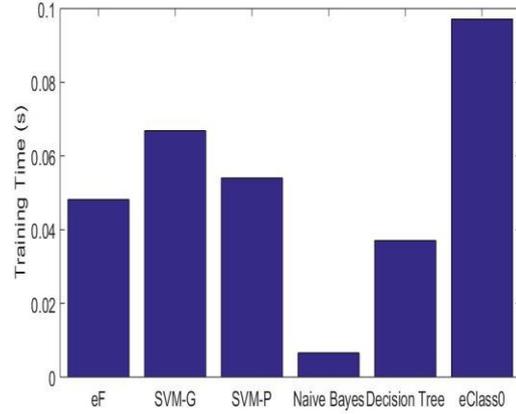


Fig.22. Average training time (in seconds)

4.4. Letter Recognition Dataset [23]

This dataset contains 20000 character images consisting of large numbers of black-and-white rectangular pixels displaying the 26 capital letters (from “A” to “Z”) of the Latin alphabet used in English language. Each image has been converted into 16 primitive numerical attributes as follows: 1) x-box: horizontal position of the box; 2) y-box: vertical position of the box; 3) width: width of the box; 4) height: height of the box; 5) onpix: total number of pixels that are “on”; 6) x-bar: mean value of x of the pixels that are “on” in the box; 7) y-bar: mean value of y of the pixels that are “on” in the box; 8) x2bar: mean value of the x variance; 9) y2bar: mean value of the y variance; 10) xybar: mean value of the x,y correlation; 11) x2ybr: mean value of the $x \cdot x \cdot y$; 12) xy2br: mean value of the $x \cdot y \cdot y$; 13) x-ege: mean edge count from left to right; 14) xegvy: correlation of x-ege with y; 15) y-ege: mean edge count from bottom to top; 16) yegvx: correlation of y-ege with x. The 16 attributes have been scaled to fit into a range of integer values from 0 through 15.

In this experiment, we, firstly, normalized the data samples by their norms before classification using equation (15). Then, the first 60% of the data samples of each class were used as a training set and the rest were used as a validation set. The confusion matrix of the result obtained by the proposed approach is depicted in Fig. 23. The ϵ MFs of the 3rd and 4th attributes (width and height), derived from the training samples are visualized in a 3D form per type per feature in Fig. 24 as an illustrative example.

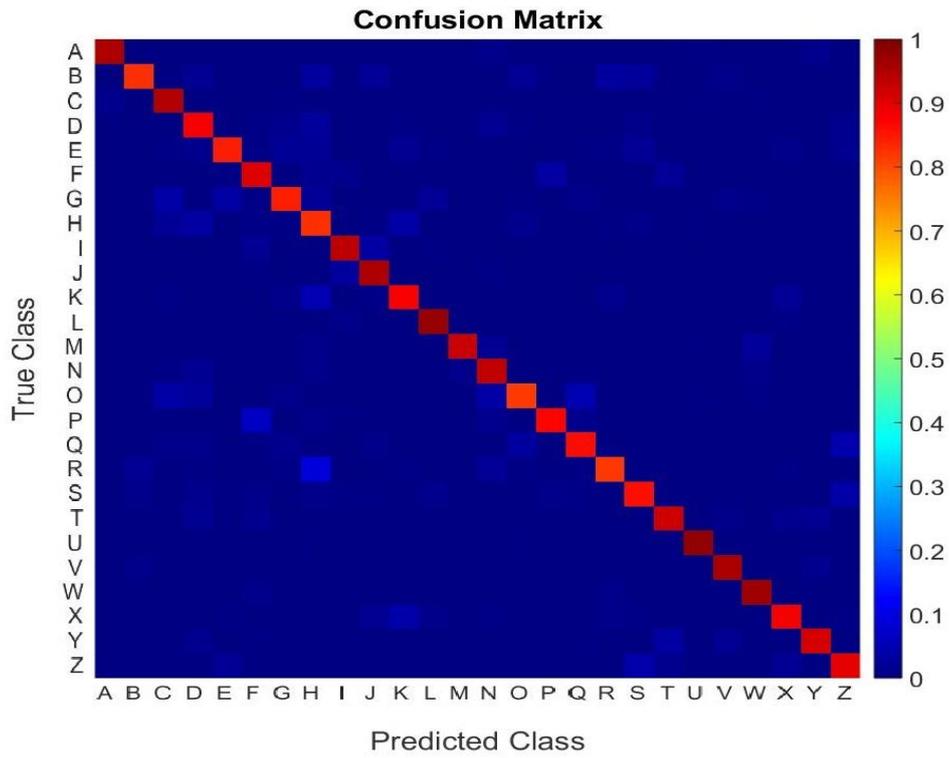
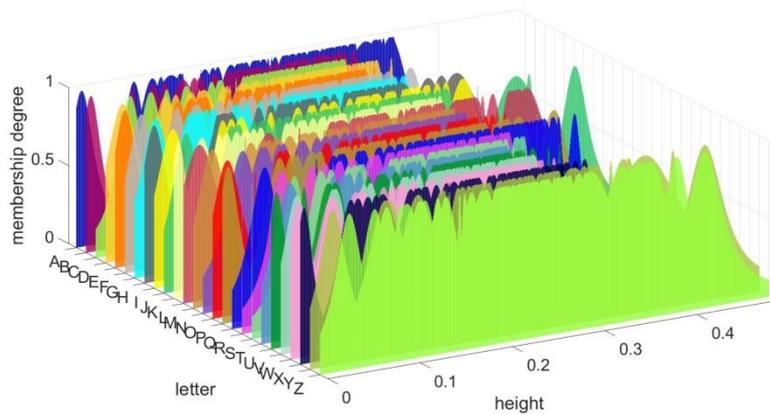
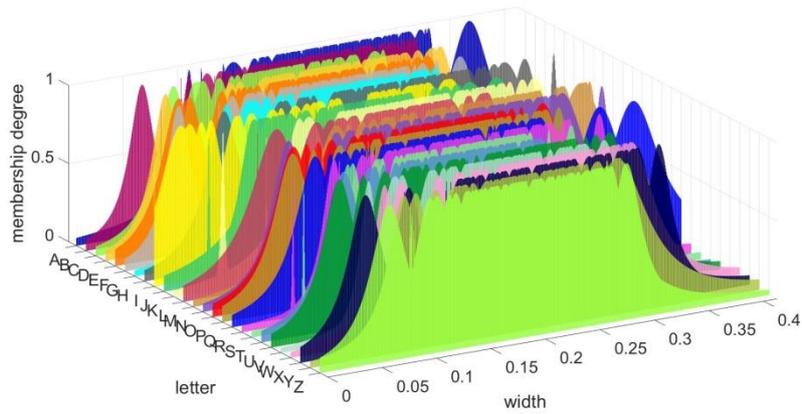


Fig. 23. The confusion matrix of the classification result of the letter recognition dataset



(b) ϵ MF of the 4th attribute- height

Fig. 24. The 3D visualization of the ϵ MFs of the letter recognition dataset

Same as the previous three benchmark problems, 10 Monte Carlo experiments are conducted by randomly selecting 60% of the data samples of each class for training the classifiers and using the rest for validating the classifiers. The average overall accuracy and the time consumption (in seconds) of the training process of the six classifiers are depicted in Fig. 25 and Fig. 26, respectively.

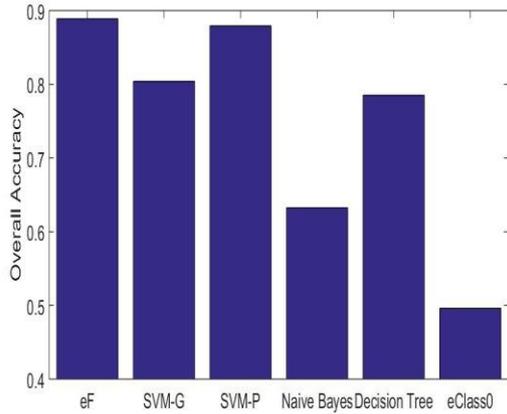


Fig.25. Overall accuracy

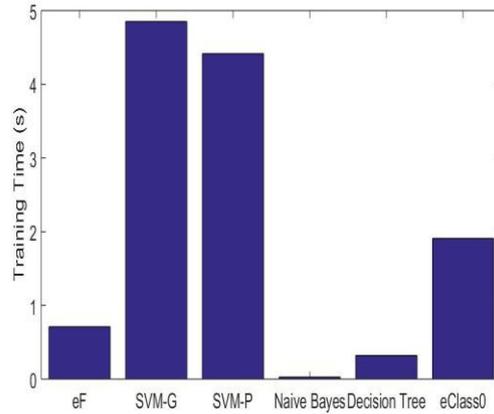


Fig.26. Average training time (in seconds)

4.5. Discussion

From the four numerical examples in subsection 4.1, 4.2, 4.3 and 4.4 we can see, that the proposed approach has a very good performance in the four benchmark classification problems. In addition, it has very high computational efficiency. In contrast, the SVM classifier with a Gaussian kernel [14], [15] shows quite good results in the complicated problems, i.e. letter recognition and banknote authentication problems, but its computational efficiency is very low. The SVM classifier with a Polynomial kernel [15] is comparable with the proposed approach in terms of accuracy, however, it requires more time for training. Unlike the SVM classifiers, the proposed ϵ F classifier has a very high transparency and clear interpretability. The naïve Bayes classifier [16] is the fastest one due to its simplicity and its performance is very high on the simpler problems, i.e. wine dataset. The classification accuracy of decision tree classifier [17] is lower than the proposed classifier in the complicated problems, but its training process is slightly faster. eClass0 classifier [18] is an FRB classifier. It is very efficient, but it only shows good result in the simplest problem (wine dataset). The most important advantages of ϵ FSs are their transparency and objectiveness as well as the ability of dealing with categorical data naturally and seamlessly (see Figs. 16 and 24).

5. Conclusion

In this paper, we introduce a new form of describing fuzzy sets, named ϵ FSs and a new form of FRB systems, named ϵ FRB grounded at the Empirical Data Analytics (EDA) framework. The proposed approach touches the fundamental question of how to build a FRB system. Two approaches (subjective and objective) for identifying ϵ FRB systems are described in this paper. Through a number of illustrative examples, we demonstrate that the proposed approach is a powerful alternative for scientists working with FRB systems in various fields and it has a strong potential.

Compared with the traditional FSs and FRB systems, the proposed approach has the following significant advantages:

- i) The ϵ FSs are derived in a transparent, data-driven way without *prior* assumptions
- ii) Effectively combines the data- and human- derived models;
- iii) It has very strong interpretability and high objectiveness;
- iv) The involvement of human experts is significantly facilitated and can be bypassed.

Numerical examples in this paper have demonstrated the high performance of the ϵ F classifier, but the applications of the proposed approach include, but are not limited to classification, control, prediction.

As a future work, we will detail the evolving ϵ FRB systems, predictors and apply it to various problems, i.e. high frequency trading, image classification, aircraft control, etc. We will also prove stability conditions for the ϵ FRB systems.

6. Acknowledgements

This work was partially supported by The Royal Society grant IE141329/2014 “Novel Machine Learning Paradigms to address Big Data Streams”.

Reference

- [1] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [2] P. Angelov, *Autonomous Learning Systems: From Data Streams to Knowledge in Real Time*. John Wiley, 2012.
- [3] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *Int. J. Man. Mach. Stud.*, vol. 7, no. 1, pp. 1–13, 1975.
- [4] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Trans. Syst. Man. Cybern.*, vol. 15, no. 1, pp. 116–132, 1985.
- [5] W. Pedrycz, “Fuzzy relational equations with generalized connectives and their applications,” *Fuzzy Sets Syst.*, vol. 10, no. 1–3, pp. 185–201, 1983.
- [6] P. Angelov and R. Yager, “A new type of simplified fuzzy rule-based system,” *Int. J. Gen. Syst.*, vol. 41, no. 2, pp. 163–185, 2011.
- [7] P. P. Angelov, X. Gu, J. Principe, and D. Kangin, “Empirical data analysis - a new tool for data analytics,” in *IEEE International Conference on Systems, Man, and Cybernetics*, 2016, pp. 53–59.
- [8] P. Angelov, X. Gu, and D. Kangin, “Empirical data analytics,” *Int. J. Intell. Syst.*, DOI

- 10.1002/int.21899, 2017.
- [9] C. C. Lee, "Fuzzy Logic in Control Systems : Fuzzy Logic Controller - Part 1," *IEEE Trans. Syst. Man Cybern.*, vol. 20, no. 2, pp. 404–418, 1990.
 - [10] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*, 2nd ed. Chichester, England: John Wiley & Sons., 1999.
 - [11] X. Gu, P. P. Angelov, and J. Principe, "Autonomous data partitioning," *under review*, 2017.
 - [12] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in *5th Annual Future Business Technology Conference*, 2008, pp. 5–12.
 - [13] "Climate Dataset in Manchester," <http://www.worldweatheronline.com>.
 - [14] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines : and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press, 2000.
 - [15] J. H. Min and Y. C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 603–614, 2005.
 - [16] C. M. Bishop, *Pattern Recognition*. New York: Springer, 2006.
 - [17] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man. Cybern.*, vol. 21, no. 3, pp. 660–674, 1990.
 - [18] P. Angelov and X. Zhou, "Evolving fuzzy-rule based classifiers from data streams," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1462–1474, 2008.
 - [19] "Wine Dataset," <https://archive.ics.uci.edu/ml/datasets/Wine>, Accessed on July 31st, 2017.
 - [20] "Banknote Authentication Dataset," <https://archive.ics.uci.edu/ml/datasets/banknote+authentication>, Accessed on July 31st, 2017.
 - [21] V. Lohweg, J. L. Hoffmann, H. Dörksen, R. Hildebrand, E. Gillich, J. Hofmann, and J. Schaede, "Banknote authentication with mobile devices," in *Proc. SPIE 8665, Media Watermarking, Security, and Forensics 2013*, 2013, pp. 866507–866507.
 - [22] "Tic-Tac-Toe Endgame Dataset," <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>, Accessed on July 31st, 2017.
 - [23] "Letter Recognition Dataset," <https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>, Accessed on July 31st, 2017.