2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Antecedent Selection in Fuzzy Rule Interpolation using Feature Selection Techniques

Ren Diao, Shangzhu Jin and Qiang Shen
Department of Computer Science
Institute of Mathematics, Physics and Computer Science
Aberystwyth University
Aberystwyth, SY23 3DB, UK
Email: {rrd09, shj1, qqs}@aber.ac.uk

***Abstract*—Fuzzy rule interpolation offers a useful means for enhancing the robustness of fuzzy models by making inference possible in systems of only a sparse rule base. However in practical applications, the rule bases provided may contain irrelevant, redundant, or even misleading antecedents, which makes the already challenging tasks such as inference and interpolation even more difficult. The majority of the techniques developed in the literature assumes equal significance of rules and their antecedents, which may lead to biased or incorrect reasoning outcomes. This paper investigates similar problems being tackled in the area of feature selection, in an attempt to identify techniques that can be applied to measure the significance of rule antecedents. In particular, two feature evaluation methods based on correlation analysis and fuzzy-rough set theory have been examined, in order to reveal their effectiveness in determining the importance of individual antecedents, and their capabilities for discovering subsets of antecedents that provide similar reasoning accuracies as a larger set of antecedents used in the original rules. In addition, the significance values measured by the proposed method are treated as the weights associated with the relevant rule antecedents, in an effort to facilitate more appropriate selection of, and interactions with the rules in performing both forward and backward fuzzy rule interpolation via scale and move transformation-based methods. Experimental studies based on a practical scenario concerning terrorist activities and also synthetic random data are conducted, demonstrating the potential and efficacy of the proposed work.**

I. INTRODUCTION

FUZZY-RULE-INTERPOLATION (FRI) [1], [2] is of particular significance for reasoning in the presence of insufficient knowledge or sparse rule bases. When a given observation has no overlap with antecedent values, no rule can be invoked in classical (fuzzy) rule-based inference, and therefore no consequence can be derived. However, the techniques of FRI can support inference in such situations. However, despite this potential, FRI techniques are relatively rarely applied in practice [3]. One of the main reasons is that real-world applications generally involve rules with a large number of antecedents, and the errors accumulated throughout the interpolation process may affect the accuracy of the final estimation. More importantly, a rule base may consist of less relevant, redundant or even misleading variables, which may further deviate the outcome of an interpolation. Several weighted FRI methods [4], [5] have been proposed in the literature, in order to remedy the loss of accuracy induced by the varying degrees of antecedent significance.

The irrelevancy and redundancy of data has been studied extensively in the area of feature selection (FS), and techniques have been developed to rank the importance of features [6], [7], [8], or to discover a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original data [9], [10], [11], [12]. A variety of feature subset search algorithms has also emerged, several of which are inspired by nature phenomena [13] or social behaviour [14], allowing quality feature subsets to be discovered without resorting to exhaustive search.

This paper presents a new approach that uses FS techniques to evaluate the importance of antecedent variables in a fuzzy rule base. Such importance degrees are referred to as the set of “antecedent significance values” hereafter. This allows subsets of informative antecedent variables to be identified via the use of feature subset search methods (e.g., harmony search-based feature selection, HSFS [14]). It helps to reduce the dimensionality of a rule base by removing irrelevant antecedent variables. An antecedent significance-based FRI technique based on scale and move transformation-based FRI (T-FRI) [15] is also proposed, which exploits the information analysed by FS, in an effort to facilitate more effective interpolation using weighted aggregation [16]. The benefits of this work are demonstrated through the process of backward FRI (B-FRI) [17], [18], [19], which is a newly identified research focus regarding FRI.

The remainder of this paper is structured as follows. Section II introduces the general ideas behind FRI, and explains the key notions and interpolation steps of T-FRI, which is the main method used to carry out the present investigation. This section also gives an outline of the B-FRI method for completeness. Section III details the developed approach which applies the existing ideas in FS to FRI, explains the antecedent significance-based aggregation procedure that is implemented using T-FRI, and discusses its potential benefits to B-FRI. In Section IV, an example scenario concerning the prediction of terrorist bombing attack is employed to showcase the procedures of the proposed work. Further, a series of simulation-based experiments have been carried out in order to verify the general performance of the present

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

approach. Finally, Section V summaries the paper and outlines several directions for future research.

II. Background of Fuzzy Rule Interpolation

This section first introduces the general principles of FRI, and provides a brief introduction of the procedures involved in T-FRI, including the definitions of its underlying key notions, and an outline of its interpolation steps. Note that the basic form of T-FRI is herein employed for the ease of presentation, which utilises two neighbouring rules of a given observation to perform interpolation. Triangular membership functions are also adopted for simplicity, which are the most commonly used fuzzy set representation in fuzzy systems. More detailed descriptions and discussions on the theoretical underpinnings behind T-FRI can be found in the original work [20].

A FRI system being investigated in this paper may be defined as a tuple $\langle R, Y\rangle$, where $R = \{r^1, \cdots, r^{|R|}\}$ is a non-empty set of finite fuzzy rules (the rule base), and Y is a non-empty finite set of variables. $Y = A \cup \{z\}$ where $A = \{a_j \mid j = 1, \cdots, |A|\}$ is the set of antecedent variables, and z is the consequent variable appearing in the rules. Without losing generality, a given rule $r^i \in R$ and an observation o^* are expressed in the following format:

$$\begin{aligned} r^i &: \text{if } a_1 \text{ is } \hat{a}_1^i \wedge \cdots \wedge a_j \text{ is } \hat{a}_j^i \wedge \cdots \wedge a_{|A|} \text{ is } \hat{a}_{|A|}^i, \\ &\quad \text{then } z \text{ is } \hat{z}^i \\ o^* &: a_1 \text{ is } \hat{a}_1^* \wedge \cdots \wedge a_j \text{ is } \hat{a}_j^* \wedge \cdots \wedge a_{|A|} \text{ is } \hat{a}_{|A|}^* \end{aligned}$$

where $\hat{a}_j^i$ represents the value (the fuzzy set) of the antecedent variable a_j in rule r^i, and $\hat{z}^i$ denotes the value of the consequent variable z for r^i. The asterisk sign $(*)$ denotes that a value has been directly observed.

A. Transformation-Based FRI

A key concept used in T-FRI is the representative value $\mathrm{rep}(\hat{a}_j)$ of a fuzzy set $\hat{a}_j$, it captures important information such as the overall location of a fuzzy set. For triangular membership functions in the form of $\hat{a}_j = (\hat{a}_{j1}, \hat{a}_{j2}, \hat{a}_{j3})$, where $\hat{a}_{j1}$, $\hat{a}_{j3}$ represent the left and right extremities of the support (with membership values 0), and $\hat{a}_{j2}$ denotes the normal point (with a membership value of 1), $\mathrm{rep}(\hat{a}_j)$ is defined as the centre of gravity of these three points:

$$\mathrm{rep}(\hat{a}_j) = \frac{\hat{a}_{j1} + \hat{a}_{j2} + \hat{a}_{j3}}{3} \tag{1}$$

More generalised forms of representative values for more complex membership functions have also been defined in [20].

The following is an outline of the T-FRI algorithm:

1) Identification of the Closest Rules: The distance between any two rules $r^p, r^q \in R$, is determined by computing the aggregated distance between all the antecedent variable values:

$$d(r^p, r^q) = \sqrt{\sum_{j=1}^{|A|} d(\hat{a}_j^p, \hat{a}_j^q)^2} \tag{2}$$

where

$$d(\hat{a}_j^p, \hat{a}_j^q) = \frac{|\mathrm{rep}(\hat{a}_j^p) - \mathrm{rep}(\hat{a}_j^q)|}{\max_{a_j} - \min_{a_j}}$$

is the normalised result of the otherwise absolute distance measure, so that distances are compatible with each other over different variable domains. The distance between a given rule r^p and the observation o^*: $d(r^p, o^*)$ may be calculated in the same manner, and the two closest rules, say r^u and r^v, are identified and used for the subsequent interpolation process.

2) Construction of the Intermediate Fuzzy Rule: The intermediate fuzzy rule r' is the starting point of the transformation process in T-FRI. It consists of a series of intermediate antecedent fuzzy sets $\hat{a}'_j$, and an intermediate consequent fuzzy set $\hat{z}'$:

$$r' : \text{if } a_1 \text{ is } \hat{a}'_1 \wedge \cdots \wedge a_j \text{ is } \hat{a}'_j \wedge \cdots \wedge a_{|A|} \text{ is } \hat{a}'_{|A|}, \text{ then } z \text{ is } \hat{z}'$$

which is a weighted aggregation of the two selected rules r^u and r^v. For each of the antecedent dimensions a_i, a ratio $\lambda_{\hat{a}_j}, 0 \leq \lambda_{\hat{a}_j} \leq 1$ is introduced, which represents the contribution of $\hat{a}_j^v$ towards the formation of $\hat{a}'_j$ with respect to $\hat{a}_j^u$:

$$\lambda_{\hat{a}_j} = \frac{d(\hat{a}_j^u, \hat{a}_j^*)}{d(\hat{a}_j^u, \hat{a}_j^v)}$$

The intermediate antecedent fuzzy set $\hat{a}'_j$ is then computed using:

$$\hat{a}'_j = (1 - \lambda_{\hat{a}_j})\hat{a}_j^u + \lambda_{\hat{a}_j}\hat{a}_j^v \tag{3}$$

The position and shape of the intermediate consequent fuzzy set $\hat{z}'$, are then calculated in the same manner to the above using the consequent fuzzy sets of the two rules $\hat{z}^u$ and $\hat{z}^v$, where the ratio $\lambda_{\hat{z}}$ is obtained by averaging the ratios of the antecedent variables:

$$\lambda_{\hat{z}} = \frac{1}{|A|}\sum_{j=1}^{|A|} \lambda_{\hat{a}_j}$$

3) Computation of the Scale and Move Parameters: The goal of a transformation process T is to scale, move (or skew) an intermediate fuzzy set $\hat{a}'_j$, so that the transformed shape coincides with that of the observed value $\hat{a}_j^*$. In T-FRI, such a process is performed in two stages:

a) the scale operation from $\hat{a}'_j$ to $\hat{a}''_j$ (denoting the scaled intermediate fuzzy set), in an effort to determine the scale ratio $s_{\hat{a}_j}$; and
b) the move operation from $\hat{a}''_j$ to $\hat{a}_j^*$ to obtain a move ratio $m_{\hat{a}_j}$. Once performed for each of the antecedent variables, the necessary parameters $s_{\hat{z}}$ and $s_{\hat{z}}$ for the consequent variable can be approximated as follows, in order to compute the final interpolation result $\hat{z}^*$.

For a triangular fuzzy set $\hat{a}'_j = (\hat{a}'_{j1}, \hat{a}'_{j2}, \hat{a}'_{j3})$, the scale ratio $s_{\hat{a}_j}$ is calculated using:

$$s_{\hat{a}_j} = \frac{\hat{a}_{j3}^* - \hat{a}_{j1}^*}{\hat{a}'_{j3} - \hat{a}'_{j1}} \tag{4}$$

which essentially expands or contracts the support length of $\hat{a}'_j$: $\hat{a}'_{j3} - \hat{a}'_{j1}$ so that it becomes the same as that of $\hat{a}^*_j$. The scaled intermediate fuzzy set $\hat{a}''_j$, which has the same representative value as $\hat{a}'_j$, is then acquired using the following:

$$\begin{cases} \hat{a}''_{j1} = \frac{(1+2s_{\hat{a}_j})a_{j1}' + (1-s_{\hat{a}_j})a_{j2}' + (1-s_{\hat{a}_j})a_{j3}'}{3} \\ \hat{a}''_{j2} = \frac{(1-s_{\hat{a}_j})a_{j1}' + (1+2s_{\hat{a}_j})a_{j2}' + (1-s_{\hat{a}_j})a_{j3}'}{3} \\ \hat{a}''_{j3} = \frac{(1-s_{\hat{a}_j})a_{j1}' + (1-s_{\hat{a}_j})a_{j2}' + (1+2s_{\hat{a}_j})a_{j3}'}{3} \end{cases}$$

The move operation shifts the position of $\hat{a}''_j$ to becoming the same as that of $\hat{a}^*_j$, while maintaining its representative value rep$(\hat{a}''_j)$. This is made possible by using a tailored move ratio $m_{\hat{a}_j}$:

$$\begin{cases} m_{\hat{a}_j} = \frac{3(\hat{a}^*_{j1} - \hat{a}''_{j1})}{\hat{a}''_{j2} - \hat{a}''_{j1}}, \text{ if } \hat{a}^*_{j1} \geq \hat{a}''_{j1} \\ m_{\hat{a}_j} = \frac{3(\hat{a}^*_{j1} - \hat{a}''_{j1})}{\hat{a}''_{j3} - \hat{a}''_{j2}}, \text{ otherwise} \end{cases}$$

The final positions of the triangle's three points are calculated as follows:

$$\begin{cases} \begin{cases} \hat{a}^*_{j1} = \hat{a}''_{j1} + m_{\hat{a}_j} \frac{\hat{a}''_{j2} - \hat{a}''_{j1}}{3} \\ \hat{a}^*_{j2} = \hat{a}''_{j2} - 2m_{\hat{a}_j} \frac{\hat{a}''_{j2} - \hat{a}''_{j1}}{3} \\ \hat{a}^*_{j3} = \hat{a}''_{j3} + m_{\hat{a}_j} \frac{\hat{a}''_{j2} - \hat{a}''_{j1}}{3} \end{cases} & \text{, if } m_{\hat{a}_j} \geq 0 \\ \begin{cases} \hat{a}^*_{j1} = \hat{a}''_{j1} + m_{\hat{a}_j} \frac{\hat{a}''_{j3} - \hat{a}''_{j2}}{3} \\ \hat{a}^*_{j2} = \hat{a}''_{j2} - 2m_{\hat{a}_j} \frac{\hat{a}''_{j3} - \hat{a}''_{j2}}{3} \\ \hat{a}^*_{j3} = \hat{a}''_{j3} + m_{\hat{a}_j} \frac{\hat{a}''_{j3} - \hat{a}''_{j2}}{3} \end{cases} & \text{, otherwise} \end{cases}$$

Note that this operation also guarantees that the resultant shape is convex and normal.

4) Scale and Move Transformation on Intermediate Consequent Fuzzy Set: After computing the necessary parameters based on all of the observed antecedent variable values, the required parameters for $\hat{z}'$ are then determined by averaging the corresponding parameter values:

$$s_{\hat{z}} = \frac{1}{|A|} \sum_{j=1}^{|A|} s_{\hat{a}_j} \quad m_{\hat{z}} = \frac{1}{|A|} \sum_{j=1}^{|A|} m_{\hat{a}_j} \tag{6}$$

A complete scale and move transformation from the initial intermediate consequent fuzzy set $\hat{z}'$ to the final interpolative output $\hat{z}^*$, may be represented concisely by: $\hat{z}^* = T(\hat{z}', s_{\hat{z}}, m_{\hat{z}})$, highlighting the importance of the two key transformations required.

B. Backward Fuzzy Rule Interpolation

Backward Fuzzy Rule Interpolation (B-FRI) [19] is a recently proposed extension to standard (forward) FRI. It allows crucial missing values that directly relate to the conclusion to be inferred, or interpolated from the known antecedent values and the conclusion. This technique supplements a conventional FRI process, and is particularly beneficial in the presence of hierarchically arranged rule bases, since a normal inference or interpolation system will be unable to proceed if certain key antecedent values (that connect the sub-rule bases) are missing. An implementation of the B-FRI concept has been developed, based on the mechanisms of T-FRI. It works by reversely approximating the scale and move transformation parameters for the variables with missing values.

Despite that both forward and backward T-FRI share the same underlying analogy-based idea, backward T-FRI has several subtle differences, such as the procedures to select the closest rules, and those to compute the transformation parameters. For instance, assume that the value of the antecedent variable a_l is missing from the observation, whilst the conclusion $\hat{z}^*$ can be directly observed. The distance measurement $\underleftarrow{d}(r^p, r^q)$ between any two rules is handled with a bias towards the consequent variable:

$$\underleftarrow{d}(r^p, r^q) = \sqrt{|A| \cdot d(\hat{z}^p, \hat{z}^q)^2 + \sum_{j=1,\ j \neq l}^{|A|} d(\hat{a}^p_j, \hat{a}^q_j)^2} \tag{7}$$

This is because the observed value for the consequent variable embeds more information, and the weight intuitively assigned is equal to the sum of all individual antecedents $|A|$.

Having identified the closest rules, the remaining steps are the same as forward T-FRI, except that the parameters for the missing antecedent: $\lambda_{\hat{a}_l}$, $s_{\hat{a}_l}$, and $m_{\hat{a}_l}$ are calculated using a set of revised but fundamentally similar formulae. For instance, the formula to calculate $\lambda_{\hat{a}_l}$ is:

$$\lambda_{\hat{a}_l} = |A| \lambda_{\hat{z}} - \frac{1}{|A|} \sum_{j=1, j \neq l}^{|A|} \lambda_{\hat{a}_j} \tag{8}$$

Here, the required parameters are obtained by subtracting the sum of the values of the given antecedents from that of the consequent (also multiplied by a weight of $|A|$). The final backward interpolation result $\hat{a}^*_l$ can then be obtained. For notional simplicity: $\underleftarrow{T}(\hat{a}'_l, s_{\hat{a}_l}, m_{\hat{a}_l})$ is used as the abbreviation that summaries the entire B-FRI process hereafter.

III. Antecedent Significance-Based FRI

This section discusses the similarities and differences between the problem domain of FS and that of FRI, and describes the approach developed that evaluates the importance of rule antecedents using FS techniques. A weighted aggregation-based approach is also introduced, which makes use of the antecedent significance values to better approximate the interpolation results. The potential benefits of the proposed technique in B-FRI are also explained.

A. From Feature Selection to Antecedent Selection

In the context of FS, following the notions employed earlier in Section II, an information system is also a tuple $\langle \mathbb{U}, A \rangle$, where $\mathbb{U} = \{x_1, \cdots, x_{|\mathbb{U}|}\}$ is a non-empty set of finite objects (commonly referred to as the universe of discourse); and $A = \{a_1, \cdots, a_{|\mathbb{A}|}\}$ is a non-empty, finite set of features such that $a : \mathbb{U} \to V_a$ for every $a \in \mathbb{A}$. Here V_a is the set of values that feature a may take. For decision systems in particular, there exists a set of decision features Z as the injunction to the input features $\mathbb{A}$.

The above definition shows that, the key distinction between a standard FS problem and FRI is the presence of the continuously-valued consequent variable z, and that there is generally no well-defined class labels (hence the need for interpolation). Therefore, only a selected few of non-class-label-dependent feature evaluators [21] can be readily adapted for FRI, including correlation-based FS (CFS) [22] and fuzzy-rough set-based FS (FRFS) [11] which support regression tasks by default. FRFS in particular, relies on the use of fuzzy similarity to differentiate between two training objects. It employs a strict equivalence relation for class labels or categorical data, but the underlying concepts (i.e., the upper and lower approximations) may also be constructed using the real-valued "decision" variable.

It is essential to clarify that, a "sparse" rule base considered in the paper is in the sense of coverage, or the distance between any given rules within the system, rather than quantity. Similarly, a rule base may consist of a very small number of rules yet still be described as dense, if the underlying knowledge is simple (typically with limited antecedent values). Furthermore, the proposed method is in principle, also applicable to standard inference scenarios, with largely overlapping rules.

Fig. 1 illustrates the proposed antecedent selection approach for FRI. To achieve antecedent selection, a given feature evaluator such as CFS or FRFS may be employed as is, once the rule base to be processed is converted into a standard, crisp-valued data set. For this, any defuzzification mechanism may be adopted, and in this paper, the representative value of a fuzzy set (Eqn. 1) is used for this purpose. The newly created, crisp-valued data set (antecedent values) are then employed to train a feature (antecedent) evaluator. This is in order to obtain a set of "feature evaluation scores", or antecedent importance measurements $\omega'_{a_j}, j = 1, \cdots, |A|$, which are subsequently normalised to yield the required significance values:

$$\{\omega_{a_1}, \cdots, \omega_{a_{|A|}}\} = \{\frac{\omega'_{a_1}}{\Sigma_{j=1}^{|A|}\omega'_{a_j}}, \cdots, \frac{\omega'_{a_{|A|}}}{\Sigma_{j=1}^{|A|}\omega'_{a_j}}\}$$

These values indicate the relevance of the underlying antecedent variables A, with respect to the values of the consequent variable z, based on the information embedded in the rule base.

A feature subset search algorithm such as HSFS may be employed to identify a quality antecedent subset $B \subseteq A$, which captures the information within the original rule base R to a reasonable (if not the maximum) extent. R may then be pruned to just maintaining the highest quality antecedent variables, thereby producing a reduced rule base (much like a reduced data set with irrelevant features removed). Subsequent tasks such as rule selection, fuzzy inference, or FRI may benefit greatly in terms of accuracy and efficiency, once such redundant and noisy antecedent variables have been eliminated.

B. Weighted Aggregation of Antecedent Significance Measures

For a given rule base R, a set of antecedent significance values: $\{\omega_{a_1}, \cdots, \omega_{a_{|A|}}\}$ may be computed, or supplied by subject experts. A weighted rule ranking strategy can be derived for the purpose of identifying the most suitable rules

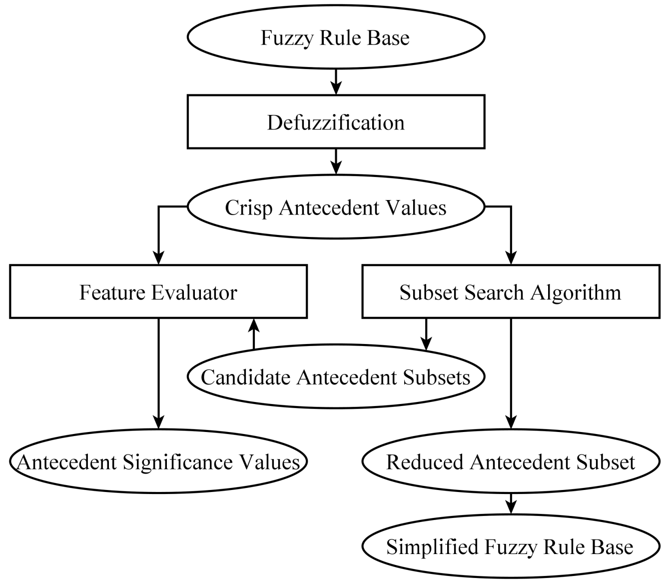

Fig. 1. Antecedent selection procedures

to perform interpolation. Recall that the standard (un-biased) formula (Eqn. 2) adopted by T-FRI for calculating the distance between any given two rules $r^p, r^q \in R$, is effectively to assume equal significance of all involved antecedent variables. A general form of weighted distance $\tilde{d}$ may be defined by:

$$\tilde{d}(r^p, r^q) = \sqrt{\sum_{j=1}^{|A|} \omega_{a_j} d(\hat{a}_j^p, \hat{a}_j^q)^2} \tag{9}$$

which takes into consideration the significance values ω_{a_j} of the antecedent variables $a_j, j = 1, \cdots, |A|$.

The use of $\tilde{d}$ allows for a more flexible selection of rules. For instance, consider the case illustrated in Fig. 2, with the assumption that a_1 and a_3 are antecedents of high significance and a_2 is irrelevant (or noisy). For a given new observation o^*, suppose that the two closest rules determined by the standard T-FRI (using un-biased distance measure) are r^1 and r^2. Also, there may exist another rule r^3 (involving dashed fuzzy sets) with values much closer to a_1 and a_3, but it has not yet been selected because its overall distance to the observation is greater than that of r^2 (or r^1), due to the value $\hat{a}_2^3$ being further away. Since in fact a_2 is of little importance, a weighted distance measurement may select r^1, r^3 to perform interpolation, and the end result $\hat{z}^*_{(1,3)}$ may provide a better estimation for this scenario, than the result obtained using r^1 and r^2.

As alternative rules may be selected via the use of weighted distance calculation, any existing FRI mechanism should therefore be modified in order to ensure consistency amongst the results interpolated using different rules. In this paper, the investigation is focused on the T-FRI method introduced in Section II-A. However, the notion of the antecedent variable significance appears to be equally applicable to other types of FRI technique, such as α-cut-based methods [1], [2], [23].

Recall the first step of T-FRI, the construction of the

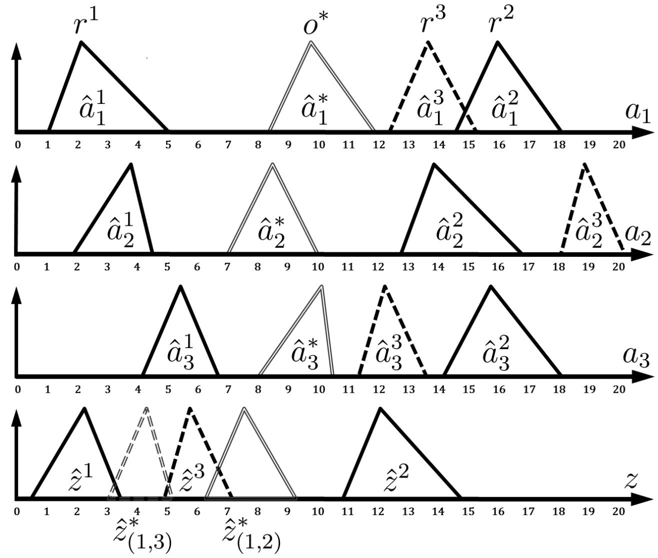

Fig. 2. Alternative rule selection using weighted distance calculation

intermediate fuzzy rule r' requires the set of intermediate antecedent fuzzy sets $\hat{a}'_j$, and the intermediate consequent fuzzy set $\hat{z}'$. A set of shift parameters $\{\lambda_{\hat{a}_1}, \cdots, \lambda_{\hat{a}_{|A|}}, \lambda_{\hat{z}}\}$ are required, in order to maintain the position (representative value) of r' on each of its antecedent dimensions. The value of $\lambda_{\hat{z}}$ plays an important role in determining the initial position of the intermediate consequent fuzzy set, which will affect the final interpolative output. For the present problem, the calculation of $\lambda_{\hat{z}}$ is modified to reflect the variations in antecedent variable significance, thereby producing a weighted shift parameter $\tilde{\lambda}_{\hat{z}}$:

$$\tilde{\lambda}_{\hat{z}} = \frac{1}{|A|}\sum_{j=1}^{|A|} \omega_{a_j}\lambda_{a_j}$$

which is then used to obtain the weighted intermediate consequent fuzzy set $\hat{z}'$. Thus, it is then necessary to apply the transformations to the intermediate consequent fuzzy set $\hat{z}'$, with weighted scale ratio $\tilde{s}_{\hat{z}}$ and move ratio $\tilde{m}_{\hat{z}}$ computed by

$$\tilde{s}_{\hat{z}} = \frac{1}{|A|}\sum_{j=1}^{|A|} \omega_{a_j}s_{\hat{a}_j} \quad \tilde{m}_{\hat{z}} = \frac{1}{|A|}\sum_{j=1}^{|A|} \omega_{a_j}m_{\hat{a}_j}$$

These are a modified version of Eqn. 6, following the same principle as that applied to the calculation of $\tilde{\lambda}_{\hat{z}}$.

Finally, a complete, weighted T-FRI procedure from $\hat{z}'$ to $\hat{z}^*$ can be readily created, by following the transformation $\hat{z}^* = \tilde{T}(\hat{z}', \tilde{s}_{\hat{z}}, \tilde{m}_{\hat{z}})$. This weighted aggregation procedure makes minimal alterations to the original T-FRI algorithm. Symbolically, it appears identical to the conventional T-FRI method, and is therefore omitted here. As such, the procedure maintains its structural simplicity and intuitive appeal, while extending the capability of T-FRI.

C. Use of Antecedent Significance in B-FRI

One of the common problems faced by a B-FRI system is the event where more than one antecedent value is missing from an observation. It is difficult to fully reconstruct, or even closely approximate multiple missing values, since there may exist a number of possible combinations of values that may each lead to the same conclusion. It is also computationally complex to perform reverse reasoning with a large number of unknowns. Antecedent selection, being a dimensionality reducing technique, may be potentially beneficial in such situations. By identifying more important antecedent variables, or by removing irrelevant antecedents altogether, a priority-based backward reasoning system may be established and greatly simplifies the problem. However, much of relevant research concerning this issue is beyond the scope of this paper.

IV. Experimentation and Discussion

This section provides a real-world scenario concerning the prediction of terrorist activities, it is used to demonstrate the procedures of the proposed antecedent significance-based approach, for both conventional T-FRI and B-FRI problems. The accuracy and efficiency of the work is further validated via a systematic evaluation using synthetic random data.

A. Terrorist Activity Prediction

Consider a practical scenario that requires the prediction of terrorist bombing *risk*. The likelihood of an explosion may be directly affected by the number of people in the area, and crowded places (high *popularity* and high travel *convenience*) are usually more likely to attract terrorist attentions. Safety precaution such as police *patrol* may also be a very important factor, the more alert and prepared a place is, the less opportunities there are for the terrorists to attack. Other aspects such as *temperature* and *humidity* may be of relevance, but their impact on the potential outcome is limited. Table I lists a few example linguistic rules that may be derived for coping with such a scenario.

TABLE I
EXAMPLE LINGUISTIC RULES FOR TERRORIST BOMBING PREDICTION (M. FOR MODERATE, V. FOR VERY)

	popularity a_1	*convenience* a_2	*patrol* a_3	*temperature* a_4	*humidity* a_5	*risk* z
r^1	V. Low	V. Low	V. High	M.	High	V. Low
r^2	V. Low	V. High	V. Low	High	Low	V. Low
r^3	M.	High	M.	Low	M. High	M.
r^4	M.	M.	M.	Low	Low	M. Low
r^5	M.	High	Low	M. Low	M. High	High
r^6	High	V. Low	High	V. Low	Low	V. Low
r^7	High	High	M.	M.	High	M. Low
r^8	High	High	V. Low	Low	Low	V. High

The antecedent significance values obtained using CFS and FRFS are presented in Table II. Both feature evaluators agree on that *temperature* and *humidity* are relatively less important than the other 3 antecedent variables. CFS in particular, assigns a weight of $\omega_{a_5} = 0.0299$ to *humidity*, signifying its relatively lack of relevancy in this rule base. The ranking of importance for the major antecedent variables is $a_3 > a_1 > a_2$, when CFS is used. The resultant ranking determined by FRFS is similar, thought it gives *convenience* (a_2) a higher significance score.

TABLE II
ANTECEDENT SIGNIFICANCE VALUES DETERMINED USING FEATURE EVALUATORS CFS AND FRFS

	w_{a_1}	w_{a_2}	w_{a_3}	w_{a_4}	w_{a_5}
CFS	0.2765	0.2461	0.3312	0.1163	0.0299
FRFS	0.2220	0.3228	0.2904	0.0833	0.0814

1) FRI Example: Suppose that a new observation o^* is present that requires interpolation, its linguistic values and the underlying semantics in terms of triangular fuzzy sets, are given in Table III. In order to illustrate and compare the outcomes of the rule selection procedure, the rules selected using the standard T-FRI process and the antecedent significance-based weighted distance metric are all provided. The results show that the two closest rules selected following the standard T-FRI process are close to the observed values on all antecedent dimensions. However, if antecedent significance values are taken into consideration, alternative rules will be selected. For the two rules selected according to CFS, large differences in values can be observed for the variable *humidity* (a_5), which is likely caused by its very low significance value, as shown previously in Table II.

The detailed calculations of the T-FRI transformations are omitted here to save space as they are easily conceived. The final interpolated result using the standard T-FRI is $\hat{z}^* = (3.1, 3.6, 4.4)$, following a transformation of $T(\hat{z}' = (3.4, 3.7, 4), s_{\hat{z}} = 2.163, m_{\hat{z}} = 0.003)$. Using weights determined by CFS, the result is $\hat{z}^* = (2.2, 2.8, 3.2)$, the weighted transformation is:

$$\tilde{T}(\hat{z}' = (2.4, 2.8, 3), \tilde{s} = 1.479, \tilde{m} = -0.101)$$

The results obtained based on FRFS is $\hat{z}^* = (2.2, 3.2, 4)$ with a corresponding weighted transformation as shown below:

$$\tilde{T}(\hat{z}' = (2.7, 3, 3.6), \tilde{s}_{\hat{z}} = 2.084, \tilde{m}_{\hat{z}} = -0.337)$$

Intuitively speaking, although the area in question may be crowded (due to the place being popular and convenient to reach), the risk of an attack can still be quite low. This is because the area is in a moderate high level of alert, and the two weather-related factors (despite being less significant): low temperature and high humidity may also discourage any potential terrorist activities. One of the example rules listed in Table I, r^7 describes a fairly similar event, where the consequent value is given as M. Low. From the analysis above, the results obtained via weighted aggregation: $(2.2, 2.8, 3.2)$ (Low) using significance values determined by CFS, and $(2.2, 3.2, 4)$ (Low) by FRFS, are both more intuitively agreeable than that produced by the standard T-FRI: $(3.1, 3.6, 4.4)$ (M. Low).

2) B-FRI Example: For the B-FRI scenario, suppose that an observation o^* with a missing value for the antecedent variable *patrol* (a_3) is given with the consequent variable *risk* (z) directly observed. Table IV lists the observation o^*, and the different rules selected by the respective approaches. Note that both CFS- and FRFS-based weighted distance metrics select the same two closest rules, both of which differ from those selected by the standard B-FRI. For the latter, the values of the required parameters for the missing antecedent variable are computed through those of the known antecedents and the consequent. For example, following Eqn. 8, $\lambda_{\hat{a}_3}$ can be calculated such that:

$$\begin{aligned}\lambda_{\hat{a}_3} &= 5\lambda_{\hat{z}} - \frac{1}{5}\sum_{j=1, j\neq 2}^{5} \lambda_{\hat{a}_j} \\ &= 5 \times 0.792 - 2.171 = 1.789\end{aligned}$$

which then constructs an intermediate fuzzy term $\hat{a}_3' = (0.2, 1.2, 2.2)$. Both $s_{\hat{a}_3}$ and $m_{\hat{a}_3}$ are computed similar to $\lambda_{\hat{a}_3}$, and finally, the backward transformation $\underleftarrow{T}$ given below is derived, which provides the final B-FRI output as V. Low:

$$\hat{a}_3^* = \underleftarrow{T}((0.2, 1.2, 2.2), 0.400, -0.172) = (0.7, 1.2, 1.5)$$

To avoid unnecessary repetition, the detailed procedures to compute the weighted B-FRI outputs are omitted. The CFS-based antecedent significance values yield the following weighted B-FRI transformation:

$$\hat{a}_3^* = \underleftarrow{\tilde{T}}((1.5, 2.5, 3.5), 0.1389, 0.0185) = (2.4, 2.5, 2.7) \text{ (Low)}$$

while the FRFS-based method calculates slightly differently, resulting in the following backward interpolative outcome:

$$\hat{a}_3^* = \underleftarrow{\tilde{T}}((1.7, 2.7, 3.7), 0.0807, 0.0714) = (2.6, 2.7, 2.8)$$

which may also be interpreted as of a linguistic meaning of Low.

Note that all of the values, except those for *patrol* and *risk*, are the same as the previous observation used to demonstrate forward T-FRI. This narrows down the reason why *risk* has jumped from Low to M. High, which is the level of *patrol* in the area. Intuitively, for a highly crowded area, if very little *patrol* is present (as suggested by the result of the standard B-FRI: V. Low), the resultant value of *risk* should become V. High. Therefore, having a Low level of *patrol* may be a more appealing approximation,

B. Systematic Evaluation

In this set of simulation-based experiments, a simple numerical test function with 15 variables ($|A| = 15$) is employed, as shown below:

$$\begin{aligned} y = &-1 + \sqrt{10x_1} + \frac{x_2}{10000} + \frac{x_3}{10000} + \frac{x_4}{10000} - 5x_5 \\ &+ \frac{x_6}{10000} - x_7 + \frac{100}{x_8+1} - \frac{x_9}{10000} + \frac{x_{10}}{10000} + \frac{x_{11}}{10000} \\ &- \frac{x_{12}}{10000} + \frac{x_{13}}{10000} - x_{14}^2 + \frac{x_{15}}{10000}\end{aligned}$$

Such a systematic test is important to validate the consistency, accuracy, and robustness of the developed approach.

TABLE III
EXAMPLE OBSERVATION (LINGUISTIC TERMS AND FUZZY SET REPRESENTATIONS), AND THE CLOSEST RULES SELECTED BY STANDARD T-FRI, AND WEIGHTED T-FRI WITH VALUES DETERMINED USING CFS AND FRFS

		popularity a_1	*convenience* a_2	*patrol* a_3	*temperature* a_4	*humidity* a_5	*risk* z
	o^*	High	High	M. High	Low	M. High	?
	o^*	$(8.0, 8.5, 9)$	$(5.8, 7.5, 8)$	$(5.0, 5.5, 6.0)$	$(1.5, 2.0, 3)$	$(5.5, 6.0, 6.5)$	?
Standard	r^1	$(8, 8.3, 8.4)$	$(8.4, 8.6, 9.1)$	$(5.4, 5.9, 6.2)$	$(3.5, 3.7, 4)$	$(6.3, 6.9, 7.2)$	$(2.9, 3, 3.3)$
Standard	r^2	$(9.7, 9.8, 10.4)$	$(4.9, 5.4, 5.5)$	$(1.7, 2.1, 2.2)$	$(1.2, 1.2, 1.8)$	$(3.7, 4.3, 4.4)$	$(4.3, 5, 5.4)$
CFS	r^1	$(8.5, 9.1, 9.8)$	$(8.1, 8.6, 9)$	$(5, 5.4, 6.1)$	$(1.8, 2.3, 2.4)$	$(2.4, 3, 3.1)$	$(2.7, 3, 3.2)$
CFS	r^2	$(6, 6.6, 6.8)$	$(5.3, 5.9, 6.1)$	$(7.4, 7.6, 7.8)$	$(1.7, 2.1, 2.8)$	$(9.2, 9.6, 10.2)$	$(1.4, 2, 2.5)$
FRFS	r^1	$(8.9, 9.3, 9.8)$	$(7.2, 7.6, 8.3)$	$(6.3, 6.3, 6.4)$	$(0.6, 0.7, 1.3)$	$(4.1, 4.5, 4.7)$	$(2.6, 3, 3.7)$
FRFS	r^2	$(6.9, 6.9, 7)$	$(3.9, 4.2, 4.7)$	$(3.6, 3.9, 4.6)$	$(2.8, 3.5, 4.3)$	$(7.9, 7.9, 8.1)$	$(2.8, 3, 3.4)$

TABLE IV
EXAMPLE OBSERVATION (BOTH LINGUISTIC TERMS AND FUZZY SET REPRESENTATIONS), AND THE CLOSEST RULES SELECTED BY STANDARD B-FRI, AND WEIGHTED B-FRI WITH VALUES DETERMINED USING CFS AND FRFS

		popularity a_1	*convenience* a_2	*patrol* a_3	*temperature* a_4	*humidity* a_5	*risk* z
	o^*	High	High	?	Low	M. High	M. High
	o^*	$(8.0, 8.5, 9)$	$(5.8, 7.5, 8)$	?	$(1.5, 2.0, 3)$	$(5.5, 6.0, 6.5)$	$(5.1, 5.8, 6.4)$
Standard	r^1	$(8.7, 9.7, 10.7)$	$(5.4, 6.4, 7.4)$	$(0.9, 1.9, 2.9)$	$(0.2, 1.2, 2.2)$	$(6.7, 7.7, 8.7)$	$(4.6, 5.6, 6.6)$
Standard	r^2	$(7.5, 8.5, 9.5)$	$(6.9, 7.9, 8.9)$	$(0.5, 1.5, 2.5)$	$(7.2, 8.2, 9.2)$	$(3.9, 4.9, 5.9)$	$(4.8, 5.8, 6.8)$
CFS	r^1	$(7.7, 8.7, 9.7)$	$(5.9, 6.9, 7.9)$	$(4.1, 5.1, 6.1)$	$(1.0, 2.0, 3.0)$	$(3.8, 4.8, 5.8)$	$(2.3, 3.3, 4.3)$
CFS	r^2	$(6.7, 7.7, 8.7)$	$(7.5, 8.5, 9.5)$	$(0.0, 0.8, 1.8)$	$(3.0, 4.0, 5.0)$	$(5.2, 6.2, 7.2)$	$(5.7, 6.7, 7.7)$
FRFS	r^1	$(7.7, 8.7, 9.7)$	$(5.9, 6.9, 7.9)$	$(4.1, 5.1, 6.1)$	$(1.0, 2.0, 3.0)$	$(3.8, 4.8, 5.8)$	$(2.3, 3.3, 4.3)$
FRFS	r^2	$(6.7, 7.7, 8.7)$	$(7.5, 8.5, 9.5)$	$(0.0, 0.8, 1.8)$	$(3.0, 4.0, 5.0)$	$(5.2, 6.2, 7.2)$	$(5.7, 6.7, 7.7)$

This is because random samples may be generated from a controlled environment, where the ground truths are also available to verify the correctness of the interpolation results. These tests share a similar underlying principle behind that of cross-validation and statistical evaluation [24], [25].

1) FRI Results: The results shown in Table V are averaged outcomes of 200 randomised runs. By employing the weighted aggregation scheme based on the antecedent significance values, both the mean error and standard deviation are considerably improved. The results obtained according to FRFS appears to have a slightly higher mean error and a wider spread, however, t-test ($p = 0.01$) shows that the difference is not statistically significant. The improvement is more evident when the original rule base is simplified by removing the redundant antecedent variables.

TABLE V
EVALUATION OF PROPOSED APPROACHES FOR STANDARD FRI

	Mean Error %	S.D. %
Standard T-FRI	7.32	6.15
Weighted by CFS	5.33	4.69
Weighted by FRFS	5.68	5.16
Reduced by CFS	3.38	3.01
Reduced by FRFS	3.33	2.63

The antecedent subset selected by CFS is $\{a_0, a_4, a_7, a_{13}\}$, a reduction of 73% in the number of variables, which achieves a mean error of 3.38%; the subset selected by FRFS is $\{a_1, a_4, a_7, a_9, a_{13}\}$, with a reduction of 67%, witch helps obtain an mean error of 3.33%. Both evaluators yield reasonable reduction results, and the interpolation error (compared to numerical function's true output) is also much lower than the standard and weighted T-FRI.

2) B-FRI Results: Given the numerical test function,a randomly selected antecedent variable is set to be missing per test subject to the constraint that this "missing" variable is drawn from the set $\{a_0, a_4, a_7, a_{13}\} \cap \{a_1, a_4, a_7, a_9, a_{13}\} = \{a_1, a_7, a_{13}\}$, which is the intersection of the two antecedent subsets identified by CFS and FRFS, respectively. This set up allows direct comparison between the different techniques.

The proposed weighted aggregation scheme, and the reduced rule base containing only the most significant antecedents (determined via antecedent selection) are then used to reconstruct the original values. In this set of experiments, the error is calculated with respect to the actual antecedent variable value, which has been intentionally removed to simulate the B-FRI environment. The mean error and standard deviation of the 200 simulated tests are given in Table VI.

The number of antecedent variables involved is quite large and presents a considerable challenge for precise backward reasoning. The original B-FRI approach achieves a 18.20% mean error, while the accuracy is slightly improved when weighted aggregation is used. Based on the simplified rule bases that are reduced through the use of either CFS or FRFS, the interpolation accuracy is notably improved, with a mean error of 8.45% and 6.95%, respectively. Furthermore, the

TABLE VI
EVALUATION OF PROPOSED APPROACHES FOR B-FRI

	Mean Error %	S.D. %
Standard B-FRI	18.20	19.40
Weighted by CFS	16.94	19.15
Weighted by FRFS	17.59	18.76
Reduced by CFS	8.45	13.70
Reduced by FRFS	6.93	13.70

quality of the output is also more stable with the standard derivation, dropping from the original 19.40% to 13.70% for both cases, demonstrating the benefits of the reduced rule base for B-FRI.

V. CONCLUSION

This paper has presented a new FRI approach that exploits FS techniques in order to evaluate the importance of antecedent variables. A weighted aggregation-based interpolation method is proposed that makes use of the identified antecedent significance values. The original rule base may also be simplified by removing the irrelevant or noisy antecedents using a feature subset search algorithm such as HSFS, and retains an antecedent subset of a much lower dimensionality. Example scenarios and systematic tests are employed to demonstrate the potential of this work, for both conventional and B-FRI problems. The resultant antecedent significance-based FRI technique is both technically sound and conceptually appealing, as humans often (automatically) screen out seemingly irrelevant antecedents, and focus on more important factors in order to perform reasoning.

The present antecedent selection approach for FRI may be further improved by considering unsupervised or semi-supervised FS methods [26], [27], [28], which have emerged recently for analysing the inter-dependencies between features without the aid of class information. Although generic in concept, the current implementation of antecedent significance-based aggregation is strongly coupled with the T-FRI method. It is worth further extending the principles behind to alternative FRI methods, thereby providing a potentially more flexible framework for efficient interpolation. Fuzzy aggregation functions [29], [30] are of particular assistance in realising such a task.

REFERENCES

[1] L. Koczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *International Journal of Approximate Reasoning*, vol. 9, no. 3, pp. 197–225, 1993.

[2] ——, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Information Sciences*, vol. 71, no. 1-2, pp. 169–201, 1993.

[3] S. Kovács, "Special issue on fuzzy rule interpolation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, p. 253, 2011.

[4] S.-M. Chen and Y.-C. Chang, "Weighted fuzzy rule interpolation based on ga-based weight-learning techniques," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 729–744, 2011.

[5] S.-M. Chen, Y.-K. Ko, Y.-C. Chang, and J.-S. Pan, "Weighted fuzzy interpolative reasoning based on weighted increment transformation and weighted ratio transformation techniques," *Fuzzy Systems, IEEE Transactions on*, vol. 17, no. 6, pp. 1412–1427, 2009.

[6] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*. Wiley-IEEE Press, 2008.

[7] I. Kononenko, E. Simec, and M. Robnik-Sikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Applied Intelligence*, vol. 7, pp. 39–55, 1997.

[8] S. Senthamarai Kannan and N. Ramaraj, "A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 580–585, Aug. 2010.

[9] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155–176, Dec. 2003.

[10] M. Hall, "Correlation-based feature subset selection for machine learning," Ph.D. dissertation, University of Waikato, Hamilton, New Zealand, 1998.

[11] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.

[12] H. Liu and H. Motoda, *Computational Methods of Feature Selection (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)*. Chapman & Hall/CRC, 2007.

[13] J. Wróblewski, "Ensembles of classifiers based on approximate reducts," *Fundamenta Informaticae*, vol. 47, no. 3-4, pp. 351–360, Oct. 2001.

[14] R. Diao and Q. Shen, "Feature selection with harmony search," *IEEE Trans. Syst., Man, Cybern. B*, vol. 42, no. 6, pp. 1509–1523, 2012.

[15] Z. Huang and Q. Shen, "Fuzzy interpolative reasoning via scale and move transformations," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 340–359, 2006.

[16] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.

[17] S. Jin, R. Diao, and Q. Shen, "Backward fuzzy interpolation and extrapolation with multiple multi-antecedent rules," in *IEEE International Conference on Fuzzy Systems*, 2012, pp. 1–8.

[18] S. Jin, R. Diao, C. Quek, and Q. Shen, "Backward fuzzy rule interpolation with multiple missing values," in *IEEE International Conference on Fuzzy Systems*, 2013.

[19] ——, "Backward fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, 2014, to appear.

[20] Z. Huang and Q. Shen, "Fuzzy interpolation and extrapolation: A practical approach," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 13–28, 2008.

[21] D. Paul, E. Bair, T. Hastie, and R. Tibshirani, "Preconditioning for feature selection and regression in high-dimensional problems," *The Annals of Statistics*, vol. 36, no. 4, pp. pp.1595–1618, 2008.

[22] Y. Cui, J. Jin, S. Zhang, S. Luo, and Q. Tian, "Correlation-based feature selection and regression," in *Advances in Multimedia Information Processing*, ser. Lecture Notes in Computer Science, G. Qiu, K. Lam, H. Kiya, X.-Y. Xue, C.-C. Kuo, and M. Lew, Eds. Springer Berlin Heidelberg, 2010, vol. 6297, pp. 25–35.

[23] D. Tikk, I. Joó, L. Kóczy, P. Várlaki, B. Moser, and T. Gedeon, "Stability of interpolative fuzzy KH controllers," *Fuzzy Sets and Systems*, vol. 125, no. 1, pp. 105–119, 2002.

[24] G. Bontempi, H. Bersini, and M. Birattari, "The local paradigm for modeling and control: from neuro-fuzzy to lazy learning," *Fuzzy Sets and Systems*, vol. 121, no. 1, pp. 59–72, 2001.

[25] L. Kuncheva, "Fuzzy versus nonfuzzy in combining classifiers designed by boosting," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 6, pp. 729–741, 2003.

[26] N. Mac Parthaláin and R. Jensen, "Fuzzy-rough set based semi-supervised learning," in *IEEE International Conference on Fuzzy Systems*, Jun. 2011, pp. 2465–2472.

[27] ——, "Measures for unsupervised fuzzy-rough feature selection," *International Journal of Hybrid Intelligent Systems*, vol. 7, no. 4, pp. 249–259, Dec. 2010.

[28] J. Zhao, K. Lu, and X. He, "Locality sensitive semi-supervised feature selection," *Neurocomputing*, vol. 71, no. 10–12, pp. 1842–1849, 2008.

[29] T. Boongoen and Q. Shen, "Nearest-neighbor guided evaluation of data reliability and its applications," *IEEE Trans. Syst., Man, Cybern. B*, vol. 40, no. 6, pp. 1622–1633, Dec. 2010.

[30] Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*, ser. Cognitive Technologies. Springer, 2010.