Aberystwyth University

Fuzzy knowledge-based prediction through weighted rule interpolation

Li, Fangyi; Li, Ying; Shang, Changjing; Shen, Qiang

Fuzzy Knowledge-Based Prediction Through Weighted Rule Interpolation

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

Abstract—Fuzzy rule interpolation (FRI) facilitates approximate reasoning in fuzzy rule-based systems only with sparse knowledge available, remedying the limitation of conventional compositional rule of inference working with a dense rule base. Most of the existing FRI work assumes equal significance of the conditional attributes in the rules while performing interpolation. Recently, interesting techniques have been reported for achieving weighted interpolative reasoning. However, they are either particularly tailored to perform classification problems only or employ attribute weights that are obtained using additional information (rather than just the given rules), without integrating them with the associated FRI procedure. This paper presents a weighted rule interpolation scheme for performing prediction tasks by the use of fuzzy sparse knowledge only. The weights of rule conditional attributes are learned from a given rule base to discriminate the relative significance of each individual attribute and are integrated throughout the internal mechanism of the FRI process. This scheme is demonstrated using the popular scale and move transformation-based FRI for resolving prediction problems, systematically evaluated on 12 benchmark prediction tasks. The performance is compared with the relevant state-of-the-art FRI techniques, showing the efficacy of the proposed approach.

Index Terms—Attribute weighting, intelligent prediction, knowledge interpolation, sparse knowledge.

I. INTRODUCTION

UNLIKE knowledge-based classification systems using inference rules that are used to determine categorical class labels for unknown data, prediction systems facilitate the forecasting of continuous-valued outcomes of a certain problem being modeled by the system. Such systems enjoy a wide range of successful real-world applications, including medical case assessment [1]; object tracking and video surveillance [2], [3]; financial trend forecasting [4]; civil industry simulation [5]; and generic problem of time series

analysis [6], [7]. A variety of techniques have been developed for knowledge-based prediction, entailing the transparency of both the system model and the inference process of running the model, amongst which are fuzzy rule-based prediction systems (e.g., [8]–[10]).

Fuzzy rule-based systems provide a powerful tool to perform modeling and reasoning tasks involving imprecision and vagueness. The use of semantics-rich rules also strengthens the inference interpretability for human users. This is often supported by employing the compositional rule of inference (CRI) [11], involving a dense fuzzy rule base which fully covers the entire problem domain. If, however, the problem domain is not completely covered by the given rules, there may exist observations that fail to activate any existing rules to compute a required prediction. Fuzzy rule interpolation (FRI) plays an approximate and useful role in such situations where only an incomplete rule base is available [12], [13]. Such a rule base is termed as a sparse rule base hereafter. FRI works with this form of sparse knowledge, attempting to reduce, if not to completely remove, the restriction of CRI for cases where no conclusion may be derived due to no rules matching a new observation. This offers an alternative way to predict an approximately inferred prediction outcome.

Many FRI methods exist (e.g., [14]–[21]) including advanced techniques that are capable of dealing with interpolation by the use of fuzzy rules that involve multiple conditional variables. There is a common problem existing in these FRI approaches, where the conditional attributes within the rules are presumed to be of equal significance for interpolation. As such, incorrect interpolated outcomes may result since different domain attributes may generally make different contributions to the prediction process. Recently, however, a number of methods have been presented for FRI working with multiple conditionals associated with different weights [22]–[26]. Nevertheless, two key questions remain to be further investigated in developing such weighted FRI: 1) how the weights are generated and 2) whether these weights are integrated within the underlying, nonweighted FRI.

Regarding the first question, one of the possible answers is to assign the predefined weights by domain experts [27]. However, this will require human intervention and, hence, adversely reduce the flexibility of the resulting fuzzy systems. Automatic weight learning schemes are obviously preferred. In particular, there has been work developed that utilizes genetic algorithms (GAs) to induce from data the weights of rule conditionals, thereby strengthening the effectiveness of FRI [22]. Yet, such techniques introduce much more additional

computation and also the required specification of many GA parameters. Alternatively, the weights may be determined by a distance measure between the information content of an observation and that of an conditional attribute within a given rule. The information concerned is related to the characteristic points of the fuzzy sets that specify the corresponding attributes, including the central point [23] or the ranking value [24] of a fuzzy set. The weights are assigned differently to each conditional fuzzy set that appears in each and every different rule, incurring significant extra computation, and reducing the interpretability of the weighted rules. Different from these, the weight learning schemes as reported in [26], form an implementation of the "wrapper" approach, mixing up FRI-based inference and learning from data.

For the second question, the existing techniques generally work by artificially creating an overall weight to each of the rules before running the weighted rules in FRI. Such weights are normally computed through aggregating the weights calculated for individual conditionals, thereby involving additional aggregation procedures. Situations become even more complicated when dealing with different fuzzy interpolative reasoning systems that may be constructed from different perspectives, coping with different aspects of the fuzzy rule model, e.g., piecewise fuzzy entropies [23] and ranking scores [24] of the fuzzy values. The resulting weights are exploited rather differently, depending on what underlying FRI mechanism is employed. Most significantly, within these techniques, the computed weights are decoupled from the internal working procedures of the particular FRI method used. This makes the interpretation of the resulting FRI process more difficult than explicitly combining the weights with the procedural steps.

Most recently, there has been an attempt to address both the aforementioned issues with initial success [28]. However, that seminal work is tailored to handle knowledge-based classification problems only. Having taken notice of this, in order to perform fuzzy rule-based prediction with sparse rule bases, it is desirable to exploit a way to effectively evaluate the attribute variables and to organically integrate the resulting attribute significance measures into FRI algorithms. This should be done by the use of only the given knowledge, in terms of the rule base without resorting to any real observations or triggering any FRI or CRI inference process.

There have been many proposals for assessing the capabilities of domain attributes in their influence upon the potential prediction outcome, including both supervised attribute evaluation methods (e.g., [29] and [30]) and unsupervised methods (e.g., [31] and [32]). Through making the assessment in a certain form of relationship (say, similarity) amongst the attributes themselves, the unsupervised approach offers more flexibility, especially for prediction problems since the consequent attribute is not required during the attribute evaluation process. The adaptation of such an conditional attribute evaluation method is employed here to weigh the relative rule conditionals. The resulting weights, learned from the given sparse rule base only, are integrated within the FRI to remedy the adverse and restrictive assumption of weights having equal significance.

The work is, herein, implemented by adapting the popular transformation-based FRI (T-FRI) [17], [33] (that only deals with rules whose conditional attributes are of equal significance) with the weights being used to modify all components of the T-FRI computation process. Comparative experimental investigations are carried out on prediction tasks of both multivariate regression and time series forecasting. The results show that the proposed approach is able to considerably reduce the adverse impact of assuming equal significance of different conditionals that has been commonly made in typical FRI techniques. This helps to significantly improve the prediction accuracy of systems using FRI.

The rest of this paper is structured as follows. Section II presents the details of the proposed fuzzy sparse rule-based interpolative scheme for prediction. Section III analyzes the computational complexity of the proposed approach. Section IV reports on a systematic, comparative evaluation over different prediction problems and discusses the experimental results. Finally, Section V summarizes the contributions of this paper and outlines interesting further research.

II. WEIGHTED FUZZY RULE INTERPOLATION FOR PREDICTION

A fuzzy sparse knowledge-based inference mechanism for prediction is presented in this section. A generic framework is first outlined, followed by a description of the algorithm for learning the weights of conditional attributes from the given sparse rule base, and an implementation of integrating such learned weights within T-FRI.

A. Framework for Sparse Rule-Based Prediction

To facilitate the illustration of the proposed work, without losing generality, an original (sparse) fuzzy rule base $R = \{r^1, r^2, \ldots, r^N\}$, where conditionals are not individually weighted, and an observation o^* are represented in the following format, respectively:

r^i : if a_1 is A_1^i and a_2 is A_2^i and $\cdots$ and a_m is A_m^i, then z is Z^i

o^* : a_1 is A_1^* and a_2 is A_2^* and $\cdots$ and a_m is A_m^*

where $a_j, j = 1, 2, \ldots, m$, are the conditional attributes; z is the consequent attribute; A_j^i denotes the fuzzy set value taken by a_j in the rule r^i; and Z^i stands for the fuzzy set value of the consequent attribute z in r^i. For simplicity, triangular membership functions are employed for explanation and evaluation in this paper.

Given R and o^*, most of the conventional fuzzy rule-based systems may be able to generate a required consequent by the use of CRI firing the matched rule(s). If, however, the rule base is sparse, where no rule matches the observation, fuzzy interpolative inference can be utilized as an alternative reasoning mechanism for deriving an estimated consequent. The proposed method integrates both the conventional CRI and a novel T-FRI technique (referred to as weighted FRI hereafter) that is guided with the weights learned and assigned to the conditionals. Through this integration, it is expected to obtain more accurate prediction results by exploiting the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: FUZZY KNOWLEDGE-BASED PREDICTION THROUGH WEIGHTED RULE INTERPOLATION 3

Algorithm 1 Sparse Rule-Based Prediction

Input: • Rule base $R = \{r^1, \ldots, r^N\}$, of N rules
- Observation $o^* = \{A_1^*, \ldots, A_m^*\}$, over m conditional attributes
- Cardinalities of fuzzy partitions $C = [c_1, \ldots, c_m]$, over attribute domains
- Lists of fuzzy values $F = [g_1, \ldots, g_m]$, where $g_i = \{f_1, \ldots, f_{c_i}\}$ per attribute
- Number of closest rules n

Output: • Prediction in crisp value
1: **for** $i = 1$ to $i = N$ **do**
2:     Matching o^* against rule r^i;
3: **end for**
4: **if** *matched with at least one rule* **then**
5:     Fire matched rule(s) using CRI to obtain required prediction Z^* for o^*;
6: **else**
7:     **if** o^* is first unmatched ever **then**
8:         Learn weights from sparse rule base R: $W = LWFR(R, C, F)$;
9:         Save W;
10:     **else**
11:         Recall attribute weights W;
12:     **end if**
13:     Execute weighted FRI to compute $Z^* = WeightedTFRI(R, o^*, n, W)$;
14: **end if**
15: Defuzzify Z^* as crisp real number;
16: **Return** Crisp-valued prediction

---

advantage of CRI for matched observations and that of the weighted FRI for those unmatched ones.

The integrated framework is shown in Algorithm 1. First of all, a check is made to determine whether the observation is matched with any rule in the given rule base. If there is at least one rule being found to match the observation, the result will be obtained by firing the matched rule(s). Otherwise, the weighted FRI as proposed below is used to make inference to estimate the consequent.

In reflection of the two research issues as raised in Section I, the weights are first learned by the use of attribute evaluation from the sparse rule base only, without requiring any observations. Then, given the rule base and the weights derived from it, a weighted FRI algorithm performs the required prediction, through weighted search of the closest neighboring rules of the observation and weighted interpolation with the selected closest rules. These are reflected in lines 8 and 13, respectively, in Algorithm 1, with the details explained in the following sections.

B. Learning Attribute Weights From Sparse Rule Base

The biggest issue of learning weights in an effort to distinguish the relative significance degrees associated with the conditionals is where the data comes from. In this paper, no additional information is assumed, other than the sparse rule base provided. In general, FRI works with a sparse rule base, and so it may be difficult to acquire sufficient example data for use in support of computing the required weights. If there were sufficient training data in the problem domain, the situation of having a sparse rule base might not exist in the first place, as such data could have been utilized to generate a dense rule

base. Thus, only the originally provided sparse rule base is used as the information source for assessing attribute weights. This requires the introduction of a method to preprocess the sparse rule base for the generation of a set of valid training instances.

The basic idea of the preprocessing is to reformulate automatically the rules in the given sparse rule base into data representations of a common structure. This is necessary because for a sparse rule-based system, different conditional attributes may appear in different rules and different rules may have different number of conditionals. Reflecting this view, the training instances are artificially generated through the following three-step procedure.

1) Identification of all conditional attributes appearing in all the rules and all (finite and fuzzy) values used to define these attributes.
2) Expansion of each rule in the sparse rule base into one that involves all conditional attributes such that if a certain conditional is not originally involved in the rule, then it is inserted into the rule with its value being set to a qualitative term, "do not care."
3) Replacement of each "do not care" with every possible fuzzy value for the corresponding attribute in each rule that contains this qualitative value, such that one rule involving L attributes of the "do not care" value ($L \geq 1$) is replaced by $\prod_{i=1}^{L} c_i$ rules, with c_i being the cardinality of the value domain of a certain conditional that does not appear in the original rule.

In so doing, within each of the expanded rules a conditional attribute that does not appear in a given rule now takes one and only one possible fuzzy value from its underlying domain. For example, if a given original rule contains just one "missing" conditional attribute, then this rule is expanded to k rules, where k is the number of the fuzzy sets that this attribute may take as its value.

The computation involved in implementing the above procedure may appear to be expensive at the first sight. However, typically there are only a relatively much smaller number of rules contained within a sparse rule base than those within a dense fuzzy rule base. Also, the replacement only takes place for those missing conditionals by filling in values taken from their value domains that are of usually a small cardinality (psychologically speaking, to have a human comprehensible prediction system, the cardinality is at most 9). Thus, the computation overheads are practically very manageable. Note that this procedure for generating artificial training instances makes logical sense. Indeed, as an attribute does not appear in the original rule conditional part, an observation will lead to the same consequent independent of what fuzzy set value may be taken by that attribute, so long as all the other conditional attributes originally appearing in the instance are matched with the observation.

This procedure of learning weights from a sparse rule base is based on the idea that is originally introduced in [28] and is formally formulated for the first time here, as presented in Algorithm 2. In particular, the above preprocessing of the rule base is summarized in the first 13 lines. As may be expected, the pool of generated training instances remains the

---

Algorithm 2 Weight Learning From Sparse Rule Base: $W = LWFR(R, C, F)$

Input:
- Rule base $R = \{r^1, \ldots, r^N\}$, of N rules
- Cardinalities of fuzzy partitions $C = [c_1, \ldots, c_m]$, over attribute domains
- Lists of fuzzy values $F = [g_1, \ldots, g_m]$, where $g_i = \{f_1, \ldots, f_{c_i}\}$ per attribute

Output:
- Normalised attribute weights W

1: Initialise training instance pool $TIP = R$;
2: **for** $i = 1$ to $i = N$ **do**
3:     Check if there are any missing conditionals in rule r^i;
4:     **if** *no missing* **then**
5:         Continue;
6:     **else**
7:         **for** each missing conditional a_k in r^i **do**
8:             Replacing r^i in *TIP* with c_k copies of r^i;
9:             Assigning a_k in each copy with one of different fuzzy values in g_k;
10:         **end for**
11:     **end if**
12: **end for**
13: Remove identical instances in *TIP*;
14: Calculate weights: weight$_{LLC}$=LLC(*TIP*) or weight$_{LS}$=LS(*TIP*);
15: Calculate normalised attribute weights:

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

16: **Return** Normalised attribute weights W

---

same as the original rule base if there is no missing antecedent attribute within any rule condition across the entire rule base. This reveals the essence of such a training instance production process, which helps to reduce the sparsity of the original rule base, by complementing the potential information without requiring any further knowledge.

Now, having created a training instance pool from a given sparse rule base, attribute significance can be estimated by the use of an unsupervised evaluation method. In this paper, the following two methods are considered as potential alternatives due to their popularity and availability: 1) local learning-based clustering (LLC) [32] and 2) Laplacian score (LS) [31]. A direct application of either will lead to a vector of weights for the attributes, denoted by weight$_{LLC}$ and weight$_{LS}$, respectively. Note that LS assigns a smaller LS for an attribute of a higher importance degree. Hence, the weight derived by the use of LS for a certain attribute a_i, $i = 1, \ldots, m$, is modified from the raw score LS_i, such that

$$\text{weight}_{LS}(a_i) = \frac{1}{1 + LS_i}. \tag{1}$$

The above is just one of the possible alternatives, other transformation, for example, e^{-LS_i}, may be employed to define the LS-based weights.

Once the weights of individual conditionals are estimated, they are normalized and ranked to better reflect their relative significance. In particular, each conditional attribute a_i can be assigned with a normalized weight as shown in line 15 of Algorithm 2, where weight$_*$ denotes either of the two types of weight (namely weight$_{LLC}$ and weight$_{LS}$). Such normalized weights signify how significant the individual conditional attributes may be in the determination of the rule consequent.

This weighting scheme is independent of any acquisition of real observations, it is purely data-driven by the use of the training instances artificially derived from the original rule base. Whilst two alternatives are provided here to offer flexibility for choice, only one approach is required to implement the following weighted FRI (i.e., just using either weight$_{LLC}$ or weight$_{LS}$ in any implementation in line 14).

C. Weighted FRI

The primary motivation for introducing weights for FRI is to significantly reduce the adverse impact caused by the assumption that all conditional attributes are of equal importance. As an implementation of the proposed framework, as indicated previously, T-FRI [17], [33] is utilized to act as the foundation for integration with attribute weights. Algorithm 3 formulates the procedure of such an implementation. The development of such a weighted T-FRI algorithm is based on the examination of how the conventional T-FRI performs nonweighted interpolative prediction. The following describes the key ideas involved in more detail.

1) Weighted Distance for Closest Rules Selection: Any FRI process starts as an observation o^* being newly presented to the fuzzy system does not activate any rule in the sparse rule base, due to no matching (or in certain FRI-based systems, due to too low-level partial matching). Then, for interpolation, n ($n \geq 2$) rules closest to the observation are sought in order to implement the interpolation. In conventional T-FRI, the selection of these closest neighboring rules is normally done via measuring and aggregating the Euclidean distances between individual conditional attributes of a given rule and their corresponding values in the observation. Now that individual attributes have been evaluated and assigned a weight, each signifying their relative significance, the distance between any rule r^i and o^* needs to be adapted accordingly. This is realized in line 2 of Algorithm 3 by defining

$$\tilde{d}(o^*, r^i, W) = \frac{1}{\sqrt{\sum_{t=1}^{m}(1 - W_t)^2}} \sqrt{\sum_{j=1}^{m} \left((1 - W_j) d(A_j^*, A_j^i) \right)^2} \tag{2}$$

with $d(A_j^*, A_j^i)$ being computed via the representative value [17] such that

$$d(A_j^*, A_j^i) = \frac{\left| \text{Rep}(A_j^*) - \text{Rep}(A_j^i) \right|}{\max_{A_j} - \min_{A_j}} \tag{3}$$

where $d(A_j^*, A_j^i)$ represents the normalized result of the otherwise absolute distance; $\max_{A_j}$ and $\min_{A_j}$ denote the maximal and minimal value of the attribute a_j, respectively; and m is the number of all conditional attributes involved in all the given rules. If triangular membership functions are used throughout the reasoning process (say, for simplicity), then the representative value of a fuzzy set may be simply calculated by averaging the vertices of the triangular membership function, such that

$$\text{Rep}(A) = \frac{v_1 + v_2 + v_3}{3} \tag{4}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: FUZZY KNOWLEDGE-BASED PREDICTION THROUGH WEIGHTED RULE INTERPOLATION

5

Algorithm 3 Weighted T-FRI $Z^* = WeightedTFRI(R, o^*, n, W)$

Input:
- Sparse rule base $R = \{r^1, \ldots, r^N\}$, of N rules
- Observation $o^* = \{A_1^*, \ldots, A_m^*\}$, over m conditionals
- Number of closest rules n
- Conditional weights $W = (W_1, \ldots, W_m)$

Output:
- Interpolated consequent Z^*
  – **Closest Rules Selection:**
1: **for** $i = 1$ to $i = N$ **do**
2:   Calculating weighted distance $\tilde{d}(o^*, r^i, W)$ between o^* and r^i;
3: **end for**
4: Select n rules of shortest distance(s);
  – **Intermediate Rule (r') Construction:**
5: Obtain weights $w_j^i, i = 1, \ldots, n, j = 1, \ldots, m$, as computed by original T-FRI to jth conditional attribute of ith selected rule, such that
$$w_j^i = \frac{1}{1 + d(A_j^*, A_j^i)}$$
6: Compute conditional attribute values of intermediate rule $A_j', j = 1, 2, \ldots, m$, by linearly aggregating corresponding weighted conditional values over selected n rules using normalized weights $\hat{w}_j^i = \frac{w_j^i}{\sum_{t=1}^n w_j^i}$;
7: Calculate weight $\tilde{w}_z^i$ for each consequent per selected closest rule, by accumulating normalized weights contributed by $\hat{w}_j^i$, such that
$$\tilde{w}_z^i = \sum_{j=1}^m W_j \hat{w}_j^i$$
8: Construct fuzzy term Z' for consequent attribute of intermediate rule, by aggregating consequent values of n closest rules $z^i, i = 1, \ldots, n$, which are, respectively, weighted by $\tilde{w}_z^i$;
  – **Scale and Move Factor Calculation:**
9: **for** each conditional attribute **do**
10:   Obtaining scale rate s_{A_j} that modifies A_j' into $\hat{A}_j'$ such that it maintains same scale as corresponding component in o^*;
11:   Obtaining move ratio m_{A_j} that modifies $\hat{A}_j'$ for it to maintain same position as corresponding component in o^*;
12: **end for**
  – **Scale and Move Transformation:**
13: Calculate overall transformation factors for Z' to ensure analogy, by aggregating corresponding weighted scale and move factors, such that
$$\tilde{s}_Z = \sum_{j=1}^m W_j s_{A_j} \qquad \tilde{m}_Z = \sum_{j=1}^m W_j m_{A_j}$$
14: Compute final interpolated outcome Z^* by applying scale and move factors to Z', such that $Z^* = T(Z', \tilde{s}_Z, \tilde{m}_Z)$;
15: **Return** Z^*

---

where v_1 and v_3 represent the two extreme points of the support of the fuzzy set and v_2 denotes the normal point where the member value reaches 1.

Those conditional attributes associated with a relatively larger weight make less contribution to the overall, aggregated distance $\tilde{d}(o^*, r^i, W)$, because the value of $(1 - W_j)d(A_j^*, A_j^i)$ is smaller for $j = 1, \ldots, m$. Thus, selecting the n closest rules using this distance measure allows the rules that involve conditionals that are more important to be selected with a higher priority.

2) Weighted Interpolation: Recall the central idea of the weighted T-FRI approach: in sharp contrast with conventional FRI techniques, the significance degrees of individual conditional attributes are captured by artificially calculated attribute weights and used to compute the (interpolated) consequent given an unmatched observation. Thus, it is naturally desirable for the resulting weights to be integrated throughout the entire interpolation process. That is, further to the procedure for the closest rules selection as discussed above, procedures for intermediate rule construction, transformation factors calculation and eventual interpolative transformation are all expected to take the weights into consideration. Details for implementing such weighted procedures are also shown in Algorithm 3. For instance, the weighting on the consequent ($\tilde{w}_z^i$) and the required scale and move factors in the weighted transformation process ($\tilde{s}_Z$ and $\tilde{m}_Z$) are now computed as described in lines 7 and 13 of the algorithm, respectively.

As a weighted extension to the conventional T-FRI that is described in [17] and [33], the general rule interpolation process of this algorithm remains the same as its original. Note that the term of *weights* is a little over-worked herein, since it has already appeared in the conventional T-FRI, namely $w_j^i, i = 1, \ldots, n, j = 1, \ldots, m$, as given in line 5 of Algorithm 3. However, these weights are assigned for the sake of the construction of the intermediate rule, through direct comparison between the conditional attributes of a rule and the observation. This is completely different from the term of attribute weight $W_j, j \in \{1, \ldots, m\}$, that is focused on, in this paper, which reveals the relative importance of each conditional attribute underpinning the original data. In particular, the weight W_j associated with a certain conditional attribute a_j is computed independent of, and fixed throughout, the interpolative process, no matter which original rule is under consideration. They are artificially calculated without acquisition of any real observations nor comparison between a given observation and any rules. Yet, in the original T-FRI, the weight w_j^i computed with respect to a certain conditional attribute is generally of a different value when a different fuzzy rule r^i is addressed.

III. ANALYSIS OF COMPUTATIONAL COMPLEXITY

This section analyzes the computational complexity of the proposed weighted T-FRI approach for prediction. Recall Algorithms 1–3, the time complexity of the overall approach can be estimated in the following. In particular, the two key sub-procedures, namely weight learning from sparse rule base and weighted T-FRI, are analyzed first, which are then collected together to present the overall computational complexity. The notations for describing the algorithmic variables involved are the same as those specified in the **Input** statements of each algorithm.

A. Time Complexity of Learning Weights From Sparse Rule Base

As shown in Algorithm 2, the initialization and result return in lines 1 and 16 cost $O(1)$ of computation time. The **for** loop in lines 2–12 repeats N times. In particular,

line 3 takes $O(m)$. Without losing generality, suppose that there are conditional attributes missing in a certain rule. In order to estimate the time complexity, in the worst case (where only one conditional is not missing from the rule), the **for** loop in lines 7–10 repeats $m - 1$ times, costing $O(c)$ for each, where $c = \max\{c_1, \ldots, c_m\}$. The computation time of line 13 involves the number of entries in the resultant training instance pool, which costs $O((\text{size(TIP)} - 1)!)$. Assume that the time complexity of the method for attribute evaluation is $T(\text{AttriEval})$, while the computation cost for normalized weights is $O(m)$. Therefore, in total, the time complexity of Algorithm 2 is $T(\text{LWFR}) = 2O(1) + N \times [O(m) + (m - 1)O(c)] + O((\text{size(TIP)} - 1)!) + T(\text{AttriEval}) + O(m) = O(Nmc) + O((\text{size(TIP)} - 1)!) + T(\text{AttriEval})$.

B. Time Complexity of Weighted T-FRI

In weighted T-FRI, lines 1–3 of Algorithm 3 cover a **for** loop which costs $N \times O(m)$ of computation time, and line 4 takes $O(N^2)$ for sorting. Lines 5–7 lead to a time cost of $O(mn)$ each, as they involve linear computation for every jth conditional attribute of the ith selected closest rule ($i = 1, \ldots, n, j = 1, \ldots, m$). Line 8 requires $O(n)$ time. Lines 9–12 form a **for** loop with each step in the loop (i.e., lines 10 and 11) taking a unit time of $O(1)$, and thus, the whole loop costs $O(m)$ of computation time. Line 13 takes $O(m)$ as the calculation of the transformation factors takes linear time with regard to the number of conditional attributes. Finally, the computation of the required interpolated result and returning it in the last two lines take $O(1)$ time each. Note that the number of the closest rules required to perform interpolation is commonly set to $n = 2$ in the existing literature. The total time complexity of weighted T-FRI is therefore, estimated to be $T(\text{WeightedTFRI}) = N \times O(m) + O(N^2) + 3O(mn) + O(n) + 2O(m) + 2O(1) = O(N(m + N))$.

C. Overall Computational Complexity

Algorithm 1 outlines the proposed fuzzy sparse rule-based prediction process, which invokes two subroutines: 1) weights learning scheme and 2) weighted T-FRI. Given the above analysis regarding the time complexity of these two subprocedures, it is ready to assess the overall computational complexity of a system implementing the entire framework. The starting **for** loop in lines 1–3 repeats N times, each of which costs $O(m)$ of computation time. The **if** statement in line 4 takes $O(m)$ as well. Firing matched rules in line 5 only requires a unit time of $O(1)$, otherwise, the worst case time complexity will reach the sum of $T(\text{LWFR}) + T(\text{WeightedTFRI})$. The close up step for defuzzification and return statement cost a unit time of $O(1)$ for each. This results in the total time complexity (in the worst case): $T_{\text{worst}} = N \times O(m) + O(m) + T(\text{LWFR}) + T(\text{WeightedTFRI}) + 2O(1) = O(Nmc) + O((\text{size(TIP)} - 1)!) + O(N(m + N)) + T(\text{AttriEval})$.

For comparison, the time complexity of the conventional T-FRI procedure [17], [33] is also checked here, which is $T(\text{TFRI}) = O(N(m+N))$. This is exactly the same as the complexity of the weighted T-FRI because the attribute weights in the weighted version are not computed within the interpolative

process itself. However, regarding the entire prediction system which employs just the original T-FRI for interpolative inference, without involving attribute weight learning, the worst total time complexity becomes: $T_{\text{worst-TFRI}} = N \times O(m) + O(m) + T(\text{TFRI}) + 2O(1) = O(mN + N^2)$, which is of course lower than that required by the weighted version and is expected.

Note that the above complexity analysis is the first systematically carried out regarding any of the T-FRI techniques, including the most recent development of weighted T-FRI for classification problems [28] (where only limited empirical case studies are performed). As such, this section is itself of significant novelty and theoretical value. Note also that the time complexity of attribute evaluation is not detailed here as the employment of such an algorithm is independent of the FRI inference process. Naturally, an evaluation method which has less time consumption is preferred for use. As can be seen in the experimental results to be shown next, the use of which evaluation method may not cause much difference upon the prediction accuracy. Hence, the choice of an attribute evaluation mechanism can be made with respect to their computational simplicity.

IV. Experimental Evaluation

The proposed fuzzy sparse knowledge-based interpolation approach for prediction is herein applied for dealing with 12 benchmark problems, including eight of which for multivariate regression and four for time series prediction. The prediction accuracies are assessed through comparison with those obtained by the conventional nonweighted T-FRI over all tasks. In addition, the performance is also compared against the weighted fuzzy interpolation method as reported in [23], which represents the state-of-the-art of FRI involving attribute weights, across the same seven problems used in that work. Note that there were eight datasets given in [23] but one of which is not available for the present investigation and, hence, only seven datasets are considered here.

A. Experimental Set-Up

The eight benchmark multivariate regression problems are taken from the popular UCI machine learning [34] and KEEL dataset repositories [35], while the four classic time series prediction problems are acquired from [6] and [36]. These 12 selected datasets involve different numbers of conditional attributes and cover various real-world problem domains, including civil engineering, energy consumption, weather forecasting, and time series prediction in industrial processes, amongst others. The properties of these datasets are summarized in Table I.

The rule bases used by each implemented fuzzy system, for both CRI and rule interpolation-based reasoning, are learned from the raw training data using the popular method of [37] (though alternative learning mechanism may be employed for this if preferred). For simplicity, the fuzzy values of all conditional attributes are represented by triangular membership functions in this experimental investigation. The partition of each conditional attribute domain into such fuzzy values is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: FUZZY KNOWLEDGE-BASED PREDICTION THROUGH WEIGHTED RULE INTERPOLATION 7

TABLE I
DATASETS USED FOR PREDICTION

Dataset	Attribute #	Instance #
Abalone [35]	7	4177
Concrete Compressive Strength [35]	8	1030
Concrete Slump Test [34]	9	103
Laser [35]	4	993
Plastic [35]	2	1650
Daily Electricity Energy [35]	6	365
Weather Izmir [35]	9	1461
Auto MPG6 [35]	5	392
Mackey-Glass Chaotic Time Series Prediction [36]	4	3000
Chemical Process Concentration Readings Prediction [6]	3	194
Chemical Process Temperature Readings Prediction [6]	3	223
Gas Furnace Prediction [6]	6	293

realized by approximating what is learned by the use of fuzzy *C*-means (FCM) [38], owing to its popularity. The number of triangular membership functions tuned by FCM is set to 6 for each conditional attribute across all datasets, making a fair and common start point for comparison. Whilst, different conditional attributes have their own underlying value domains, they are normalized to the common range of 0–1 before fuzzification. In terms of rule consequent, to reflect the nature of prediction problems and also to maintain the consistency in the experiments, the consequent learned in all prediction rules are represented by isosceles triangular fuzzy sets with each having $1/5$ of its domain range.

The prediction performance is measured using the root mean square error (RMSE) as defined by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{c}(y_i^* - y_i)^2}{c}} \qquad (5)$$

where y_i^* and y_i represent the predicted and target outcomes of testing samples t_i, $i = 1, 2, \ldots, c$, respectively, and c stands for the cardinality of the testing dataset. To obtain a defuzzified value as the predicted outcome, the classical defuzzification method that uses the centroid of the area under the output fuzzy set is employed. Generally, the smaller the RMSE values are, the more accurate the prediction is. To avoid potential influence of noise in judging the prediction quality, the testing results presented below are the averaged ones using ten times fivefold cross validation per dataset.

Throughout all the experiments carried out, the implementation of LLC- and LS-based attribute weighting methods adopts the existing component tools for feature evaluation from Feature Selection Library (MATLAB Toolbox) [39]. If desired, several parameters of these two methods may be tuned in order to potentially optimize the solution for each particular problem. However, for fair comparison, the experiments carried out herein do not attempt to exhaustively tune the parameters but use the default values embedded in the toolbox.

B. Experimental Results

Results are to be presented and discussed in comparison to alternative approaches, supported by statistical analysis.

1) Prediction Accuracy: Table II shows the averaged prediction errors directly computed using (5), and the corresponding standard deviation (SD) values. In this table, the column under the heading of *Non-Weighted* lists the calculated RMSEs for the testing data obtained, by the use of CRI working together with the original nonweighted T-FRI. The middle two columns, LLC and LS, list the RMSEs achieved by the proposed approach with the conditionals weighted using either LLC or LS, respectively. Last but not least, the column of *AVG_Proposed* shows the averaged prediction RMSEs between the two attribute weighted T-FRI methods. From these prediction RMSEs, it can be seen that across all datasets, the proposed approach outperforms the conventional T-FRI (that has now been strengthened with the use of CRI). This general result is not affected by the use of either of the two attribute weighting methods, comparing the RMSEs obtained by using LLC- and LS-weighted approaches.

The above results are measured on the predicted outcomes over different problem domains, showing different orders of the error scale. To facilitate a better comparison amongst different methods across all datasets, the RMSE and SD values in Table II are normalized into the range of [0, 1] per dataset, with the averaged values calculated across all datasets being presented in Table III. A clearer comparison can now be made regarding the relative performances of the different methods investigated. Using either of the weighted rule interpolation-based prediction systems, the averaged RMSE is much smaller in relation to that achievable by the nonweighted T-FRI. This indicates that introducing weights to individual rule conditional attributes leads to more accurate prediction, and that the weights obtained by artificially learning from the original sparse rule bases are effective for distinguishing the contributions of their corresponding attributes upon the prediction outcome. Moreover, the relatively lower SD values in Table II (those figures following the RMSEs), obtained by the use of weighted FRI systems in almost all datasets, further demonstrate the robustness of the proposed work, which are further verified in Table III. Interestingly, this superior prediction performance conforms to the general results achievable by running the weighted rule interpolative reasoning system that is tailored for classification problems (see [28] for detail).

Apart from the prediction error and its SD, it is important to investigate whether the improvement of the attribute weighted approach over the nonweighted FRI is of statistical significance. Table IV presents the *p*-values (in the range of [0, 1]) returned from the statistical pairwise *t*-test between the attribute weighted (i.e., LLC- and LS-based) T-FRI and the conventional nonweighted T-FRI. Given the null hypothesis that there is no significant difference between the two compared approaches, small values of *p* indicate doubt regarding such a hypothesis. As can be seen from this table, both LLC and LS weighting methods lead to rather small *p*-values for almost all datasets. In most cases, the test results reject the null hypothesis at a quite low significance level. In this table, the asterisk sign (∗) indicates that the improvement made by the LLC/LS-weighted T-FRI over the nonweighted T-FRI is validated at the 5% significance level (as commonly

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE TRANSACTIONS ON CYBERNETICS

TABLE II
AVERAGE RMSE AND SD IN TEN TIMES FIVEFOLD CROSS VALIDATION

Dataset	Non-Weighted	LLC	LS	AVG_Proposed
Abalone	2.5124 ± 0.0714	2.4315 ± 0.0607	2.4963 ± 0.0658	2.4639 ± 0.0632
Concrete Compressive Strength	9.1938 ± 0.4948	8.6094 ± 0.5598	8.9908 ± 0.5207	8.8001 ± 0.5402
Concrete Slump Test	4.2263 ± 0.3960	3.9849 ± 0.3231	3.3304 ± 0.6013	3.6576 ± 0.4622
Laser	12.4856 ± 1.7795	12.0280 ± 1.4927	12.3016 ± 1.7496	12.1648 ± 1.6211
Plastic	2.2664 ± 0.1483	2.1878 ± 0.1188	2.1957 ± 0.1599	2.1917 ± 0.1393
Daily Electricity Energy	0.4726 ± 0.0456	0.4539 ± 0.0433	0.4621 ± 0.0356	0.4580 ± 0.0394
Weather Izmir	2.3232 ± 0.1077	2.2700 ± 0.1027	2.2873 ± 0.1126	2.2786 ± 0.1076
Auto MPG6	3.0698 ± 0.3662	2.9062 ± 0.2438	2.9275 ± 0.3268	2.9168 ± 0.2853
Mackey-Glass Chaotic Time Series Prediction	0.0432 ± 0.0047	0.0398 ± 0.0040	0.0384 ± 0.0030	0.0391 ± 0.0035
Chemical Process Concentration Readings Prediction	0.3331 ± 0.0225	0.3140 ± 0.0185	0.3169 ± 0.0213	0.3154 ± 0.0199
Chemical Process Temperature Readings Prediction	0.4221 ± 0.0329	0.3611 ± 0.0326	0.3736 ± 0.0189	0.3673 ± 0.0257
Gas Furnace Prediction	0.6527 ± 0.0484	0.6202 ± 0.0354	0.6226 ± 0.0374	0.6214 ± 0.0364

TABLE III
COMPARISON ON AVERAGED RMSE AND SD ACROSS DATASETS

	Non-Weighted	LLC	LS	AVG_Proposed
RMSE	1.0000	0.0851	0.2897	0.1869
SD	0.7904	0.2780	0.5252	0.3999

TABLE IV
P-VALUE IN STATISTICAL PAIRWISE t-TEST ANALYSIS

Dataset	LLC	LS
Abalone	$0.0386(*)$	$7.25 \times 10^{-4}(*)$
Concrete Compressive Strength	$0.0442(*)$	$0.0005(*)$
Concrete Slump Test	$0.0039(*)$	$0.0027(*)$
Laser	0.2707	$1.00 \times 10^{-6}(*)$
Plastic	$2.00 \times 10^{-5}(*)$	0.0960
Daily Electricity Energy	$2.98 \times 10^{-9}(*)$	0.3530
Weather Izmir	$5.24 \times 10^{-11}(*)$	0.2988
Auto MPG6	$0.0433(*)$	$3.85 \times 10^{-10}(*)$
Mackey-Glass Chaotic Time Series Prediction	$0.0197(*)$	$0.0323(*)$
Chemical Process Concentration Readings Prediction	$0.0001(*)$	$0.0032(*)$
Chemical Process Temperature Readings Prediction	$0.0021(*)$	$1.91 \times 10^{-4}(*)$
Gas Furnace Prediction	$0.0154(*)$	$0.0139(*)$

TABLE V
COMPARISON WITH CC ON RMSE IN TEN TIMES FIVEFOLD CROSS VALIDATION

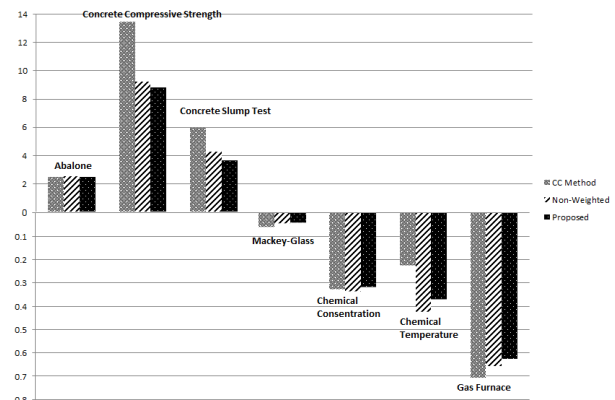| Dataset	CC	Non-Weighted	Proposed
Abalone	2.45	2.5124	2.4639
Concrete Compressive Strength	13.44	9.1938	8.8001
Concrete Slump Test	5.91	4.2263	3.6576
Mackey-Glass Chaotic Time Series Prediction	0.0597	0.0432	0.0391
Chemical Process Concentration Readings Prediction	0.3248	0.3331	0.3154
Chemical Process Temperature Readings Prediction	0.2241	0.4221	0.3673
Gas Furnace Prediction	0.7035	0.6527	0.6214

Fig. 1. Comparison with CC on RMSE across Datasets.

used), implying that statistically, the attribute weighted T-FRI significantly outperforms the original nonweighted version.

2) Comparison With State-of-the-Art Weighted FRI: This part of experimental study compares the proposed work with the state-of-the-art weighted FRI mechanism, which is reported in [23] and is referred to as the CC method (or simply, CC) below. Table V and Fig. 1 show the comparative results of RMSEs over seven prediction problems that have been used by CC, including both multivariate regression and time series prediction tasks. Note the different scales used to present the results in Fig. 1, in an effort to reduce the impact of significant differences in the output domains of different problems. For fair comparison, the settings regarding the partition of input and output attributes follow the same definition as indicated in the original work of [23].

To minimize any potential bias against the use of a particular attribute evaluation method, the averaged performance between the two implementations of the proposed approach is shown here, in the *Proposed* column (which is taken from Table II). As empirically proven in [23], CC already outperforms six classical nonweighted and weighted FRI techniques in dealing with these seven prediction problems. In particular, the conventional T-FRI (which is denoted as the HS method in [23] without including the use of CRI) has been shown to be of less accurate performance amongst the competitors. Still, the present fuzzy sparse rule-based inference scheme, by integrating the general CRI for those matched observations and the weighted T-FRI for the unmatched ones

TABLE VI
COMPARISON WITH CC ON RMSE ACROSS DATASETS

Dataset	CC	Non-Weighted	Proposed
Abalone	0.0000	1.0000	0.2227
Concrete Compressive Strength	1.0000	0.0848	0.0000
Concrete Slump Test	1.0000	0.2524	0.0000
Mackey-Glass Chaotic Time Series Prediction	1.0000	0.1990	0.0000
Chemical Process Concentration Readings Prediction	0.5310	1.0000	0.0000
Chemical Process Temperature Readings Prediction	0.0000	1.0000	0.7232
Gas Furnace Prediction	1.0000	0.3812	0.0000
Average	0.6472	0.5596	0.1351

(using the weights learned from the original sparse rule base alone), produces much more accurate results for five problems, basically ties one with CC and only underperforms with respect to CC for the dataset "chemical temperature." These results can be seen in Table V and also from Fig. 1.

To examine the overall relative performance across all seven datasets that the compared systems have been run on, as with Section IV-B1, normalization on RMSEs is carried out per dataset. The resultant averaged relative RMSEs between the different approaches are shown in Table VI. It reaffirms that the proposed approach has the smallest error in five datasets out of the seven, whilst in the other two cases it still beats the conventional nonweighted T-FRI. As a whole, in comparison to CC, the averaged relative RMSE is significantly lower (0.1351 versus 0.6472 out of a universal maximum of 1.0). The relative error reduction of $0.5121 (= 0.6472 - 0.1351)$ stands for an over 50% increase in prediction accuracy overall. In addition, this table also shows that with an averaged RMSE reduction of $0.0876 (= 0.6472 - 0.5596)$, combining CRI and conventional T-FRI helps to improve the performance of nonweighted T-FRI to supersede that of CC, although this can be expected to certain extent given the employment of CRI. Collectively, these results positively reflect the significant potential of the proposed work.

V. CONCLUSION

This paper has presented a fuzzy sparse knowledge-based interpolative reasoning scheme for performing prediction tasks, where a weighted FRI is embedded. Without requiring any observation or running the underlying FRI, the proposed method can automatically determine the relative importance of rule conditional attributes by the use of the given sparse rule base only. This approach integrates such learned weights explicitly with all computational steps involving in the interpolation process. The implemented work also enables fuzzy prediction systems to use the conventional CRI and the weighted FRI technique. The T-FRI is employed to realize the proposed approach. The implemented system has been checked against the use of 12 benchmark prediction tasks, involving both multivariate regression and time series, systematically outperforming both the nonweighted T-FRI and the state-of-the-art weighted FRI technique.

Apart from T-FRI, many other FRI methods are also effective in performing reasoning given a sparse rule base. Thus, it would be interesting to further develop the ideas proposed herein for use with those techniques. Also, the problem of the curse of dimensionality may arise due to the production of the artificial training instances from the given rule base, as the number of missing rule conditionals increases despite only a sparse rule base is involved. Thus, it is desirable to increase the algorithmic efficiency while revising the work. Potential solutions to this include: 1) to exploit feature selection techniques (e.g., [40]) to restrain the learning process and 2) to explore link-based analysis tools (e.g., [41]) to better associate and refine the rules and rule conditions.

Finally, the proposed weighted T-FRI currently works on a static rule base. Yet, a volume of intermediate fuzzy rules are typically generated while executing rule interpolation. From this, the ideas of [42] can be exploited to enrich the rule base by refining and promoting these intermediate rules, gaining efficiency by allowing for more direct rule-firing without running the interpolation procedure. In particular, the attribute weights in this paper may help leading to a weighted assembly of additional rules, thereby improving the performance of the emerged rule base by considering different importance levels amongst the rule conditionals. Nonetheless, in general, any addition or removal of certain original rules will affect the weights induced from the given rule base, which in turn, will affect the interpolated results. The exact influence upon the interpolative reasoning process remains a piece of important further research.

REFERENCES

[1] E. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY, USA: Springer, 2008.

[2] B. I. Ahmad, J. K. Murphy, P. M. Langdon, and S. J. Godsill, "Bayesian intent prediction in object tracking using bridging distributions," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 215–227, Jan. 2016.

[3] X. Sun *et al.*, "Camera-assisted video saliency prediction and its applications," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2520–2530, Sep. 2018.

[4] K.-S. Shin, T. Lee, and H.-J. Kim, "An application of support vector machines in bankruptcy prediction model," *Expert Syst. Appl.*, vol. 28, no. 1, pp. 127–135, 2005.

[5] T. Wang, Z. Han, J. Zhao, and W. Wang, "Adaptive granulation-based prediction for energy system of steel industry," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 127–138, Jan. 2018.

[6] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[7] M. Xu, M. Han, T. Qiu, and H. Lin, "Hybrid regularized echo state network for multivariate chaotic time series prediction," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2018.2825253.

[8] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.

[9] P.-C. Chang and C.-H. Liu, "A TSK type fuzzy rule based system for stock price prediction," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 135–144, 2008.

[10] N. K. Kasabov and Q. Song, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Apr. 2002.

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

[12] L. 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.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE TRANSACTIONS ON CYBERNETICS

[13] L. Kóczy and K. Hirota, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Inf. Sci.*, vol. 71, nos. 1–2, pp. 169–201, 1993.

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

[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] C. Chen *et al.*, "Rough-fuzzy rule interpolation," *Inf. Sci.*, vol. 351, pp. 1–17, Jul. 2016.

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

[18] 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.

[19] 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.

[20] Y. Yam, M. 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.

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

[22] 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.

[23] 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, Feb. 2016.

[24] S.-H. Cheng, S.-M. Chen, and C.-L. Chen, "Fuzzy interpolative reasoning based on ranking values of polygonal fuzzy sets and automatically generated weights of fuzzy rules," *Inf. Sci.*, vol. 325, pp. 521–540, Dec. 2015.

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

[26] 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.

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

[28] F. Li, C. Shang, Y. Li, J. Yang, and Q. Shen, "Fuzzy rule-based interpolative reasoning supported by attribute ranking," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2758–2773, Oct. 2018.

[29] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, nos. 1–2, pp. 155–176, 2003.

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

[31] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 507–514.

[32] 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.

[33] 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.

[34] D. Dheeru and T. Karrai. (2017). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[35] J. Alcalá-Fdez *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, nos. 2–3, pp. 255–287, 2011.

[36] R. S. Crowder, "Predicting the Mackey-Glass time series with cascade correlation learning," in *Proc. Connectionist Models Summer School*, 1990, pp. 117–123.

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

[38] J. C. Bezdeks, R. Ehrlich, and W. Full, "FCM: The fuzzy *c*-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.

[39] G. Roffo, S. Melzi, and M. Cristani. *Feature Selection Library (MATLAB Toolbox)*. Accessed: Aug. 10, 2018. [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/56937-feature-selection-library

[40] 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.

[41] T. Boongoen, Q. Shen, and C. Price, "Disclosing false identity through hybrid link analysis," *Artif. Intell. Law*, vol. 18, no. 1, pp. 77–102, 2010.

[42] N. Naik, R. Diao, and Q. Shen, "Dynamic fuzzy rule interpolation and its application to intrusion detection," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 1878–1892, Aug. 2018.

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

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

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

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

Changjing Shang received the Ph.D. degree in computing and electrical engineering from Heriot-Watt University, Edinburgh, U.K., in 1995.

She was with Heriot-Watt University, Loughborough University, Loughborough, U.K., and Glasgow University, Glasgow, U.K. She is a University Research Fellow with the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. She has published extensively, and supervised over ten Ph.D./PDRA in the areas of pattern recognition, data mining and analysis, space robotics, and image modeling and classification.

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 appointed as the Chair of computer science and the Pro Vice-Chancellor Faculty of Business and Physical Sciences, Aberystwyth University, Aberystwyth, U.K. He has authored two research monographs and over 380 peer-reviewed papers.

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