Fuzzy Rule Based Interpolative Reasoning Supported by Attribute Ranking

Fangyi Li, Changjing Shang , Ying Li , Jing Yang, and Qiang Shen

Abstract—Using fuzzy rule interpolation (FRI) interpolative reasoning can be effectively performed with a sparse rule base where a given system observation does not match any fuzzy rules. While offering a potentially powerful inference mechanism, in the current literature, typical representation of fuzzy rules in FRI assumes that all attributes in the rules are of equal significance in deriving the consequents. This is a strong assumption in practical applications, thereby, often leading to less accurate interpolated results. To address this challenging problem, this paper employs feature selection (FS) techniques to adjudge the relative significance of individual attributes and therefore, to differentiate the contributions of the rule antecedents and their impact upon FRI. This is feasible because FS provides a readily adaptable mechanism for evaluating and ranking attributes, being capable of selecting more informative features. Without requiring any acquisition of real observations, based on the originally given sparse rule base, the individual scores are computed using a set of training samples that are artificially created from the rule base through an innovative reverse engineering procedure. The attribute scores are integrated within the popular scale and move transformation-based FRI algorithm (while other FRI approaches may be similarly extended following the same idea), forming a novel method for attribute ranking-supported fuzzy interpolative reasoning. The efficacy and robustness of the proposed approach is verified through systematic experimental examinations in comparison with the original FRI technique over a range of benchmark classification problems while utilizing different FS methods. A specific and important outcome that is supported by attribute ranking, only two (i.e., the least number of) nearest adjacent rules are required to perform accurate interpolative reasoning, avoiding the need of searching for and computing with multiple rules beyond the immediate neighborhood of a given observation.

Index Terms—Attribute ranking, feature selection (FS), fuzzy rule interpolation (FRI), interpolative reasoning, reverse engineering.

I. INTRODUCTION

FUZZY rule interpolation (FRI) enables fuzzy rule-based reasoning systems to perform inference with a sparse rule base [1], [2]. It addresses the key limitation of conventional fuzzy rule-based systems that work using compositional rule of inference (CRI) [3], where no conclusion may be drawn if none of the rules in the given rule base matches a new observation. Resolving real-world problems frequently involves the use of such sparse rule bases where all the given rules cannot fully cover the entire problem domain. FRI techniques play a significant role in such situations explicitly, where an estimation is able to be made by computing an interpolated consequent for the observation which matches no rules.

A number of FRI methods have been proposed and improved in the literature (e.g., [4]–[19]). However, conventional FRI approaches assume that the domain attributes appearing in the rule antecedents are of equal significance in the implementation of interpolation. This can lead to inaccurate or incorrect interpolated results. FRI methods that exploit rules with weighted antecedents have therefore been introduced to remedy the adverse side-effects of this equal significance assumption [20]–[23]. For example, genetic algorithms (GA) have been applied to learn the weights of rule antecedents in support of FRI [24], but this incurs a substantial increase in computation overheads and requires the setting of many additional GA parameters. Also, in [25], a weighted fuzzy interpolative reasoning method is proposed by employing weighted increment and weighted ratio transformations, entailing automatic tuning of the optimal weights of the antecedent attributes. Similar methods to this are reported in [25] and [26], all of which follow a "wrapper" scheme, where the attribute weighting procedure is enabled by firing the underlying FRI for the given training samples. A different approach is given in [27] by exploring piecewise fuzzy entropies of the fuzzy sets, with the weights assigned differently to each antecedent fuzzy set involved in different fuzzy rules, thereby, working at the expense of significant computation. An alternative work is to subjectively predefine the weights on the antecedents of the rules by experts, but this may restrict the adaptivety of the rules and therefore, the flexibility of the resulting fuzzy reasoning system [21], [28].

A common issue shared by most of the aforementioned weighting schemes is to aggregate the weights computed for individual antecedent attributes, in an effort to assign an overall weight to each rule prior to its use in interpolation. Yet, the resultant weights are utilized in rather different ways dependent upon

which underlying FRI technique is used. In these techniques, the weights are not organically integrated with the internal working of the FRI method. In terms of their typical applications, these weighted FRI techniques (e.g., [24], [27]) are typically tailored to problems such as multivariate regression and prediction. Little work has been done in developing weighted FRI to perform classification tasks (which this paper is focused on).

Feature selection (FS) [29], [30] aims to discover a minimal subset of features that are most predictive of a given outcome. It generally follows a four-step procedure: generation, evaluation, termination, and validation. Feature subsets are generated via a certain search procedure amongst the family of subsets of the original feature set. These feature subsets are then evaluated individually with regard to a given quality measure. The process of searching for a reduced feature subset is terminated if the measured quality degree reaches a satisfactory level. Finally, a selected feature subset is validated with respect to the application problem at hand. In developing effective FS mechanisms, much work has been carried out regarding the second step that evaluates the quality of a candidate feature subset [31]–[34], including those directly assessing and ranking individual features [35]–[37]. For any reasoning system (be it fuzzy or boolean), different ranking scores of features or domain attributes imply different contributions of them to the inference outcome. Inspired by this observation, a novel weighted FRI approach is proposed here, consolidating upon the initial ideas presented in [38], where a feature evaluation method is integrated within the FRI procedure to score the significance of individual rule antecedents. This is different from existing techniques for rule interpolation that involve weights (e.g., [39]–[41]), which construct an interpolated result by weighted aggregation of rule consequents, where rule importances are ranked using Euclidean distance between rule antecedents and a given observation.

In developing this new approach to fuzzy rule-based interpolative reasoning, an innovative reverse engineering process is introduced to artificially convert a given sparse rule base into a set of training samples. This is accomplished for the sake of computing the required ranking scores without the need of acquiring any real observations. This results in a attribute ranking-guided FRI method, implemented on the basis of the popular scale and move transformation-based FRI (T-FRI) [5] (although the same idea appears to be applicable to other FRI techniques). To ensure the proposed approach does not rely on a certain specific FS technique, this approach is systematically evaluated using five different feature ranking algorithms. Comparative studies demonstrate that this research helps minimize the adverse impact of the equal significance assumption made in the conventional FRI techniques, significantly improving the accuracy of the results of fuzzy interpolative reasoning. This paper also shows that supported by attribute ranking, only two (i.e., the least number of) nearest adjacent rules are required to perform accurate interpolative reasoning. This helps increase computational efficiency, without the need of searching for and operating on multiple rules beyond the immediate neighborhood of a given observation.

The remainder of this paper is organized as follows. Section II outlines the relevant background of T-FRI and reviews five popular FS approaches, each of which may be adopted for attribute ranking. Section III presents the proposed FRI method that is guided with attribute rankings. Section IV shows the results of a systematic comparative experimental evaluation. Finally, Section V concludes this paper and points out interesting issues for further research.

II. BACKGROUND

This section presents the relevant background work, including an outline of FRI based on scale and move transformations and a brief description of selected FS methods to be used for attribute ranking.

A. Transformation-Based FRI (T-FRI)

A fuzzy rule-based system essentially contains two key elements $\langle R, Y \rangle$ in describing a given problem: a nonempty finite set of domain attributes $Y = A \cup \{z\}$, where $A = \{A_j | j = 1, 2, \ldots, m\}$ represents the set of antecedent attributes and z stands for the consequent, and a nonempty set of finite fuzzy rules $R = \{r^1, r^2, \ldots, r^N\}$. In conventional FRI, a given rule $r^i \in R$ and an observation o^* are generally expressed as follows:

$$r^i : \text{if } a_1 \text{ is } A_1^i \text{ and } a_2 \text{ is } A_2^i \text{ and } \cdots \text{ and } a_m \text{ is } A_m^i$$
$$\text{then } z \text{ is } z^i$$
$$o^* : \ a_1 \text{ is } A_1^* \text{ and } a_2 \text{ is } A_2^* \text{ and } \cdots \text{ and } a_m \text{ is } A_m^* \quad (1)$$

where A_j^i represents the (fuzzy) value of the antecedent attribute a_j in the rule r^i, and z^i denotes the value of the consequent attribute z in r^i.

Given a sparse rule base R and an observation o^*, T-FRI works by running a computational process as highlighted in Fig. 1, involving four core procedures as summarized in the following.

1) Selection of Closest Rules: This procedure is required as the basis upon which to perform FRI, when o^* does not match any of the rules in the rule base. It searches for a certain number of rules that are closest to the observation. The distance between an observation o^* and a rule r^q, or the distance between any two rules $r^p, r^q \in R$, is determined by computing the aggregated distances between all the corresponding values of the shared attributes between them

$$d(v, r^q) = \frac{1}{\sqrt{m}} \sqrt{\sum_{j=1}^{m} d(A_j^v, A_j^q)^2} \quad (2)$$

where v is o^* or r^p (so A_j^v is A_j^* or A_j^p), depending on whether the distance is between an observation and a rule or between two rules. So, the n closest rules to o^* are those rules leading to the n smallest values of this distance measurement.

In the above definition

$$d(A_j^v, A_j^q) = \frac{\left| \text{Rep}(A_j^v) - \text{Rep}(A_j^q) \right|}{\max_{A_j} - \min_{A_j}} \quad (3)$$

representing the normalized result of the otherwise absolute distance, where $\max_{A_j}$ and $\min_{A_j}$ denote the maximal and minimal value of the attribute a_j, respectively. This normalization is to ensure that all distance measures are compatible with each
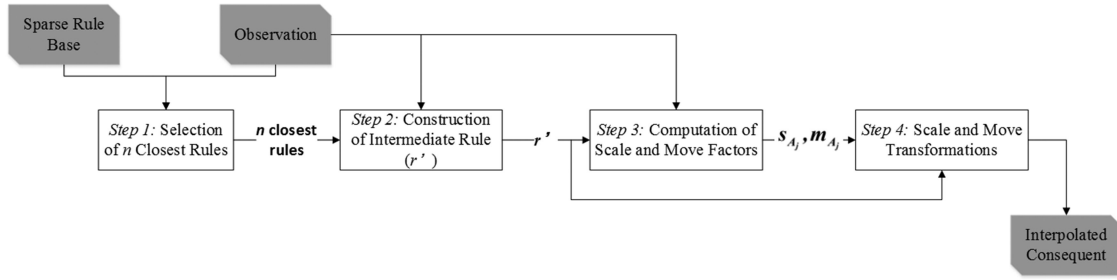
Fig. 1.   Framework of T-FRI.

other over different attribute domains. The notation $\text{Rep}(A_j)$ regarding a fuzzy set A_j in this formula represents an important concept in T-FRI, termed representative value of the fuzzy set. It reflects the key information on the overall location of A_j in its domain and also, its geometric shape. For instance, given an arbitrary polygonal fuzzy set $A = (a_1, a_2, \ldots, a_n)$ where $a_i, i = 1, 2, \ldots, n$ denote the vertices of the polygonal, its representative value $\text{Rep}(A)$ is defined by

$$\text{Rep}(A) = \sum_{i=1}^{n} w_i a_i \qquad (4)$$

where w_i is the weight assigned to the vertex a_i per i.

For computational simplicity, many fuzzy rule-based systems (including this paper) have adopted triangular membership functions to define fuzzy sets while representing attribute values. A triangular membership function is denoted in the form of $A_j = (a_{j1}, a_{j2}, a_{j3})$, with a_{j1}, a_{j3} denoting the left and right extremity of the support and a_{j2} the normal point of the fuzzy set. That is, the membership values of a_{j1} and a_{j3} are equal to 0 and the membership value of a_{j2} equals to 1. For such a fuzzy set A_j, $\text{Rep}(A_j)$ is simply defined as follows (though its center of gravity may also be used as an alternative if preferred):

$$\text{Rep}(A_j) = \frac{a_{j1} + a_{j2} + a_{j3}}{3}. \qquad (5)$$

The definition of representative values for more complex membership functions can be found in [6].

2) Construction of Intermediate Fuzzy Rule: From the preceding procedure, n closest rules to a given observation can be chosen which have the minimal distances amongst all the rules to the observation. From this, an intermediate fuzzy rule r' can be constructed, forming the starting point of the transformation process in T-FRI. In most applications of T-FRI, n is taken to be 2 purely for computational efficiency, but often at the expense of interpolative accuracy.

The construction procedure computes the antecedent fuzzy sets $A'_j, j = 1, \ldots, m$ and the corresponding consequent fuzzy set Z' of the intermediate rule:

r' : if a_1 is A'_1 and a_2 is A'_2 and $\cdots$ and a_m is A'_m, then z is Z'

which is a weighted aggregation of the n closest rules. Let $w^i_j, i \in \{1, \ldots, n\}$, denote the weight to which the jth antecedent of the ith fuzzy rule contributes to the construction of the jth antecedent A'_j of the intermediate fuzzy rule

$$w^i_j = \frac{1}{1 + d(A^i_j, A^*_j)} \qquad (6)$$

where $d(A^i_j, A^*_j)$ is calculated as per (3). Then

$$A'_j = A''_j + \delta_{A_j}(\max_{A_j} - \min_{A_j}) \qquad (7)$$

with

$$A''_j = \sum_{i=1,\ldots,n} \hat{w}^i_j A^i_j \qquad (8)$$

where $\hat{w}^i_j$ is the normalized weight and δ_{A_j} is a constant (termed the shift factor of A_j), defined respectively by

$$\hat{w}^i_j = \frac{w^i_j}{\sum_{t=1,\ldots,n} w^t_j} \qquad (9)$$

$$\delta_{A_j} = \frac{|\text{Rep}(A^*_j) - \text{Rep}(A''_j)|}{\max_{A_j} - \min_{A_j}}. \qquad (10)$$

The consequent value of the intermediate rule is constructed in the same manner as above

$$Z' = Z'' + \delta_z(\max_z - \min_z) \qquad (11)$$

where Z'' is the weighted aggregation of the consequent values of the n closest rules $Z^i, i = 1, \ldots, n$

$$Z'' = \sum_{i=1,\ldots,n} \hat{w}^i_z Z^i \qquad (12)$$

with $\hat{w}^i_z$ being the mean of the normalized weights associated with the antecedents $\hat{w}^i_j$ in each rule

$$\hat{w}^i_z = \frac{1}{m} \sum_{j=1}^{m} \hat{w}^i_j \qquad (13)$$

and $\max_z$ and $\min_z$ in (11) are the maximal and minimal values of the consequent attribute, and the shift factor δ_z of the consequent is the mean of $\delta_{A_j}, j = 1, \ldots, m$

$$\delta_z = \frac{1}{m} \sum_{j=1}^{m} \delta_{A_j}. \qquad (14)$$

3) Computation of Scale and Move Factors: The goal of a transformation process T in T-FRI is to scale and move an intermediate fuzzy set A'_j, such that the transformed shape and

representative value coincide with those of the observed value A_j^*. This process is implemented in two stages: First, scale operation from A_j' to $\hat{A}_j'$ (denoting the scaled intermediate fuzzy set), and second, move operation from $\hat{A}_j'$ to A_j^*. For this purpose, the required scale rate s_{A_j} and move ratio m_{A_j} are determined in this step. It computes and records all such scale rates and move ratios for use in the subsequent and final procedure to obtain the required consequent value. Unfortunately, it is difficult to have a generic closed form representation of these transformation factors as they are dependent upon the fuzzy membership functions used.

For this work, triangular fuzzy sets are used throughout. Given such a fuzzy set $A_j' = (a_{j1}', a_{j2}', a_{j3}')$, the scale rate s_{A_j} is

$$s_{A_j} = \frac{a_{j3}^* - a_{j1}^*}{a_{j3}' - a_{j1}'} \tag{15}$$

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

$$\hat{a_{j1}'} = \frac{(1 + 2s_{A_j})a_{j1}' + (1 - s_{A_j})a_{j2}' + (1 - s_{A_j})a_{j3}'}{3}$$

$$\hat{a_{j2}'} = \frac{(1 - s_{A_j})a_{j1}' + (1 + 2s_{A_j})a_{j2}' + (1 - s_{A_j})a_{j3}'}{3}$$

$$\hat{a_{j3}'} = \frac{(1 - s_{A_j})a_{j1}' + (1 - s_{A_j})a_{j2}' + (1 + 2s_{A_j})a_{j3}'}{3}. \tag{16}$$

Similarly, while dealing with triangular fuzzy sets, the move operation shifts the position of $\hat{A}_j'$ to becoming the same as that of A_j^*, while maintaining its representative value $\text{Rep}(\hat{A}_j')$. This is achieved using the move ratio as follows:

$$m_{A_j} = \begin{cases} \frac{3(a_{j1}^* - \hat{a_{j1}'})}{a_{j2}' - \hat{a_{j1}'}}, & \text{if } a_{j1}^* \ge \hat{a_{j1}'} \\ \frac{3(a_{j1}^* - \hat{a_{j1}'})}{a_{j3}' - \hat{a_{j2}'}}, & \text{otherwise.} \end{cases} \tag{17}$$

4) Scale and Move Transformation: After calculating the necessary scale and move factors (i.e., s_{A_j} and $m_{A_j}, j = 1, \dots, m$), this procedure completes the T-FRI process, deriving the required consequent of Z^*. This follows the intuition of similar observations leading to similar consequents, a heuristic fundamental to analogical approximate reasoning. For this, the transformation factors on the antecedent attributes are aggregated. In the conventional T-FRI, this is implemented by averaging them as

$$s_z = \frac{1}{m}\sum_{j=1}^{m} s_{A_j} \qquad m_z = \frac{1}{m}\sum_{j=1}^{m} m_{A_j}. \tag{18}$$
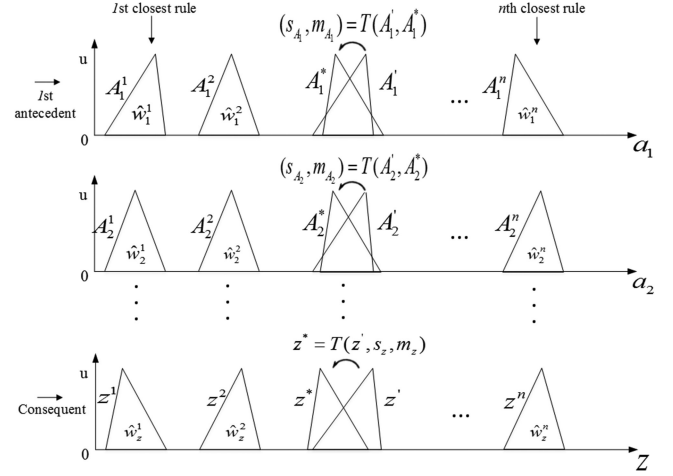

Fig. 2. FRI via scale and move transformations.

This entails the computation of scaled $\hat{Z}' = (\hat{z_1'}, \hat{z_2'}, \hat{z_3'})$ as

$$\hat{z_1'} = \frac{(1 + 2s_z)z_1' + (1 - s_z)z_2' + (1 - s_z)z_3'}{3}$$

$$\hat{z_2'} = \frac{(1 - s_z)z_1' + (1 + 2s_z)z_2' + (1 - s_z)z_3'}{3}$$

$$\hat{z_3'} = \frac{(1 - s_z)z_1' + (1 - s_z)z_2' + (1 + 2s_z)z_3'}{3} \tag{19}$$

where $Z' = (z_1', z_2', z_3')$ is the fuzzy value of the intermediate consequent previously computed. From this, again, by analogy to the transformation required for the antecedent to match the observation, move transformation is applied, resulting in the final required interpolated consequent $Z^* = (z_1^*, z_2^*, z_3^*)$ as

$$z_1^* = \hat{z_1'} + m_z\gamma$$

$$z_2^* = \hat{z_2'} - 2m_z\gamma \qquad \gamma = \begin{cases} \frac{\hat{z_2'} - \hat{z_1'}}{3}, & \text{if } m_z \ge 0 \\ \frac{\hat{z_3'} - \hat{z_2'}}{3}, & \text{otherwise} \end{cases}$$

$$z_3^* = \hat{z_3'} + m_z\gamma. \tag{20}$$

The entire scale and move transformation process can be graphically illustrated as shown in Fig. 2. For conciseness, such a process can be collectively represented by: $Z^* = T(Z', s_z, m_z)$, emphasizing on the significance of both scale and move transformations.

B. Attribute Evaluation Within FS

FS aims to choose a minimal subset of domain attributes that are the most relevant to the target concept or decision. It preserves the original meaning of the selected attributes while reducing their overall dimensionality. In FS, an evaluation function is used to measure how good a subset of attributes are regarding the potential solution to the problem at hand, if they are utilized. This offers a natural way to evaluate the relative significance of an attribute. If systematically carried out across all domain attributes, the use of such a function will enable the ranking of the attributes with regard to the underlying quality criteria. Existing evaluation functions in the literature can be

generally grouped into categories that reflect the criteria adopted to judge attribute quality, including those based on measures over distance, information, dependence, consistency, etc. [29]. The following presents a brief introduction to five of these that are popularly used and readily available, which will be adopted to implement the attribute ranking task in the subsequent development.

1) Information Gain (IG): IG is defined via Shannon entropy in information theory to measure the expected reduction in uncertainty caused by partitioning the values of an attribute [35], [42]. Given a collection of examples $U = \{O, A\}$, $o_i \in O$ is an object that is represented with a group of attribute $A = \{a_1, \ldots, a_l\}$ and a certain class label i. Shannon entropy of O is defined by

$$H(O) = -\sum_{i=1}^{m} p_i \log_2 p_i \qquad (21)$$

where p_i is the proportion of O whose elements are each labeled as the class i. IG upon a particular attribute $a_k, k \in \{1, \ldots, l\}$, is then defined by

$$\text{IG}(O, a_k) = H(O) - \sum_{v \in \text{Value}(a_k)} \frac{|O_v|}{|O|} H(O_v) \qquad (22)$$

where $\text{Value}(a_k)$ is the set of all possible values for the attribute a_k, O_v is the subset of O where the value of a_k is equal to v, and $|\cdot|$ denotes the cardinality of a set. Obviously, a quality attribute should lead to a high IG value.

2) Relief-F: Relief-F [36] works by exploiting distance measures. Each individual attribute is assigned a cumulative weight computed over a predefined number of sample data selected from a given training dataset. Attributes with a weight above a certain threshold become selected elements of the attribute subset sought. A weight is assigned on the basis of the following intuition: Instances that belong to a similar class should be closer together than those in a different class. Suppose that near_hit represents an instance that is closest to a certain training instance x under consideration, with both belonging to the same class, and that near_miss represents an instance that is closest to x but in a different class. The cumulative weight associated with a given attribute is then computed by

$$w_i = w_{i-1} - d(x, \text{near\_hit})^2 + d(x, \text{near\_miss})^2, i = 1, \ldots, I \qquad (23)$$

where $w_0 = 0$, I stands for the number of training iterations, and $d(., .)$ is typically implemented with Euclidean metric.

3) Laplacian Score (LS): LS [37] is another distance measure-based evaluation function. It is calculated for each individual attribute to reflect its capability of locality preserving. The definition of an LS is inspired by an observation that the data points being related to the same topic should be close to each other. Let LS_k denote the LS measure of a certain attribute a_k. Then, it is computed by

$$\text{LS}_k = \frac{\sum_{ij} (f_{ki} - f_{kj})^2 S_{ij}}{\text{Var}(f_k)} \qquad (24)$$

where f_{ki} and f_{kj} denote the value of a_k within the instance x_i and that within x_j, respectively, $\text{Var}(f_k)$ is the estimated variance of a_k, and S_{ij} represents the neighborhood relationship between the instances x_i and x_j, such that

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}, & \text{if } x_i \text{ and } x_j \text{ are nearest neighbors} \\ 0, & \text{otherwise.} \end{cases} \qquad (25)$$

A quality attribute should be of a small LS.

4) Local Learning-Based Clustering for FS (LLCFS): LL-CFS [34] performs attribute selection within the framework of the local learning-based clustering algorithm [43]. It computes a weight and assigns it to each attribute while performing a clustering task. Typically, the weights are thinly distributed if the dataset contains much redundancy, with a weight of zero indicating that the corresponding attribute is dispensable; only those attributes associated with a weight of a significant magnitude are selected. Incidentally, such an FS approach is termed wrapper-based in the literature, as opposed to the other techniques outlined herein which follow the so-called filter-based approach to FS [30]. LLCFS works by iteratively executing the following two steps until convergence: first, estimating the weights for the attributes using the intermediate clustering result, and second, updating the clustering given the weighted attributes. As such, the weights are estimated iteratively during the clustering process.

5) Rough Set-Based FS (RSFS): As a dependence measure-based attribute reduction method, RSFS [44] discovers the dependencies between attributes using the granularity structure inherent in data. Given the attribute subsets P and Q, the dependence degree of Q on P is defined as

$$\gamma_P(Q) = \frac{|\text{POS}_P(Q)|}{|U|} \qquad (26)$$

where U is a nonempty set of finite objects and $\text{POS}_P(Q)$ is termed the positive region, which is defined by

$$\text{POS}_P(Q) = \bigcup_{X \in U/\text{IND}(Q)} P_*(X)$$

$$P_*(X) = \bigcup_{x \in U} \{[x]_P : [x]_P \subseteq X\} \qquad (27)$$

where $X \in U/\text{IND}(Q)$, representing one of the equivalence classes partitioning U though the Q-indiscernibility relation

$$\text{IND}(Q) = \{(x, y) \in U^2 | \quad \forall a \in Q, a(x) = a(y)\}.$$

$P_*(X)$ determines the *P-lower approximation* of X in rough set theory, which is the union of the equivalence classes of the P-indiscernibility relations that are completely included in X. The positive region so defined contains all objects of U that can be classified as the classes of $U/\text{IND}(Q)$ using only the information conveyed by those attributes in P.

As can be seen from the above, FS methods using IG, Relief-F, LS, and LLCFS directly weigh and hence, rank features (and they may follow a filter or wrapper based approach). However, RSFS takes a different scheme where the quality of an attribute subset is evaluated at a time, instead of that of an individual
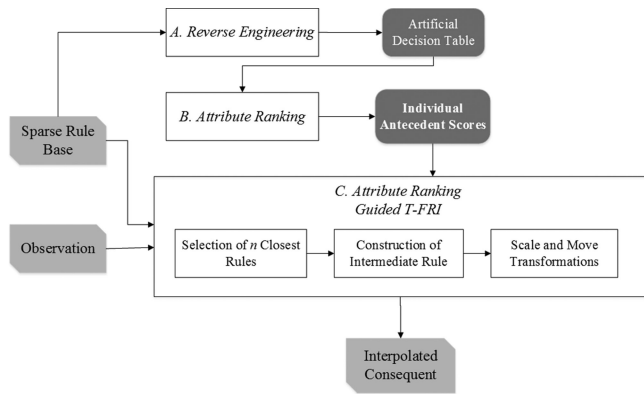
Fig. 3. Attribute ranking-supported T-FRI, with attribute rankings used in all key procedures of conventional T-FRI.

attribute. These different styles of FS mechanism are all considered here in order to demonstrate the generality of the proposed approach, as illustrated below.

III. FRI GUIDED BY ATTRIBUTE RANKING

This section presents a novel approach for FRI that is guided by attribute ranking techniques, with the framework as illustrated in Fig. 3. Note that any of the five different FS methods outlined in Section II-B can be employed to perform the ranking, in order to obtain the relative significance of individual attributes.

A. Reverse Engineering for Sparseness Reduction

In conventional T-FRI algorithms, the first key stage is the selection of n closest fuzzy rules when an observation is presented with no direct matching rules available in the rule base. In such work, all rule antecedents are assumed to be of equal significance while searching for the subset of closest rules; there is no assessment regarding the relative importance or ranking of these antecedents. This may reflect a seemingly important practical issue in which typically the fuzzy rules that are provided by domain experts or learned from historical data (which constitute the rule base) are of the form as shown in (1). That is, there is no information available on the relative significance of individual antecedent attributes. This is a premier reason that existing approaches to T-FRI commonly assume the use of this format of knowledge representation.

Fortunately, the evaluation functions embedded in the FS techniques offer an effective ranking mechanism to address this problem. However, while utilizing the evaluation function of a certain FS method to differentiate the significance of attributes, data is required to act as the training instances for computing the ranking scores. Yet, in general, at the stage of performing FRI, no sufficient example data are available for use to facilitate such computation. Nevertheless, every T-FRI system has a sparse rule base as indicated in Section II-A. This set of rules can be translated into a man-made decision table, forming a set of artificially created training samples, where each row represents either a rule in the given rule base or an artificial rule generated

from a given rule. Note that in data-driven learning, rules are learned from data samples. The work here is done through a reverse engineering process of data-driven learning, translating rules back to data.

1) Reverse Engineering Procedure: The question is how to create such artificial rules. In general, a fuzzy reasoning system with a sparse rule-base may involve rules that employ different antecedent attributes and a different number of antecedent attributes in different rules. To be able to systematically implement the reverse engineering procedure to obtain a training decision table, all rules are reformulated into a common representation using the following procedure: First, all possible antecedent attributes that appear in any given rule are identified, together with the value domains of these attributes. Then, each given rule is expanded iteratively into one that involves all domain attributes. The expansion is implemented such that if a certain antecedent attribute is not originally involved in a given rule, then that rule is replaced by q artificial rules, with q being the cardinality of the value domain of that attribute. In so doing, each expanded rule involves all domain attributes and each attribute in the rule takes one and only one possible value from its domain.

This reverse engineering procedure can be logically justified: For a given rule in the sparse rule base, if an attribute is missing from the rule antecedent, then the rule will have the same consequent value independent of what fuzzy value that attribute may take, provided that all those attributes appearing in the rule are satisfied regarding their respectively specified value. The presumption of the value domains being finite and discrete is also justifiable, given that only fuzzy rules are considered here, where each attribute takes values from a (normally small) collection of fuzzy sets. In particular, the proposed reverse engineering procedure works with a sparse rule base, which typically involves a much smaller number of rules than the usual fuzzy rule-based systems. Besides, only those missing antecedents are to be filled with the possible fuzzy sets taken from their value domains. These factors jointly help restrain the adverse impact of the curse of dimensionality possibly caused by converting individual rules in the sparse rule base into artificial training samples. Fundamentally, it is recognized, however, that in so doing, the underlying problem may be significantly reduced but not completely removed and therefore, work remains to develop a more efficient mechanism in implementing this approach.

2) Illustration of Reverse Engineering: A simple example may help illustrate the idea of this procedure. Suppose that the sparse rule base consists of the following two rules only, each involving one different antecedent attribute, x or y, and the common consequent attribute z as follows:

r_1: if x is A_1, then z is C_1
r_2: if y is B_2, then z is C_2

where x takes values from the domain $\{A_1, A_2\}$ and y from $\{B_1, B_2, B_3\}$.

Following the two-step reverse engineering procedure, first, all possible antecedent attributes involved in the problem are identified, these are x and y, together with their value domains as indicated above. Then, the artificial decision table as of Table I can be constructed. This is because there are two

TABLE I
EXAMPLE REVERSE ENGINEERED DECISION TABLE

Variables Artificial Rules	x	y	z
r^1	A_1	B_1	C_1
r^2	A_1	B_2	C_1
r^3	A_1	B_3	C_1
r^4	A_1	B_2	C_2
r^5	A_2	B_2	C_2

antecedent attributes in question, of which x has two possible values (A_1 and A_2) and y has three alternatives (B_1, B_2, B_3). Without losing generality, suppose that the first given rule is used to construct a part of the emerging artificial decision table first. As y is missing in r_1, which means if x is satisfied (with the value A_1), this rule is satisfied and hence, the consequent attribute z will have the value C_1 no matter which value y takes. That is, r_1 can be expanded by three artificial rules, resulting in r^1, r^2, and r^3 in Table I, for each of which y takes one of its three possible values. Similarly, r^4 and r^5 can be constructed to expand the original rule r_2.

3) Inconsistency in Artificial Decision Table: When the reverse engineering procedure is applied to a given (sparse) rule base, the resultant artificially constructed decision table may include logically inconsistent rules where certain rules may have the same antecedent but different consequents. For instance, in the above illustrative example, r^2 and r^4 in Table I may appear to be inconsistent. This does not matter as the eventual rule-based inference, including rule interpolation, does not use these artificially generated rules, but the original sparse rule base. They are created just to help assess the relevant significance of individual variables through the estimation of their respective ranking scores. It is because there are attributes which may lead to potentially inconsistent implications in a given problem that it is possible to distinguish their relative importance to the problem, or their potential power in influencing the derivation of the consequent.

B. Scoring of Individual Attributes

Suppose that an artificial decision table has been derived from a given sparse rule base via reverse engineering. Then, any of the five feature ranking methods reviewed in Section II-B may be applied to evaluate the relative significance of individual antecedent attributes.

1) Scoring Methods: As indicated previously, four of those (namely, IG, Relief-F, LS, and LLCFS) can be directly applied to assess individual attributes, each resulting in a vector of weighting scores associated with those attributes. For easy referencing, these score vectors are denoted as Score_{IG}, $\text{Score}_{\text{Relief-F}}$, Score_{LS}, and $\text{Score}_{\text{LLCFS}}$, respectively. Note that the LS-based FS method seeks those attributes of the smallest LS(s) for selection. Thus, the ranking score of the LS for a rule antecedent attribute $a_i, i = 1, \ldots, m$, can be defined by

$$\text{Score}_{\text{LS}}(a_i) = \frac{1}{1 + \text{LS}_i}. \tag{28}$$

Also as indicated earlier, the RSFS method conducts FS based on evaluating attribute subsets, instead of individual attributes. To obtain individual attribute scores using such a technique, the evaluation procedure needs to be modified, which is done in this paper as follows. It is known that the dependence degree $\gamma_P(Q)$ captures the dependence of an attribute subset Q on another subset P. Suppose that the subset Q contains the single consequent attribute z and the subset P contains just one certain antecedent attribute $a_i, i = 1, \ldots, m$ of a rule in the sparse rule base. As such, the general form of the dependence degree $\gamma_{\{a_i\}}(\{z\})$ between two subsets of attributes as per (26) degenerates to the one that assesses the importance degree of each individual antecedent attribute upon which the consequent depends

$$\text{Score}_{\text{RSFS}}(a_i) = \gamma_{\{a_i\}}(\{z\}) = \frac{\left|\text{POS}_{\{a_i\}}(\{z\})\right|}{|U|}. \tag{29}$$

This is of course, what RSFS exactly does in the first round during its iterative process of adding attributes to the emerging selected feature subset (starting from an empty set), determining which attribute is individually speaking, the best to be selected. It means that to obtain attribute scoring vector using the evaluation function of RSFS, only one iteration of the FS algorithm is needed to be run.

2) Attribute Weighting: Having computed the scores of individual attributes, using either of the aforementioned five scoring methods, a normalized relative weighting scheme can be readily introduced. Thus, all antecedent attributes employed in the rules of a given sparse rule base can be ranked, each (say, the attribute a_i) being associated with a weight

$$W_i = \frac{\text{Score}_*(a_i)}{\sum_{t=1,\ldots,m} \text{Score}_*(a_t)} \tag{30}$$

where Score_* denotes any of the five types of score (namely, one of the following: Score_{IG}, $\text{Score}_{\text{Relief-F}}$, Score_{LS}, $\text{Score}_{\text{LLCFS}}$, and $\text{Score}_{\text{RSFS}}$).

Given their underlying definition, the resulting normalized values have a natural appeal to be interpreted as the relative significance degrees of the individual rule antecedent attributes, in the determination of the corresponding rule consequent. Therefore, they can be used to act as the weights associated with each individual antecedent attribute in the original sparse rule base. Of course, for any implementation in modifying conventional nonweighted T-FRI, one and just one of the five types of the weight is required. From this viewpoint, this paper presents a range of choices regarding the weighting methods that may be utilized to support and refine fuzzy interpolative reasoning, as described in the following.

C. Weighted T-FRI

From the above, weights can be computed and associated with rule antecedent attributes to indicate their relative significance in deriving the consequent. From this, T-FRI can be modified as shown in Fig. 3, involving the following three key computational stages.

1) Weight-guided Selection of n Closest Rules: Suppose that an observation is present which does not entail a direct match

with any rule in the sparse rule base. Thus, a neighborhood of n ($n \geq 2$) closest rules of the observation is required to be chosen in order to perform rule interpolation. The conventional approach to making this choice is by exploiting the Euclidean distance measured through aggregating the distances between individual antecedent attributes of a given rule and the corresponding attribute values in the observation, as per (2). Now that the weights of individual attributes have been obtained with a scoring mechanism (derived from the use of the evaluation function in an FS method), the distance between a given rule r^p and the observation o^* needs to be updated accordingly, such that

$$
\begin{aligned}
\tilde{d}(r^p, o^*) &= \frac{1}{\sqrt{\sum_{t=1}^{m}\left(\frac{1-W_t}{m-1}\right)^2}} \sqrt{\sum_{j=1}^{m}\left(\left(\frac{1-W_j}{m-1}\right)d(A_j^p, A_j^*)\right)^2} \\
&= \frac{1}{\sqrt{\sum_{t=1}^{m}(1-W_t)^2}} \sqrt{\sum_{j=1}^{m}\left((1-W_j)d(A_j^p, A_j^*)\right)^2}
\end{aligned}
\tag{31}
$$

where $d(A_j^p, A_j^*)$ is calculated according to (3) and m is the total number of rule antecedents in the rule base and $m \geq 2$ (since there is no need to assign any weight if all rules in the rule base involves just the same single antecedent attribute). The term $(m-1)$ in the first part of this formula is for local weight normalization purpose, but it is cancelled out in the overall equation. In so doing, those n closest rules whose antecedent attributes are deemed more significant (than the rest) will be selected with priority. This is because such attributes will make less contribution (i.e., $(1-W_j)d(A_j^p, A_j^*), j = 1, \ldots, m$) to the aggregated distance $\tilde{d}(r^p, o^*)$ given their relatively larger weight values.

The computation of the distance $\tilde{d}(r^p, o^*)$ is carried out to measure the relative closeness of the rules to the observation. Under the condition where there is no rule matching the given observation, the attribute ranking-supported T-FRI is triggered. Hence, the aggregated distance is calculated in terms of (31) between the individual elements of the observation and each rule antecedent, respectively. From this, those n rules that have resulted in the n smallest distance values are selected.

Note that the normalization term $\frac{1}{\sqrt{\sum_{t=1}^{m}(1-W_t)^2}}$ in the above is a constant and therefore, can be omitted in the process of executing FRI. This is because selecting the closest rules only requires information on the relative distance measures.

2) Weighted Construction of Intermediate Rule: With the weighting method introduced previously, all antecedent attributes can be ranked with respect to their estimated relative significance, in terms of their potential implication upon the derivation of the (interpolated) consequent. This allows for the development of a computational method to implement an improved version of T-FRI, where weights are integrated in all calculations during the transformation process, including the initial construction of the intermediate rule. Without unnecessarily detailing the entire construction process of the weighted intermediate rule, which is similar to that of the conventional

approach (see Section II-A2), only the weighting on the consequent and the shift factor during the modified process are presented as follows:

$$
\tilde{\hat{w}}_z^i = \sum_{j=1}^{m} W_j \hat{w}_j^i, \qquad \tilde{\delta}_z = \sum_{j=1}^{m} W_j \delta_{A_j}.
\tag{32}
$$

Obviously, these will degenerate to those computed as per (13) and (14), when all attributes are equally regarded in terms of their significance.

3) Weighted Transformation: Given the above method for constructing the weighted intermediate rule, the scale and move factors originally provided in (18) now become

$$
\tilde{s}_z = \sum_{j=1}^{m} W_j s_{A_j}, \qquad \tilde{m}_z = \sum_{j=1}^{m} W_j m_{A_j}.
\tag{33}
$$

From this, if an observation that does not match any rule in the sparse rule base is presented, an interpolated fuzzy value Z^* for the consequent attribute can be obtained by computing the transformation $T(\tilde{Z}', \tilde{s}_z, \tilde{m}_z)$, in the exactly same way as given in Section II-A4. Importantly, when all antecedent attributes are assumed to be of equal significance, namely when all weights are equal, the above modified fuzzy rule-based interpolative process degenerates to the conventional T-FRI. Mathematical proof for this is straightforward but is omitted here to save space.

Note that in the above description, no specification of which attribute ranking mechanism to use is made. Indeed, the proposed technique is independent of the FS method to be adopted for attribute scoring. Any of the attribute ranking methods available may be taken to assess the relative significance of individual antecedent attributes. Thus, the proposed ranking-guided FRI offers flexibility in its implementation. Section II-B has outlined five effective attribute evaluation functions (that are used in popular FS systems), each of which may be adopted to implement the ranking mechanism.

IV. EXPERIMENTAL EVALUATION

This section presents a systematic experimental evaluation of the proposed approach for T-FRI supported by attribute ranking. It first reports on the results of performing pattern classification over ten benchmark datasets. Classification results are compared with those obtained by: first, the state-of-the-art T-FRI; and second, the standard rule-based reasoning via the application of CRI [3], without involving rule interpolation but directly firing those (fully or partially) matched rules. Then, the robustness and effectiveness of the new approach is also empirically demonstrated by observing the following.

1) The analysis of confusion matrices obtained for a specified case study.
2) The classification accuracy in relation to the number of the closest rules selected for interpolation.
3) The consistency and efficiency of utilizing different FS methods in supporting T-FRI.

Finally, the improvement of the classification performance following the weighted approach is further illustrated by fine tuning the experimental settings.

TABLE II
DATASETS USED FOR CLASSIFICATION

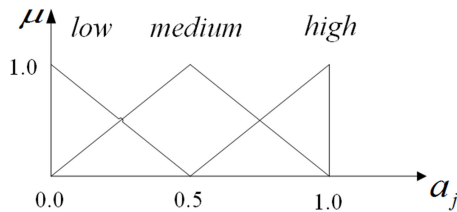| Dataset	Attributes #	Classes #	Instances #
Iris	4	3	150
Diabetes	8	2	768
Phoneme	5	2	5404
Appendicitis	7	2	106
Magic	10	2	1902
NewThyroid	5	3	215
Banana	2	2	5300
Haberman	3	2	306
BUPA	6	2	345
Hayes–Roth	4	3	160

Fig. 4.   Membership functions defining values of antecedent attributes.

A. General Experimental Setup

1) Datasets: Ten benchmark datasets (for classification problems) are taken from the UCI machine learning [45] and knowledge extraction based on evolutionary learning dataset repositories [46]. The details of these are summarized in Table II.

2) Experimental Methodology: As indicated previously, for simplicity, triangular membership functions are employed here unless otherwise stated. They are used to represent the fuzzy values of the antecedent attributes. For each problem, the consequent attribute is designed to take a singleton fuzzy set (which is equivalent to a discrete crisp value), representing a certain class label. While different antecedent attributes have their own underlying value domains, these domains are normalized to be within the common range of 0 to 1 and consisting of three qualitatively distinct fuzzy values, as shown in Fig. 4. Such a simple fuzzification is used in the main body of the experiments for simplicity as well as for fair comparison, with no optimization of the value domain carried out. Of course, if fine-tuned membership functions are available and used, the classification performance can be expected to further improve (as to be illustrated later).

Experimental results are obtained by averaging the outcomes of $10\times$ ten-fold cross validation per dataset. The rules used to perform both CRI and interpolative reasoning are learned from the raw data by the use of the classical method of [47], after fuzzification. This rule induction technique is employed, herein forming a common ground for fair comparison. However, if preferred, more advanced rule induction mechanisms (e.g., those implemented with evolutionary or memetic algorithms [48]) may be exploited to produce a more compact rule base (but this is beyond the scope of this paper).

On average, 20% of the learned rules are purposefully removed randomly in order to make the resultant rule base sparser,

and hence, to validate the effectiveness of rule interpolation. Attribute weights are derived from the use of one and each of those five ranking methods introduced previously, using the artificial training data generated from the sparse rule base via the reverse engineering procedure. Classification performance is assessed in terms of accuracy over the testing data. In each test, a testing sample is checked against the rules within the rule base first. If there is no rule matching the observation, FRI is applied to make inference, using both the conventional T-FRI and the attribute ranking-supported T-FRI to facilitate comparison.

The main body of this experimental study is based on the use of $n = 2$ closest rules to perform rule interpolation. However, a series of experiments are also carried out by varying the number of the closest rules selected for interpolation (see Section IV-B3). In particular, $10\times$ ten-fold cross validation is adopted for each of the five different cases where the number of the closest rules selected is set to 2, 3, 4, 5, 6, respectively.

B. Results and Discussions

1) Classification Accuracy: Table III shows the average classification accuracies, and standard deviations (SD), which are calculated by averaging the $10\times$ ten-fold cross validation, for each of the seven compared approaches. In this table, CRI is the column showing the accuracies achievable using CRI based on the sparse rule base; Ori lists the accuracies obtained using the conventional T-FRI, with the rest naming schemes used being obvious and self-explanatory (e.g., IG stands for the accuracies achieved by the proposed approach with the antecedent attributes in the rules weighted by their corresponding IGs); and AVG_Guided presents the accuracies obtained by averaging the performances of the five attribute ranking-guided T-FRI methods.

The comparison with CRI is included herein to demonstrate the power of FRI in general and that of ranking-guided FRI in particular in performing fuzzy reasoning, both of which significantly outperform the use of CRI in all the problems that involve a sparse rule base. This may be expected since a fuzzy system implemented with CRI alone cannot draw any conclusion when an observation does not match any of the rules in the rule base. As already indicated, no attempt is made to optimize the fuzzification of any attribute domains. Thus, the classification rates are generally not very impressive. However, this is not the point of this experimental investigation. The point is to compare the relative performances of different approaches, with a common ground for fair comparison. The improvement achievable by employing learned membership functions (from training samples) will be shown later.

The use of any of the five attribute ranking-guided methods has been shown to enable the corresponding fuzzy reasoning system to outperform the system using the conventional T-FRI. This indicates that individual rule antecedent attributes do make different contributions to the classification, and that the ranking scores obtained by FS techniques offer positive means for discovering such differences. Interestingly, the narrow-banded SD values (those numbers following the classification accuracy)

TABLE III
AVERAGE CLASSIFICATION ACCURACIES (%) WITH SD IN $10\times$ TEN-FOLD CROSS VALIDATION

Dataset	CRI	Ori	IG	Relief-F	LLCFS	LS	RSFS	AVG_Guided
Iris	42.66 ± 0.11	79.33 ± 0.05	90.66 ± 0.03	92.66 ± 0.03	91.33 ± 0.04	90.00 ± 0.04	91.33 ± 0.03	91.20 ± 0.03
Diabetes	36.71 ± 0.08	58.59 ± 0.06	66.66 ± 0.05	63.54 ± 0.04	61.97 ± 0.04	61.84 ± 0.07	58.98 ± 0.05	62.60 ± 0.05
Phoneme	30.66 ± 0.08	57.10 ± 0.06	67.33 ± 0.05	64.78 ± 0.05	64.59 ± 0.05	60.47 ± 0.07	61.67 ± 0.07	63.77 ± 0.06
Appendicitis	32.08 ± 0.10	52.00 ± 0.16	69.72 ± 0.15	66.91 ± 0.19	57.72 ± 0.17	59.45 ± 0.14	66.72 ± 0.17	62.11 ± 0.16
Magic	34.76 ± 0.09	55.84 ± 0.05	64.35 ± 0.03	59.77 ± 0.05	58.94 ± 0.06	60.89 ± 0.06	55.84 ± 0.04	59.96 ± 0.04
NewThyroid	53.87 ± 0.18	53.87 ± 0.18	58.09 ± 0.17	58.61 ± 0.16	70.32 ± 0.22	56.21 ± 0.18	55.77 ± 0.17	59.40 ± 0.18
Banana	36.01 ± 0.08	56.41 ± 0.06	59.75 ± 0.04	57.83 ± 0.04	58.53 ± 0.04	57.71 ± 0.03	59.73 ± 0.04	58.71 ± 0.04
Haberman	55.44 ± 0.10	74.26 ± 0.11	78.83 ± 0.08	79.15 ± 0.11	81.11 ± 0.09	78.50 ± 0.09	79.15 ± 0.12	79.35 ± 0.10
BUPA	19.43 ± 0.07	48.72 ± 0.09	62.03 ± 0.08	58.84 ± 0.09	57.35 ± 0.11	55.69 ± 0.09	55.40 ± 0.09	57.86 ± 0.09
Hayes-Roth	36.87 ± 0.11	46.87 ± 0.10	60.00 ± 0.11	60.62 ± 0.13	54.37 ± 0.11	56.25 ± 0.11	58.75 ± 0.12	58.00 ± 0.12
Average	37.84 ± 0.10	58.30 ± 0.10	67.74 ± 0.08	66.27 ± 0.08	65.63 ± 0.09	63.71 ± 0.08	64.34 ± 0.09	65.54 ± 0.08

given in Table III further demonstrate that the performance of the proposed work is robust.

Examining more closely, those methods based on directly assessing individual attributes (namely, IG, Relief-F, LLCFS, and LS) achieve more significant improvements, with the best average accuracy being obtained by IG-guided T-FRI (having an average improvement of 9.44% over all ten datasets than that of Ori). The remaining one, RSFS, adopts the technique of (attribute) subset selection. As shown in Section III-B, ranking attributes with such a technique requires modification of the underlying FS algorithm. Nevertheless, the RSFS-based FRI has a comparable improvement over the conventional T-FRI to the average performance of the other four, again indicating the robustness of the innovative approach proposed in this paper. Collectively, these results also show the generality of attribute ranking-guided approach in which the use of a very different FS method retains the improved performance.

As also can be seen from Table III, both FRI approaches (the original and the attribute ranking-guided) significantly outperform the standard fuzzy reasoning based on CRI, and the results are more stable with a relatively lower SD values. Of course, such an obviously poorer classification accuracy obtained by the use of CRI can be expected as it fires matched or partially matched rules only while facing the problem of sparseness of the rule base. This strongly demonstrates the effectiveness of fuzzy interpolative reasoning, especially for the proposed approach owing to its further enhanced performance over the conventional T-FRI.

2) Analysis on Confusion Matrices: Apart from the overall classification accuracies, it is practically interesting to investigate the statistical properties of the classification performance in terms of *true positive* (TP), *true negative, false positive*, and *false negative* (FN). Without overwhelming the examination while having a focused discussion, the *Haberman* dataset is taken as an example to run such an investigation. Tables IV–X show the confusion matrices computed by the use of seven comparative approaches (averaged over 10×10 fold cross validation), respectively. Table XI lists the averaged performance of the five different implementations of the attribute ranking-guided method. Despite the fact that this dataset contains samples that are distributed in a imbalanced manner (which increases the difficulties in performing accurate classification), these tables

TABLE IV
CONFUSION MATRIX OF CRI

		Classified	
		Positive	Negative
Actual	Positive	50.16%	26.40%
	Negative	18.15%	5.28%

TABLE V
CONFUSION MATRIX OF ORIGINAL T-FRI

		Classified	
		Positive	Negative
Actual	Positive	68.40%	8.79%
	Negative	16.93%	5.86%

TABLE VI
CONFUSION MATRIX OF IG-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	72.64%	4.56%
	Negative	16.61%	6.19%

TABLE VII
CONFUSION MATRIX OF RELIEF-F-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	72.64%	4.88%
	Negative	15.96%	6.51%

clearly show the superior performances achieved by the proposed approach to the original T-FRI, leaving alone the CRI.

Importantly, these tables both individually and collectively reveal that the classification accuracy achieved by the use of the attribute ranking-guided T-FRI is led by the significant

TABLE VIII
CONFUSION MATRIX OF LLCFS-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	75.90%	3.58%
	Negative	15.31%	5.21%

TABLE IX
CONFUSION MATRIX OF AN LS-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	72.32%	4.88%
	Negative	16.61%	6.18%

TABLE X
CONFUSION MATRIX OF RSFS-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	72.96%	4.56%
	Negative	16.28%	6.19%

TABLE XI
CONFUSION MATRIX OF AVG_GUIDED-T-FRI

		Classified	
		Positive	Negative
Actual	Positive	73.29%	4.49%
	Negative	16.15%	6.06%

TABLE XII
AVERAGE CLASSIFICATION ACCURACY (%) VERSUS NUMBER OF CLOSEST
RULES USED FOR INTERPOLATION

Dataset	Methods	Number of Closest Rules (n)					SD
		2	3	4	5	6	over n
BUPA	Ori	48.72	53.95	53.05	51.88	51.57	1.98
	IG	62.03	57.95	54.74	53.91	51.56	4.05
	Relief-F	58.84	55.95	51.89	52.18	50.71	3.38
	LLCFS	57.35	57.10	56.52	54.18	54.74	1.43
	LS	55.69	55.68	55.68	53.05	51.03	2.11
	RSFS	55.40	55.37	55.97	51.86	50.99	2.30
Hayes–Roth	Ori	46.87	49.37	48.12	48.75	48.12	0.93
	IG	60.00	58.12	58.12	55.00	53.12	2.76
	Relief-F	60.62	56.25	56.25	55.00	51.25	3.35
	LLCFS	54.37	52.50	53.75	55.00	53.75	0.92
	LS	56.25	54.37	53.75	53.75	54.37	1.02
	RSFS	58.75	58.75	56.25	52.50	50.00	3.89
Appendicitis	Ori	52.00	52.18	53.00	51.09	52.09	0.68
	IG	69.72	62.09	62.18	62.18	62.18	3.38
	Relief-F	66.91	64.18	62.36	64.18	62.27	1.88
	LLCFS	57.72	56.81	56.81	55.91	55.81	0.78
	LS	59.45	55.63	54.72	53.81	53.91	2.32
	RSFS	66.72	63.81	62.91	60.18	60.18	2.74
Phoneme	Ori	57.10	54.16	57.54	58.91	59.19	2.01
	IG	67.33	64.93	63.45	64.63	65.08	1.40
	Relief-F	64.78	62.89	62.91	63.71	63.82	0.77
	LLCFS	64.59	61.56	60.99	60.65	61.02	1.61
	LS	60.47	61.28	60.28	61.47	62.43	0.86
	RSFS	61.67	61.34	61.76	61.82	60.39	0.59

reduction of FNs and the substantial increase in TPs. These results form a sharp contrast with those obtainable by the use of the original T-FRI and more remarkably, with those by CRI. This is of practical significance because for many real-world applications, not only the overall classification rates should be high, but also FNs should be minimized while TPs are maximized. This is of particular importance for medical applications as with the situation of this dataset (which summarizes the cases on the survival of patients who had undergone surgery for breast cancer—if a patient died within five year of the surgery, then the case is regarded as positive, or if the patient survived for five years or longer, then it is a negative case). For such problems, FNs can be extraordinarily damaging.

Fortunately, the implementations with the proposed approach, all lead to much reduced FNs (with an average rate of 4.49% over the range of 3.58% to 4.88%, as compared to 8.79% returned by the conventional T-FRI and 26.40% by CRI). This is in addition to the remarkable improvements over the TP rates (an average of 73.29% over the range of 72.32% to 75.90%, as opposite to 68.40% by the original T-FRI and a mere 50.16% by CRI).

3) Number of Closest Rules: Up till now, all experimental results reported in the existing literature regarding the use of T-FRI have been based on the use of two closest rules (i.e., $n = 2$) to perform interpolation. The choice of using two rules is for computational simplicity. Hypotheses have been given previously in which a larger neighborhood (i.e., more than 2 closest rules) may lead to generally more accurate interpolated outcomes. It is, therefore, interesting to investigate the level of change in classification accuracy with regard to varying the number of the closest rules selected for fuzzy rule-based interpolative reasoning.

Considering the computational effort required for such an experimental investigation, only a subset of the previously listed ten benchmark datasets (namely, BUPA, Hayes–Roth, Appendicitis, and Phoneme) are randomly used to conduct this study. Table XII presents the experimental results, with the summary of these plotted in Fig. 5. Again, the accuracies shown in this table are calculated by averaging the results obtained in $10\times$ ten-fold cross validation.

Over the range of $n, n \in \{2, \ldots, 6\}$ that are examined, running both the conventional T-FRI and the attribute ranking-supported T-FRI always results in a substantial improvement (in terms of the classification accuracy) over the performance achievable by running CRI with direct rule-firing (which is shown in Table III and is irrelevant to the n). Importantly, each of the proposed five attribute-guided T-FRI methods consistently outperforms the conventional T-FRI for almost all datasets and all settings of n. The results in Table XII further demonstrate the robustness of the proposed work given that the SD values
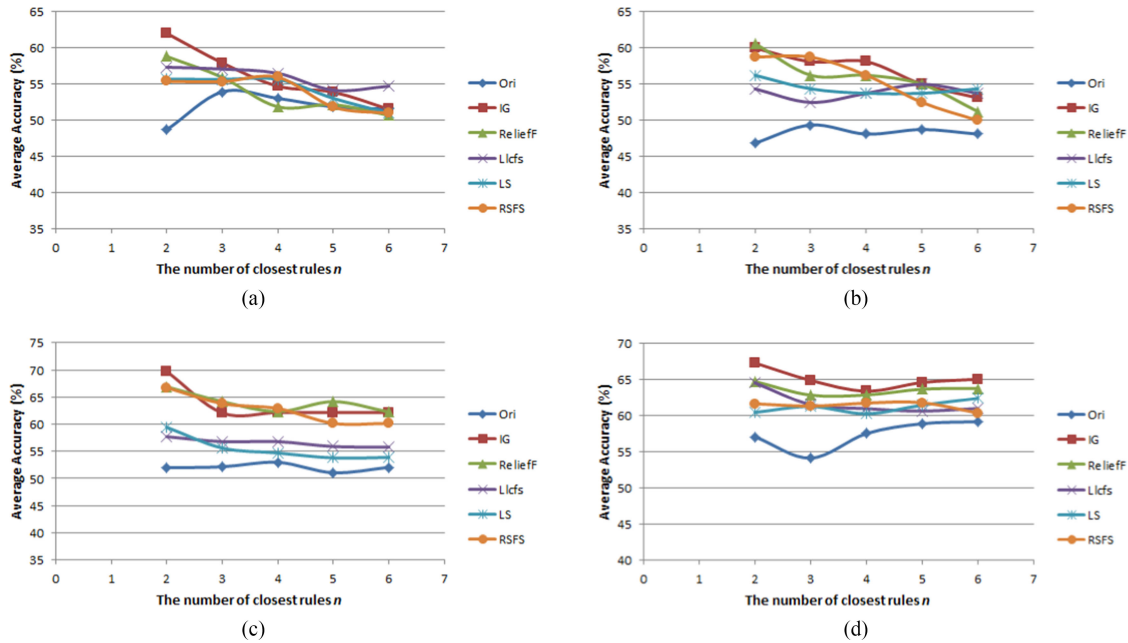
Fig. 5.   Accuracy variation with number of closest rules for four datasets: (a) BUPA. (b) Hayes–Roth. (c) Appendicitis. (d) Phoneme.

of the classification accuracy across all n values are rather small.

Surprisingly (and very positively in support of the present approach), as a larger n is assumed, little improvement can be gained for any of the five attribute ranking-supported methods. In fact, the performance may even deteriorate as n increases. The best performance is actually achieved when the number of selected closest rules is the smallest (i.e., 2). This indicates that the weighting scheme facilitates the determination of the best neighboring rules to be taken at the earliest opportunity. This result empirically negates the hypothesis commonly made about T-FRI in which more rules used for interpolation would lead to significantly better results. It also helps avoid the use of a larger n in applications of the weighted T-FRI, thereby, reducing the computational complexity that would otherwise be increased due to the requirement of searching for and running with more rules for interpolation.

4) Consistency and Efficiency of Ranking Methods: There is one exception in the above results regarding the *Phoneme* dataset where the classification accuracy achieved using an LS-guided T-FRI is eventually increasing as the number of closest rules goes up, though this variation is not significant. Therefore, a further investigation has been conducted to forensically examine the ranking scores that are obtained by the use of the five different evaluation functions. The results are presented in Table XIII.

As can be seen from this table, the first four attribute ranking methods consistently agree on that the fourth antecedent attribute plays the most significant role in deciding on the consequent, with a much higher ranking scores obtained. Three out of these (IG, Relief-F, and RSFS) put the first antecedent attribute in the second place, with LLCFS ranking it at the third place. The only one method which is out of the tune is LS,

TABLE XIII
ATTRIBUTE WEIGHTS AND RANKINGS USING DIFFERENT RANKING SCHEMES
FOR PHONEME DATASET

Methods	Antecedent Weights					Rankings
IG	0.2852	0.0792	0.0125	0.5724	0.0507	[4 1 2 5 3]
Relief-F	0.1326	0.0414	0	0.7286	0.0973	[4 1 5 2 3]
LLCFS	0.0001	0	0	0.7416	0.2583	[4 5 1 2 3]
RSFS	0.0016	0.0016	0.0016	0.9938	0.0016	[4 1 2 3 5]
LS	0	0.4541	0.0988	0.1995	0.2476	[2 5 4 3 1]

which ranks the first antecedent attribute at the bottom, with a zero score signifying its relative lack of relevancy in this rule base. This is a very different result from the great majority, implying that the LS algorithm may underperform in deriving an appropriate ranking for this particular dataset. As such, it may explain the reason that the FRI guided with the LS method achieves a relative poor performance when the number of closest rules is two and the overall different trend of the classification accuracy while varying n in this dataset case, as shown in Fig. 5.

The introduction of ranking scores of antecedent attributes in support of weighted rule interpolation may lead to additional computational overheads overall (albeit it ensuring that only the smallest number of closest rules are needed). Table XIV shows the corresponding average testing time recorded for classification over testing samples when the number of closest rules is increasing, together with the SD value over n. In this table, the column of Max Increase lists the maximum increase of the testing time observed while increasing the number of closest rules n.

Generally, there is a slight increase in time consumption when involving more closest rules in the implementation of rule interpolation for all T-FRI methods. However, while the attribute

TABLE XIV
AVERAGE TESTING TIME (SECOND) VERSUS NUMBER OF CLOSEST RULES

Dataset	Methods	Number of Closest Rules (n)					SD over n	Max Increase
		2	3	4	5	6		
BUPA	Ori	0.1584	0.1556	0.1587	0.1524	0.1624	0.0037	0.0100
	IG	0.1429	0.1445	0.1513	0.1473	0.1506	0.0036	0.0084
	Relief-F	0.1465	0.1448	0.1459	0.1538	0.1498	0.0036	0.0090
	LLCFS	0.1411	0.1473	0.1461	0.1463	0.1540	0.0046	0.0129
	LS	0.1420	0.1431	0.1446	0.1509	0.1502	0.0041	0.0089
	RSFS	0.1417	0.1433	0.1458	0.1464	0.1512	0.0036	0.0095
Hayes–Roth	Ori	0.0448	0.0425	0.0446	0.0439	0.0467	0.0015	0.0042
	IG	0.0386	0.0417	0.0413	0.0426	0.0425	0.0016	0.0040
	Relief-F	0.0394	0.0401	0.0403	0.0417	0.0408	0.0008	0.0023
	LLCFS	0.0404	0.0417	0.0414	0.0435	0.0426	0.0011	0.0031
	LS	0.0414	0.0415	0.0417	0.0422	0.0429	0.0006	0.0008
	RSFS	0.0406	0.0410	0.0420	0.0427	0.0426	0.0009	0.0021
Appendicitis	Ori	0.0334	0.0377	0.0391	0.0378	0.0394	0.0024	0.0060
	IG	0.0323	0.0350	0.0354	0.0364	0.0381	0.0021	0.0058
	Relief-F	0.0343	0.0347	0.0369	0.0368	0.0367	0.0012	0.0025
	LLCFS	0.0349	0.0352	0.0367	0.0375	0.0371	0.0011	0.0026
	LS	0.0347	0.0372	0.0368	0.0367	0.0369	0.0010	0.0022
	RSFS	0.0375	0.0368	0.0361	0.0369	0.0371	0.0005	0.0014
Phoneme	Ori	1.4078	1.4159	1.4448	1.4769	1.4911	0.0366	0.0833
	IG	1.3954	1.4335	1.4486	1.4728	1.4818	0.0343	0.0864
	Relief-F	1.3502	1.3798	1.3959	1.4303	1.4153	0.0312	0.0801
	LLCFS	1.3665	1.4047	1.4321	1.4328	1.4524	0.0332	0.0859
	LS	1.3440	1.3672	1.4023	1.4113	1.4256	0.0335	0.0816
	RSFS	1.3415	1.3900	1.3864	1.3958	1.4258	0.0302	0.0843

ranking-guided T-FRI employs the weights in all of the key stages of interpolation (including the selection of the closest rules, the construction of the intermediate rule, the calculation of weighted transformation factors, and the execution of weighted transformations), there is no significant increase in the time consumed by the weighted T-FRI as compared to that by the original T-FRI. This, together with the above observed general consistency amongst the use of the attribute ranking schemes, once again shows the efficacy of the proposed approach.

5) Use of Learned Membership Functions: As indicated previously, the classification performance in terms of accuracy is not very impressive, even though the proposed work improves it significantly over the conventional approaches. However, this is expected as the quantity space used to depict the value domains of all the attributes across all datasets is so simplistic (recall Fig. 4), without any optimization (which is purposefully designed so as to enable systematic investigations over a wide range of experimental settings). Such unbiased specification of the domain values allows fair comparison to be made between different fuzzy reasoning techniques. Besides, an average of 20% of the learned rules are deliberately removed randomly, in order to have a rule base that is rather sparse. This makes the domain knowledge, represented in terms of fuzzy rules, rather incomplete, which in turn, makes the classification task a challenge for any learning classifier and hence, leads to less accurate classification. Nevertheless, it is interesting to empirically verify what if an (at least partially) optimized quantity space is utilized.

Fuzzy C-Means (FCM) [49] is one of the most widely used fuzzy clustering algorithms. It works by assigning a membership degree to each data sample corresponding to a certain cluster center based on the relative distance between the cluster center and that sample. The closer to the cluster center, the higher the membership degree to which the sample is deemed to belong to the corresponding cluster. Thus, the clustering outcome on a given dataset reveals the distribution of the membership functions for the underlying attributes. Owing to its popularity, FCM is herein adopted to perform fuzzification, learning the membership functions for the antecedent attributes. However, any optimization of the membership functions is directly influenced by the dataset itself. Without overly complicating the experimental investigation, only the simple *Iris* dataset is used in this specific study (on the effect of using learned fuzzy sets). Fig. 6 shows the membership functions generated using FCM. The optimal number of clusters for each antecedent attribute is selected by the method of [50], resulting in four clusters for the first antecedent attribute, two for the third, and three for each of the remaining two.

Table XV presents the classification results using the FCM-returned membership functions. For comparison, it also lists those that are obtained by the use of evenly distributed fuzzification based on the entries given in Table III. As expected, a better fuzzification leads to a better classification. Individually speaking, each weighted method that uses FCM-learned membership functions beats their corresponding opponent (that employs just the simple quantity space of Fig. 4 for each antecedent attribute). Collectively, this leads to an averaged enhancement of 1.87% (=93.07% − 91.20%) for the FS-supported T-FRI methods. Importantly, this is on top of the already achieved substantial improvement of the FS-supported T-FRI over the
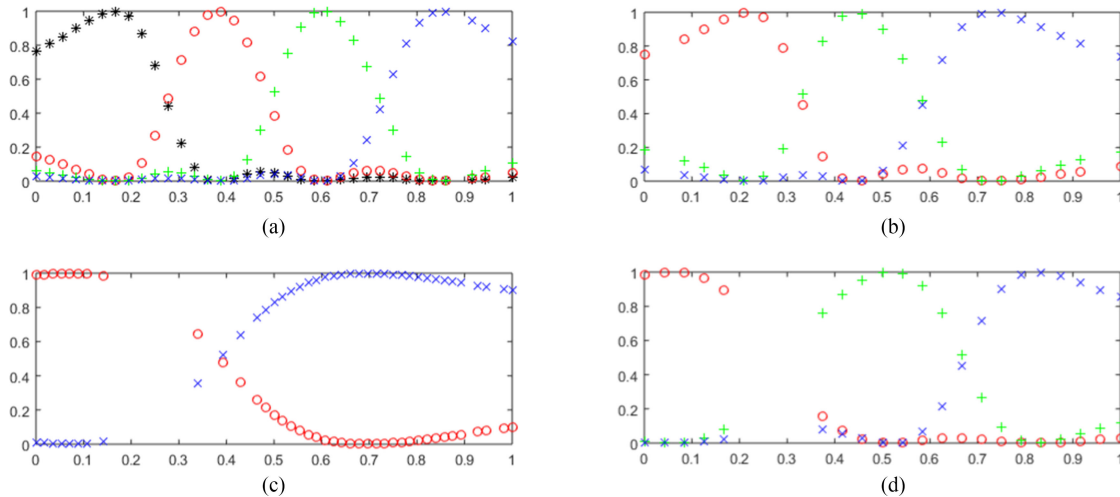
Fig. 6. Membership functions learned with FCM for Iris dataset, respectively, plotted in sub-figures (a)– (d) for attributes 1–4.

TABLE XV
ACCURACIES (%) VS. SPECIFICATION OF MEMBERSHIP FUNCTIONS FOR IRIS DATASET

Fuzzification Method	CRI	Ori	IG	Relief-F	LLCFS	LS	RSFS	AVG_Guided	Improvement over CRI	Improvement over Ori
Evenly Distributed	42.66	79.33	90.66	92.66	91.33	90.00	91.33	91.20	48.54 (114%)	11.87 (15%)
FCM-learned	68.00	88.00	95.99	93.33	92.00	90.67	93.33	93.07	25.07 (37%)	5.07 (6%)

conventional T-FRI and CRI-based classification methods, as also highlighted in this table.

It may be recognized that the improved classification rate is still not so high as the highest possible as reported in the literature regarding this simple dataset [51], where a fully trained learning classifier is adopted with the fuzzy sets involved having been comprehensively optimized. However, it must be noticed that the present high accuracy is attained with an average of 20% rules having been randomly taken out of the learned rule base. This demonstrates the great potential of the proposed FRI approach in dealing with real-world problems, where typically only partial and imprecise knowledge is available.

V. CONCLUSION

This paper has presented a novel FRI approach that significantly reinforces the power of fuzzy interpolative reasoning, by exploiting attribute ranking techniques to help determine the relative importance of rule antecedent attributes involved in a sparse rule base. The approach is general since it allows for any established ranking method to be utilized to score the attributes, leading to a flexible weighting scheme for FRI. This paper has provided five different attribute ranking methods for attribute weighting, based on popular FS techniques in the relevant literature. This paper has also proposed an innovative reverse engineering procedure, through which the ranking scores can be calculated from an artificial decision table derived from the original rules, without requiring additional observations to be made. The proposed work has been systematically evaluated on ten benchmark classification tasks.

Collectively, the experimental results presented have clearly demonstrated the efficacy and robustness of the attribute ranking-supported approach to fuzzy rule interpolative reasoning. In particular, the weighted interpolative methods have been shown to entail remarkably improved classification accuracy over both conventional T-FRI and CRI-based fuzzy reasoning techniques. This has been achieved using a very simple fuzzification mechanism. The experimental investigations have also confirmed that any of the existing popular FS techniques may be employed to evaluate and score the antecedent attributes, without adversely affecting the classification outcome nor considerably increasing the computational time complexity. The results have further illustrated that better performance can be obtained by fine tuning the membership functions that define the antecedent attributes given a particular practical problem.

In addition to the aforementioned advantages over conventional T-FRI techniques, the attribute ranking-supported T-FRI methods have systematically proven to only require the least number of the closest rules to carry out interpolation (with respect to a given observation that does not match any existing rule in the sparse rule base). Overall, as the most appropriate closest rules are selected in terms of the relative significance of domain attributes, better results are obtained using fewest rules possible, thereby, minimizing the complexity in both rule searching and rule firing.

The work in this paper is developed on the basis of the popular T-FRI algorithm, the proposed approach appears to permit other FRI techniques to be integrated with the attribute ranking and reverse engineering methods in a similar manner. For instance, popular FRI algorithms such as those reported in [10],

[39], [52] also involve multidimensional input spaces, which may be combined with the proposed attribute ranking scheme, thereby, creating potentially more effective multidimensional FRI methods. This hypothesis remains to be tested, with an expectation that a generalized weighting scheme for fuzzy interpolative reasoning can be developed, making fuzzy rule-based inference more flexible. Also, this paper assumes that a sparse rule base is given. It would be interesting to investigate which of the many data-driven rule induction techniques available may be employed to generate an improved rule base, and to study how such a learning mechanism may be blended with the reverse engineering procedure to provide a stronger attribute ranking scheme. Additionally, how the reverse engineering procedure may be efficiently implemented to minimize the adverse impact of the curse of dimensionality forms another piece of further research. Finally, the current approach presumes the use of a fixed (sparse) rule base. The most recently proposed mechanism for dynamic FRI [53] should be integrated with it to allow the collection and refinement of any intermediate fuzzy rules and interpolated results, in order to enrich the rule base and avoid unnecessary interpolation on the fly.

REFERENCES

[1] L. T. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *Int. J. Approx. Reason.*, vol. 9, no. 3, pp. 197–225, 1993.

[2] L. T. Kóczy and K. Hirota, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Inf. Sci.*, vol. 71, no. 1, pp. 169–201, 1993.

[3] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 1, pp. 28–44, Jan. 1973.

[4] P. Baranyi, L. T. Kóczy, and T. D. Gedeon, "A generalized concept for fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 6, pp. 820–837, Dec. 2004.

[5] Z. Huang and Q. Shen, "Fuzzy interpolative reasoning via scale and move transformations," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 40–359, Apr. 2006.

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

[7] S. Jin, R. Diao, C. Quek, and Q. Shen, "Backward fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1682–1698, Dec. 2014.

[8] L. Yang and Q. Shen, "Adaptive fuzzy interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1107–1126, Dec. 2011.

[9] Y. Yam, M. L. Wong, and P. Baranyi, "Interpolation with function space representation of membership functions," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 3, pp. 398–411, Jun. 2006.

[10] D. Tikk and P. Baranyi, "Comprehensive analysis of a new fuzzy rule interpolation method," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 3, pp. 281–296, Jun. 2000.

[11] P. Baranyi, D. Tikk, Y. Yam, L. T. Kóczy, and L. Nadai, "A new method for avoiding abnormal conclusion for α-cut based rule interpolation," in *Proc. 1999 IEEE Int. Fuzzy Syst. Conf. Proc.*, 1999, vol. 1, pp. 383–388.

[12] L. Yang, F. Chao, and Q. Shen, "Generalised adaptive fuzzy rule interpolation," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 839–853, Aug. 2017.

[13] L. Yang and Q. Shen, "Closed form fuzzy interpolation," *Fuzzy Sets Syst.*, vol. 225, pp. 1–22, 2013.

[14] C. Chen, N. M. Parthaláin, Y. Li, C. Price, C. Quek, and Q. Shen, "Rough-fuzzy rule interpolation," *Inf. Sci.*, vol. 351, pp. 1–17, 2016.

[15] Y. C. Chang, S. M. Chen, and C. J. Liau, "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1285–1301, Oct. 2008.

[16] S. M. Chen, Y. C. Chang, and J. S. Pan, "Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 Gaussian fuzzy sets and genetic algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 3, pp. 412–425, Jun. 2013.

[17] S. M. Chen, S. H. Cheng, and Z. J. Chen, "Fuzzy interpolative reasoning based on the ratio of fuzziness of rough-fuzzy sets," *Inf. Sci.*, vol. 299, pp. 394–411, 2015.

[18] W. H. Hsiao, S. M. Chen, and C. H. Lee, "A new interpolative reasoning method in sparse rule-based systems," *Fuzzy Sets Syst.*, vol. 93, no. 1, pp. 17–22, 1998.

[19] Z. Wu, M. Masaharu, and Y. Shi, "An improvement to Kóczy and Hirota's interpolative reasoning in sparse fuzzy rule bases," *Int. J. Approx. Reason.*, vol. 15, no. 3, pp. 185–201, 1996.

[20] R. Diao, S. Jin, and Q. Shen, "Antecedent selection in fuzzy rule interpolation using feature selection techniques," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2014, pp. 2206–2213.

[21] Y. C. Chang and S. M. Chen, "A new method for multiple fuzzy rules interpolation with weighted antecedent variables," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2008, pp. 76–81.

[22] S. M. Chen, L. W. Lee, and V. R. Shen, "Weighted fuzzy interpolative reasoning systems based on interval type-2 fuzzy sets," *Inf. Sci.*, vol. 248, pp. 15–30, 2013.

[23] S. M. Chen and S. I. Adam, "Weighted fuzzy interpolated reasoning based on ranking values of polygonal fuzzy sets and new scale and move transformation techniques," *Inf. Sci.*, vol. 435, pp. 184–202, 2018.

[24] 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, Aug. 2011.

[25] 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," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 6, pp. 1412–1427, Dec. 2009.

[26] S. M. Chen and Y. C. Chang, "Weighted fuzzy interpolative reasoning for sparse fuzzy rule-based systems," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9564–9572, 2011.

[27] S. M. Chen and Z. J. Chen, "Weighted fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on piecewise fuzzy entropies of fuzzy sets," *Inf. Sci.*, vol. 329, pp. 503–523, 2016.

[28] Y. Li, D. Huang, and L. Zhang, "Weighted fuzzy interpolative reasoning method," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2005, vol. 5, pp. 3104–3108.

[29] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 3, pp. 131–156, 1997.

[30] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454. New York, NY, USA: Springer, 2012.

[31] 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.

[32] Y. Cui, J. S. Jin, S. Zhang, S. Luo, and Q. Tian, "Correlation-based feature selection and regression," in *Proc. Pacific-Rim Conf. Multimedia*, 2010, pp. 25–35.

[33] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, no. 1, pp. 155–176, 2003.

[34] H. Zeng and Y. M. Cheung, "Feature selection and kernel learning for local learning-based clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1532–1547, Aug. 2011.

[35] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowl.-Based Syst.*, vol. 24, no. 7, pp. 1024–1032, 2011.

[36] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.*, 1994, pp. 171–182.

[37] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. 18th Int. Conf. Neural Inf. Process. Syst.*, 2005, pp. 507–514.

[38] F. Li, C. Shang, Y. Li, and Q. Shen, "Guiding fuzzy rule interpolation with information gains," in *Advances in Computational Intelligence Systems*. New York, NY, USA: Springer, 2017, pp. 165–183.

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

[40] S. Kovács and L. T. Kóczy, "The use of the concept of vague environment in approximate fuzzy reasoning," *Fuzzy Set Theory Appl.*, vol. 12, pp. 169–181 1997.

[41] Y. Yam and L. T. Kóczy, "Representing membership functions as points in high-dimensional spaces for fuzzy interpolation and extrapolation," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 6, pp. 761–772, Dec. 2000.

[42] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997, pp. 55–60.

[43] M. Wu and B. Schölkopf, "A local learning approach for clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1529–1536.

[44] R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 73–89, Feb. 2007.

[45] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases," University of California-Irvine, Irvine, CA, USA, 1998. [Online]. Available: https://archive.ics.uci.edu/ml/datasets.html

[46] J. AlcaláFdez *et al.*, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Multiple-Valued Logic Soft Comput.*, vol. 17, no. 2/3, pp. 255–287, 2011. [Online]. Available: http://sci2s.ugr.es/keel/datasets.php

[47] L. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414–1427, Nov./Dec. 1992.

[48] K. Balázs, J. Botzheim, and L. T. Kóczy, "Comparison of various evolutionary and memetic algorithms," in *Integrated Uncertainty Management and Applications.* New York, NY, USA: Springer, 2010, pp. 31–442,.

[49] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2/3, pp. 191–203, 1984.

[50] M. S. Chen and S. W. Wang, "Fuzzy clustering analysis for optimizing fuzzy membership functions," *Fuzzy Sets Syst.*, vol. 103, no. 2, pp. 239–254, 1999.

[51] L. S. Riza, C. N. Bergmeir, F. Herrera, and J. M. Ben ítez Sánchez, "frbs: Fuzzy rule-based systems for classification and regression in R," *J. Statist. Softw.*, vol. 65, pp. 1–30, 2015.

[52] K. W. Wong, D. Tikk, T. D. Gedeon, and L. T. Kóczy, "Fuzzy rule interpolation for multidimensional input spaces with applications: A case study," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 809–819, Dec. 2005.

[53] N. Naik, R. Diao, and Q. Shen, "Dynamic fuzzy rule interpolation and its application to intrusion detection," *IEEE Trans. Fuzzy Syst.*, doi: 10.1109/TFUZZ.2017.2755000.

Fangyi Li received the B.Sc. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2014. She is currently working toward the Ph.D. degree in computational intelligence jointly with the School of Computer Science and Engineering, Northwestern Polytechnical University, and also with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Aberystwyth, U.K.

Her research interests include fuzzy interpolative reasoning, machine learning, and image classification.

Changjing Shang received the Ph.D. degree in computing and electrical engineering from Heriot-Watt University, Edinburgh, U.K., in 1995.
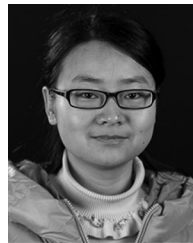
She is currently a University Research Fellow with the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Aberystwyth, U.K. Prior to joining Aberystwyth, she was with Heriot-Watt, Loughborough and Glasgow Universities. She has published extensively, and supervised more than ten Ph.Ds/PDRAs in the areas of pattern recognition, data mining and analysis, space robotics, and image modeling, and classification.

Ying Li received the Ph. D. degree in circuits and systems from the Xidian University, Xian, China, in 2002.

Since 2003, she has been with the School of Computer Science, Northwestern Polytechnical University, Xian. She is currently a Full Professor with Northwestern Polytechnical University. Her research interests include computation intelligence, image processing, and signal processing. She has published extensively in these areas.

Jing Yang received the B.Sc. degree in information and computing science from Taiyuan Normal University, Taiyuan, China, in 2012, and the M.Sc. degree in computational mathematics from Northwestern Polytechnical University, Xian, China, in 2015. She is currently working toward the Ph.D. degree jointly with the School of Computer Science and Engineering, Northwestern Polytechnical University and the Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Aberystwyth, U.K.

Her research interests include fuzzy systems, deep learning, and image super-resolution.

Qiang Shen received the Ph.D. degree in computing and electrical engineering from Heriot-Watt University, Edinburgh, U.K., in 1990, and the D.Sc. degree in computational intelligence from Aberystwyth University, Aberystwyth, U.K., in 2013.

He is currently the Chair of Computer Science and the Director of the Institute of Mathematics, Physics and Computer Science, Aberystwyth University. He has authored two research monographs and more than 370 peer-reviewed papers.

Dr. Shen was the recipient of an Outstanding Transactions Paper Award from the IEEE.