Interpolation with Just Two Nearest Neighboring Weighted Fuzzy Rules

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

Published in:
IEEE Transactions on Fuzzy Systems

DOI:
10.1109/TFUZZ.2019.2928496

Publication date:
2020

Citation for published version (APA):
Li, F., Shang, C., Li, Y., Yang, J., & Shen, Q. (2020). Interpolation with Just Two Nearest Neighboring Weighted Fuzzy Rules. *IEEE Transactions on Fuzzy Systems*, *28*(9), 2255 - 2262. Article 8762115. https://doi.org/10.1109/TFUZZ.2019.2928496

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

Interpolation with Just Two Nearest Neighbouring Weighted Fuzzy Rules

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

***Abstract*—Fuzzy rule interpolation (FRI) enables sparse fuzzy rule-based systems to derive an interpolated conclusion using neighbouring rules, when presented with an observation that matches none of the given rules. The efficacy of FRI has been further empowered by the recent development of weighted FRI techniques, particularly the one that introduces attribute weights of rule antecedents from the given rule base, removing the conventional assumption of antecedent attributes having equal weighting or significance. However, such work was carried out within the specific transformation-based FRI mechanism. This short paper reports the results of generalising it through enhancing two alternative representative FRI methods. The resultant weighted FRI algorithms facilitate the individual attribute weights to be integrated throughout the corresponding procedures of the conventional unweighted methods. With systematical comparative evaluations over benchmark classification problems, it is empirically demonstrated that these algorithms work effectively and efficiently using just two nearest neighbouring rules.**

***Index Terms*—Fuzzy interpolative reasoning, weighted rule interpolation, attribute weights, nearest neighbouring rules.**

I. Introduction

Fuzzy rule interpolation (FRI) plays a powerful role in performing inference within sparse rule-based reasoning systems [1]. It facilitates the derivation of an approximate consequent for an observation which has no matching rules, by the use of its neighbouring rules. Whilst the FRI literature has seen many methods (e.g., [2]–[4]) being proposed, most of which share a common assumption that the rule antecedents are of equal significance while performing rule interpolation. A recent focus of developing FRI techniques is to relax this assumption, by introducing weights to the individual antecedent attributes, such as [5], [6]. Nevertheless, these weighted FRI approaches require additional information for calculating the weights other than that contained within the sparse rule base. Besides, the resultant weights are not systematically integrated within the internal structure of the underlying FRI algorithm.

Most recently, a weighted interpolative reasoning scheme has been reported [7], where the weights of individual antecedent attributes are learned from the given knowledge (i.e., the sparse rule base) in support of attribute ranking. Such weights are explicitly integrated with the procedures of the popular scale and move transformation-based FRI (T-FRI) [3]. This has led to an outstanding performance in tackling classification problems, as empirically shown. In particular, an important finding is that only two (i.e., the minimal number of) neighbouring rules are required for the weighted T-FRI to perform, significantly reducing the computational overheads caused by otherwise running rule interpolation with more rules.

Given this exciting empirical outcome for weighted T-FRI, it is interesting to investigate whether the discovery that "least number of neighbouring rules does better" is common to other FRI methods if a similar weighting scheme is adopted. Fortunately, the weights learning mechanism as proposed in [7] is independent of the FRI process, which works by exploiting the sparse rule base only. Inspired by this observation, this short paper presents a further development that enhances two other commonly used FRI algorithms (namely, those first presented in [1] and [4]), by following the ideas of [7]. The resultant weighted FRI methods are systematically evaluated via addressing ten benchmark classification problems, in comparison with their corresponding unweighted originals. The improvement of classification accuracies is highlighted and more importantly, it is demonstrated that the best performance is achieved when the number of the nearest neighbouring rules required to perform the weighted FRI is indeed the smallest.

The rest of this paper is structured as follows. Section II reviews the relevant background, including the weighted T-FRI as proposed in [7] and the two selected representative FRI methods to be enhanced with the weighting approach. Section III presents the modification of those two FRI methods with the use of attribute weights. Section IV discusses the systematically compared experimental results. Section V draws the conclusion and gives an outline of further research.

II. Background

For completion, this section first shows the basic notations used throughout the paper, followed by a summary of the weighted T-FRI method as proposed in [7] and then, by an overview of the two representative FRI methods that will be extended with the inclusion of attribute weights.

A. Basic Notations

To support a consistent illustration of FRI, without losing generality, suppose that a (sparse) fuzzy rule base $R = \{r^1, r^2, \ldots, r^N\}$, where multiple antecedent attributes are

F. Li and J. Yang are with School of Computer Science and Engineering, Northwestern Polytechnical University, China and Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, UK. E-mail: {fal2, jiy6}@aber.ac.uk

C. Shang and Q. Shen are with Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, UK. E-mail: {cns, qqs}@aber.ac.uk

Y. Li is with School of Computer Science and Engineering, Northwestern Polytechnical University, China. E-mail: lybyp@nwpu.edu.cn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

involved but not individually weighted initially, and an observation o^* are represented in the following format:

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 B^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 rule antecedent attributes; z is the consequent attribute; A_j^i and A_j^* denote the fuzzy set values taken by a_j in the rule r^i and o^*, respectively; and B^i represents the fuzzy set value of the consequent attribute z in r^i. For simplicity, triangular membership functions are employed for all FRI methods involved in this work, where the fuzzy sets are represented by their characteristic points (CPs). Namely, fuzzy values A_j^i, A_j^*, B^i, and the consequent B^* to be computed ($i = 1, 2, \ldots, N, j = 1, 2, \ldots, m$) are expressed by $(a_{j1}^i, a_{j2}^i, a_{j3}^i)$, $(a_{j1}^*, a_{j2}^*, a_{j3}^*)$, (b_1^i, b_2^i, b_3^i), and (b_1^*, b_2^*, b_3^*), respectively, where the first and third CP stand for the two extreme points of the support with a membership value of 0 and the middle one stands for the normal point of the fuzzy set with a membership of 1.

B. Summary of Weighted T-FRI

In [7], T-FRI is extended with rules involving individual attribute weights, which are learned from the given sparse rule base by exploiting attribute ranking techniques. The attribute weighting scheme is enabled by an innovative reverse engineering procedure, which reduces the sparsity of the given rule base by generating an artificial training decision table. The essential idea is to reformulate all rules in the sparse rule base into a common representation, where each (possibly) missing value of any rule antecedent is replaced by one of the alternative fuzzy values from its domain. All these reformulated rules, artificial or original, are collated for evaluation of the relative significance degrees of the individual attributes.

In particular, the weights of the attributes are individually measured using a certain feature ranking method (which is implemented by modifying the feature evaluation mechanism extracted from a given feature selection technique). It has been shown in [7] that the underlying approach of weighted T-FRI is robust as different types of feature selection method may be adopted for such use without significantly affecting the level of performance improvement over the conventional unweighted T-FRI.

The primary motivation of introducing weights to rule antecedent attributes is to minimise the adverse effect of assuming all attributes having equal significance (which is typically made in conventional FRI methods, but is often impractical). In weighted T-FRI, individual attribute weights are integrated with every procedure of the unweighted T-FRI algorithm, including: the selection of the nearest neighbouring (aka. the closest) rules, the construction of intermediate rules, and the computation of scale and move transformation factors. Here, the nearest rules to an unmatched observation are selected to build the intermediate rule, based on similarity, to form the basis upon which to perform interpolative transformations. In implementation, all computational steps in the original T-FRI, which involve evenly calculated average of the attribute values, are now improved by a weighted aggregation of the corresponding components.

Detailed computation mechanisms are however, beyond the scope of this short paper but can be found in [7]. Note that when all rule antecedents are assumed to be of equal importance (i.e., all attribute weights are of the same value), the weighted T-FRI degenerates to its original unweighted version.

C. Outline of KH Rule Interpolation

The KH rule interpolation (named after its inventors [1]) offers an initial proposal for fuzzy interpolative reasoning through manipulating α-cut distances. It has been subsequently developed to address sparse rule interpolation involving multiple rules with multiple antecedent variables [8], [9].

When a given observation fails to match any rule in the sparse rule base for firing, an interpolated consequent is constructed basically by performing a linear aggregation of the rule consequents of the neighbouring rules closest to the observation. The aggregation operation complies with the general principle of similarity-based analogical reasoning, such that

The closer a rule's antecedent A^i (which is a logical aggregation of individual attribute values A_j^i) to the observation o^*, the closer the rule's consequent B^i to the outcome B^* that corresponds to o^*.

The similarity measure employed is specified by the use of fuzzy distances defined between a rule antecedent and the observation. That is, the smaller distance between A^i and o^* is, the more similar they are, thereby B^i is deemed to potentially make more contribution to the consequent being sought.

Thanks to the piecewise linear property presumed by KH interpolation, given triangular membership functions, the interpolated result $B^* = (b_1^*, b_2^*, b_3^*)$ can be determined with its two α-cut (i.e., $\alpha = 0, 1$), resulting in the three CPs taking the values of

$$b_t^* = \frac{\sum_{i=1}^{n} \frac{1}{\sqrt{\sum_{j=1}^{m} (a_{jt}^i - a_{jt}^*)^2}} b_t^i}{\sum_{i=1}^{n} \frac{1}{\sqrt{\sum_{j=1}^{m} (a_{jt}^i - a_{jt}^*)^2}}} \tag{1}$$

where n is the number of the neighbouring rules used for interpolation, m is the number of attributes in the rule, and $t = 1, 2, 3$.

D. Outline of CCL Rule Interpolation

The CCL rule interpolation (again, named after its inventors [4]) offers an alternative means for fuzzy interpolative reasoning that exploits the areas of the fuzzy sets involved in the rules and the (unmatched) observation. The idea is to preserve the logically consistent properties with respect to the ratios of fuzziness, which is determined by the areas of fuzzy sets. The core computations are summarised below in relation to the use of triangular fuzzy membership functions.

First, the normal point b_2^* of the (to be) interpolated consequent B^* is defined by linear interpolation such that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

$$b_2^* = \sum_{i=1}^{n} W_i b_2^i \tag{2}$$

in which n is the number of selected rules for interpolation, and W_i is the aggregated rule weight, which is calculated by

$$W_i = \frac{\sum_{j=1}^{m} w_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}}, \qquad w_{ij} = 1 - \left| \frac{a_{j2}^i - a_{j2}^*}{max_{a_{j2}} - min_{a_{j2}}} \right| \tag{3}$$

where $max_{a_{j2}}$ and $min_{a_{j2}}$ are used for normalisation, denoting the maximal and minimal value within $\{a_{j2}^i | i = 1, 2, \ldots, n\}$.

From this, the left triangular area $S_L(B^*)$ (i.e., the part of the geometrical area of a triangular fuzzy set on the left hand side of the normal point) and the right triangular area $S_R(B^*)$, of the fuzzy set B^* are calculated by Eqn. (4) (see next page), where in the subscript of $S_K(*)$, $K \in \{L, R\}$.

Finally, the left and right extreme points of the support for the interpolated result B^* are derived from the resulting triangular areas as follows:

$$b_1^* = b_2^* - 2S_L(B^*), \qquad b_3^* = b_2^* + 2S_R(B^*) \tag{5}$$

III. Improvement of FRI with Attribute Weighting

In this section, the two representative unweighted FRI methods, respectively outlined in Section II-C and II-D are generalised by integrating the weights of rule antecedent attributes within the underlying FRI procedures. Note that the mechanism for learning the attribute weights $AW_j, j = 1, 2, \ldots, m$, from the given sparse rule base remains exactly the same as that presented in [7] and hence, is omitted here.

A. Weighted KH Rule Interpolation

The attribute weights learned from a given sparse rule base reveal the relative significance degrees of the individual antecedent attributes, in terms of their potential in deriving the consequent given an observation. The main issue of embedding such weights within an FRI method is how to adapt the original computational mechanism of the unweighted FRI [7]. This is in order to ensure that the individually weighted attributes are aggregated in a way to better reflect their respective contributions in the interpolation process of the consequent.

As the attribute weights are learned independently of the interpolative reasoning process, all that is needed to develop a weighted version of the KH interpolation method is to modify its procedures that involve the use of α-cut distances by considering the weights accordingly. This can be carried out so that the distances are measured by taking into consideration of the relevant significance degrees of the attributes. Thus, the original unweighted KH interpolation can be extended in a straightforward manner, by computing the CPs of the interpolated consequent as per Eqn. (1) through the following weighted calculation:

$$\tilde{b_t^*} = \frac{\sum_{i=1}^{n} \frac{1}{\sqrt{\sum_{j=1}^{m} AW_j (a_{jt}^i - a_{jt}^*)^2}} b_t^i}{\sum_{i=1}^{n} \frac{1}{\sqrt{\sum_{j=1}^{m} AW_j (a_{jt}^i - a_{jt}^*)^2}}}, \quad t = 1, 2, 3 \tag{6}$$

Note that if the assumption of attributes having equal significance is applied, that is $AW_j, j = 1, 2, \ldots, m$ are of the same value, the above formula degenerates to the original version, i.e., Eqn (1). As such, this weighted KH method is a generalised version of the original, still working as previously in the event where no weighting scheme is applicable or necessary.

B. Weighted CCL Rule Interpolation

The original CCL FRI procedure as per Section II-D can be generalised in a similar manner to the above. In particular, the attribute weights are integrated in the construction of the normal point b_2^* and also, in the computation of the triangular area $S_K(B^*)$ of the interpolated consequent.

In particular, the normal point b_2^* can be specified by the weighted aggregation of rule consequents of the selected neighbouring rules, where the rule weights W_i of Eqn. (3) are redefined by normalising the aggregated weight of each entire rule antecedent per rule. Note that in the original CCL method, the aggregation of rule weights is implemented by arithmetic average. Thus, the modified rule weight $\tilde{W_i}$ is now extended to

$$\tilde{W_i} = \frac{\sum_{j=1}^{m} AW_j w_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} AW_j w_{ij}} \tag{7}$$

Intuitively, the average operation imposed over the rule antecedents also needs to be applied to the computation of the interpolated consequent fuzzy set. This leads to the corresponding modification of the area of the interpolated consequent fuzzy set, from Eqn. (4) to Eqn. (8). In this extension, the attribute weights $AW_j, j = 1, 2, \ldots, m$, are different from the weighting terms w_{ij} used in the original method which are still required to be computed in the same way as the original. Together, they are used to construct modified overall rule strengths. In effect, AW_j adjusts w_{ij} to better reflect the contribution of each individual antecedent attribute in relation to its significance, towards the calculation of the overall rule weight in deriving the consequent.

As with the weighted KH method, the above newly introduced rule weight $\tilde{W_i}$ and interpolated consequent area $\tilde{S_K}(B^*)$ also degenerate back to their original counterparts in the unweighted version if all attribute weights are equal, in terms of their relative significance. Indeed, in this case, $AW_j = 1/m, \forall j = 1, 2, \ldots, m$.

IV. Experimental Evaluation

This section presents a systematic experimental comparison among the proposed weighted KH, CCL and T-FRI, against their originals that do not involve individual attribute weights. The comparative investigation is performed over ten benchmark classification problems, most of which are of multiple class labels. The changes of classification accuracy with respect to the number of nearest neighbouring rules selected for interpolation are examined, demonstrating the efficacy of weighted FRI algorithms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

$$S_K(B^*) = \begin{cases} \left(\sum_{j=1}^{m} S_K(A_j^*)\right) \times \left(\sum\limits_{\substack{i=1,\\ \exists j S_K(A_j^i)>0}}^{n} W_i \times \frac{S_K(B^i)}{\sum_{j=1}^{m} S_K(A_j^i)} \right), \text{ if } \exists ij S_K(A_j^i) > 0 \\ \frac{\sum_{j=1}^{m} S_K(A_j^*)}{m}, \text{ if } \forall ij S_K(A_j^i) = 0 \end{cases} \tag{4}$$

$$\tilde{S_K}(B^*) = \begin{cases} \left(\sum_{j=1}^{m} AW_j S_K(A_j^*)\right) \times \left(\sum\limits_{\substack{i=1,\\ \exists j S_K(A_j^i)>0}}^{n} AW_i \times \frac{S_K(B^i)}{\sum_{j=1}^{m} AW_j S_K(A_j^i)} \right), \text{ if } \exists ij S_K(A_j^i) > 0 \\ \sum_{j=1}^{m} AW_j S_K(A_j^*), \text{ if } \forall ij S_K(A_j^i) = 0 \end{cases} \tag{8}$$

A. Experimental Set-up

1) Datasets: Ten benchmark classification datasets are taken from KEEL (Knowledge Extraction based on Evolutionary Learning) [10] and UCI machine learning [11] dataset repositories, with details summarised in Table I.

TABLE I
DATASETS USED FOR CLASSIFICATION

Dataset	#(Attributes)	#(Classes)	#(Instances)
Diabetes	8	2	768
Phoneme	5	2	5404
Magic	10	2	1902
Haberman	3	2	306
Hayes-Roth	4	3	160
Page-blocks	10	5	5472
Ecoli	7	8	336
Red Wine Quality	11	11	1599
Wireless Indoor Localisation	7	4	2000
User Knowledge Modelling	5	4	403

2) Experimental methodology: As indicated previously, the proposed weighted KH, CCL and also T-FRI methods and their original versions (those given in [9], [4], and [3] respectively) adopt triangular membership functions to represent fuzzy values. A primitive three-valued fuzzy partition (as shown in Fig. 1) is employed after normalisation over all datasets, for fair comparison as well as for illustrative simplicity.

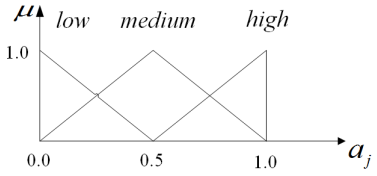

Fig. 1. Membership functions defining values of antecedent attributes.

The comparative experiments are performed via 5 times 10-fold cross validation per dataset. The rule base for each problem is learned from the training data. The classical rule induction technique of [12] is employed to generate an initial rule base, where 40% of the learned rules are purposefully removed randomly, resulting in a rather sparser rule base to better evaluate the performance of each FRI method. The attribute weights are then derived from the resultant sparse rule base, by the use of information gain (IG) for scoring each individual rule antecedent. Note that only IG is employed herein to compute attribute weights, because it has been shown in [7] that any of the popular feature ranking methods may be utilised to perform attribute weighting without incurring much performance deviation.

For testing, each new observation is checked against the rules in the rule base first, the consequent is calculated by aggregating the outcomes of firing the matched rules. If however, no matching is found, FRI methods are applied to derive an interpolated consequent (only one FRI method is applied at once of course, weighted or not).

Further to the comparative studies carried out between weighted FRI methods and their original unweighted ones, a series of experiments are conducted to investigate the variation of classification accuracy in relation to the number n of the nearest neighbouring rules selected for interpolation. For consistency, as with the work in [7], five different cases are compared regarding the cases where n is set to 2, 3, 4, 5, 6, respectively (whilst it makes little sense, both computationally and intuitively, to use any larger number of rules for interpolation). Also, for fair comparison, the selection scheme for the nearest neighbouring rules as described in [13] is employed to determine the closest rules that are required to implement the interpolation, across all six (three weighted and three original) methods compared.

B. Results and Discussion

1) Effectiveness of weighted FRI: Table II shows the classification accuracies calculated by averaging the 5 times 10-fold cross validation, for each of the six methods: three originals and three extended methods enhanced with the weighting scheme. The performances of weighted methods are directly compared against those of their originals, where two nearest neighbouring rules to the testing observation are selected for interpolation (unless otherwise stated). The results are presented in the column of *Weighted* and that of *Ori*, respectively.

As indicated previously, a significant portion (40%) of rules are randomly removed from the original learned rule base for each classification problem, in order to thoroughly compare the performance of weighted interpolation against the unweighted. In so doing, more opportunities may be generated for those

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

TABLE II
AVERAGE CLASSIFICATION ACCURACIES (%) BY INTERPOLATION WITH TWO NEAREST NEIGHBOURING RULES

Dataset	FRI	Sparser Rule Base		Full Rule Base	
		Ori	Weighted	Ori	Weighted
Diabetes	T-FRI	61.19	68.98	63.85	65.13
	KH	59.07	65.11	62.37	62.87
	CCL	60.92	66.83	63.91	64.61
Phoneme	T-FRI	53.89	65.85	65.51	66.18
	KH	58.07	60.79	64.22	64.41
	CCL	63.21	66.48	65.83	65.93
Magic	T-FRI	63.86	69.16	69.03	69.60
	KH	64.39	67.17	69.37	69.71
	CCL	65.71	68.43	69.51	69.74
Haberman	T-FRI	72.49	77.19	74.02	74.54
	KH	69.21	73.39	72.99	73.52
	CCL	70.91	74.77	73.93	74.46
Hayes-Roth	T-FRI	46.87	61.00	55.25	56.62
	KH	47.75	58.00	54.87	56.75
	CCL	46.37	55.75	55.50	56.75
Page-blocks	T-FRI	66.77	72.13	69.78	69.80
	KH	65.18	69.93	70.13	70.15
	CCL	66.71	72.07	69.76	69.76
Ecoli	T-FRI	59.52	65.96	62.50	65.86
	KH	59.89	64.04	62.71	65.81
	CCL	61.56	65.90	63.12	66.31
Red Wine Quality	T-FRI	52.98	57.37	52.54	53.89
	KH	52.52	53.92	52.25	53.62
	CCL	52.73	53.33	52.44	53.32
Wireless Indoor Localization	T-FRI	76.36	79.89	79.34	80.03
	KH	77.50	78.85	79.90	80.87
	CCL	75.59	77.18	78.84	79.85
User Knowledge Modeling	T-FRI	74.85	82.54	74.97	78.24
	KH	69.63	75.24	71.14	73.96
	CCL	70.53	74.22	71.52	74.18
Average	T-FRI	62.87	70.01	66.68	67.99
	KH	62.32	66.64	65.99	67.17
	CCL	63.42	67.50	66.43	67.49

observations that find no rules to match. However, for potential practical applications, it is also desirable to investigate how much better the proposed methods do their job than their originals if more rules are available. For this purpose, Table II shows not only the comparative results obtained when the FRI methods are applied using artificially created sparser rule bases, but also the outcomes when they work with the entire learned rule bases.

As reflected in this table, by comparing the two right-most columns, not very significant improvement is gained by the weighted FRI methods if the full rule bases are employed. This can be expected as most new observations may match certain rules to fire in the first place. However, when the number of samples requiring interpolation becomes large, as per the situation of running a sparser rule base, each of the three weighted FRI methods significantly outperforms its corresponding unweighted method for almost all datasets.

Table III lists the (rounded) average numbers of testing samples that are unmatched by the sparse rule bases and those unmatched by the original rules. Despite the fact that there are significantly larger numbers of unmatched rules in the cases where a sparse rule base is employed, the average classification accuracies (across the ten datasets) obtained using the weighted methods beat those achievable using the full original rule bases. From the perspective of obtaining improved classification accuracy rates, this clearly demonstrates the potential of the present work.

TABLE III
AVERAGE NUMBER OF TESTING SAMPLES FOR INTERPOLATION

Dataset	Samples Requiring Interpolation in Sparser Rule Base / Total	Samples Requiring Interpolation in Full Rule Base / Total
Diabetes	58 / 77	31 / 77
Phoneme	259 / 540	52 / 540
Magic	96 / 190	45 / 190
Haberman	10 / 31	2 / 31
Hayes-Roth	9 / 16	3 / 16
Page-blocks	207 / 547	8 / 547
Ecoli	21 / 33	15 / 33
Red Wine Quality	144 / 160	99 / 160
Wireless Indoor Localization	146 / 200	78 / 200
User Knowledge Modeling	28 / 40	18 / 40

More particularly, the average improvements of the weighted T-FRI, weighted KH and weighted CCL on all ten datasets over the unweighted ones are measured to be 7.14%, 4.32%, and 4.08%, respectively. This is statistically significance as verified by pairwise *t*-tests, which result in low p values as listed in the third column of Table IV. Again, these results show that the weighted FRI methods significantly enhance the interpolative performance of the unweighted ones, and that such superior performance is attained under the condition that only two nearest neighbouring rules are employed for interpolation.

2) Efficiency of weighted FRI: The previous work on weighted T-FRI (see [7]) produced a surprising and very positive result, discovering that the use of the minimum number of nearest neighbouring rules does better for such rule interpolation. Inspired by that discovery, this part of the experimental investigation systematically looks into the effect of varying the number of neighbouring rules used for interpolation across all three weighted methods. The investigation is carried out for all aforementioned ten datasets, using five different numbers of closest rules.

Note that attribute weights can also be exploited to help modify the selection procedure for the nearest neighbouring rules (see [7] for details). Thus, in order to thoroughly examine the implication of the weighting scheme upon both the procedure for closest rules selection and that for rule interpolation, the experiments on classification results are herein purposefully designed to cover the following all four cases, for each particular FRI approach (be it T-FRI, KH or CCL): unweighted selection with unweighted interpolation, unweighted selection with weighted interpolation, weighted selection with unweighted interpolation, and weighted selection with weighted interpolation. These are denoted as $S_{\bar{w}}I_{\bar{w}}$, $S_{\bar{w}}I_w$, $S_wI_{\bar{w}}$ and S_wI_w respectively. Of course, if the number of neighbouring rules is set to two, then the first and the last become exactly the same as those denoted by *Ori* and *Weighted* as previously given in Table II that run on a sparse rule base.

Tables V, VI and VII (with Tables VI and VII being the continuations of Table V due to the limit of the physical space)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

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

Dataset	FRI	Ori vs. Weighted ($n = 2$)	$n = 2$ vs. $n = 3$ (Weighted FRI)
Diabetes	T-FRI	8.50×10^{-6}	4.58×10^{-4}
	KH	1.44×10^{-6}	0.0254
	CCL	3.84×10^{-6}	3.16×10^{-5}
Phoneme	T-FRI	3.26×10^{-6}	1.20×10^{-4}
	KH	6.60×10^{-6}	0.1156
	CCL	1.29×10^{-4}	5.36×10^{-4}
Magic	T-FRI	4.02×10^{-5}	0.0211
	KH	1.68×10^{-5}	0.6869
	CCL	6.14×10^{-7}	0.0719
Haberman	T-FRI	1.77×10^{-5}	0.0018
	KH	1.88×10^{-4}	0.0074
	CCL	3.48×10^{-5}	9.40×10^{-4}
Hayes-Roth	T-FRI	1.66×10^{-5}	0.0300
	KH	2.98×10^{-7}	0.0155
	CCL	2.91×10^{-5}	0.1575
Page-blocks	T-FRI	1.62×10^{-5}	1.29×10^{-4}
	KH	8.26×10^{-5}	6.82×10^{-5}
	CCL	2.50×10^{-5}	7.24×10^{-4}
Ecoli	T-FRI	1.84×10^{-6}	5.96×10^{-5}
	KH	2.80×10^{-4}	6.22×10^{-4}
	CCL	5.24×10^{-5}	1.52×10^{-4}
Red Wine Quality	T-FRI	2.03×10^{-6}	0.0025
	KH	0.0559	0.0152
	CCL	0.0423	0.2836
Wireless Indoor Localization	T-FRI	1.79×10^{-5}	0.0019
	KH	9.55×10^{-4}	0.0089
	CCL	0.0020	0.1940
User Knowledge Modeling	T-FRI	1.50×10^{-5}	2.26×10^{-4}
	KH	0.0026	0.0075
	CCL	0.0030	8.13×10^{-4}

present the results of this set of experiments, with the examined range of n set to $\{2, 3, \ldots, 6\}$. This is partly to facilitate direct comparison with the state-of-the-art results provided in [7], and partly to reflect the practical consideration where using more than six closest rules to perform interpolation is of little intuitive appeal, both in terms of computational complexity and of classification result interpretability. Over this entire range, the accuracies obtained by the use of weighted interpolation generally outperform those by the unweighted for all three FRI approaches. That is in most cases, the results achieved by $S_{\bar{w}}I_w$ are improved over $S_{\bar{w}}I_{\bar{w}}$, while S_wI_w does better than $S_wI_{\bar{w}}$. These improvements further demonstrate the effectiveness of the weighted FRI methods proposed here.

Figure 2 plots the changing trend of classification accuracy in relation to the number of neighbouring rules used. As n goes up from the minimum (i.e., $n = 2$), the accuracies drop, sometimes sharply, for all three weighted FRI methods with the weighted interpolation supported by weighted rule selection (i.e., S_wI_w). This behaviour of weighted FRI for the *Magic* and *Red Wine Quality* datasets is slightly less obvious, but increasing n does not help to improve the classification performance either.

The observation that the results of S_wI_w with any FRI approach when $n = 2$ beat those when $n = 3$ is further validated by pairwise *t*-test in Table IV, with *p* values shown in the fourth column of this table. These experimental results indicate that the reduction of classification accuracies when the number of the nearest neighbouring rules is increased from 2 to 3 is statistically significant for almost all FRI methods across all datasets.

TABLE V
AVERAGE CLASSIFICATION ACCURACIES (%) VS. NUMBER OF NEAREST NEIGHBOURING RULES USED FOR DIFFERENT FRI

Dataset	FRI	n	$S_{\bar{w}}I_{\bar{w}}$	$S_{\bar{w}}I_w$	$S_wI_{\bar{w}}$	S_wI_w
Diabetes	T-FRI	2	61.19	66.50	63.69	**68.98**
		3	63.64	63.77	62.44	63.09
		4	65.15	67.13	63.87	66.00
		5	64.68	65.41	63.12	65.02
		6	65.12	66.76	64.21	65.90
	KH	2	59.07	63.81	60.21	65.11
		3	63.52	63.49	61.95	62.03
		4	63.00	64.67	61.77	63.08
		5	64.46	64.51	62.50	62.87
		6	64.28	**65.37**	62.45	63.49
	CCL	2	60.92	64.90	62.71	**66.83**
		3	62.24	62.63	59.87	60.31
		4	63.65	64.74	61.80	63.15
		5	63.39	63.65	61.72	61.56
		6	64.01	65.00	61.93	63.28
Phoneme	T-FRI	2	53.89	56.43	62.15	**65.85**
		3	55.09	55.22	58.57	62.37
		4	54.78	57.08	61.09	64.51
		5	55.59	56.74	57.75	64.52
		6	56.11	58.51	60.86	65.33
	KH	2	58.07	59.53	59.21	**60.79**
		3	59.50	59.50	59.38	59.38
		4	59.58	59.78	59.47	59.42
		5	59.34	59.40	59.35	59.39
		6	60.03	60.12	59.91	60.19
	CCL	2	63.21	64.31	65.68	**66.48**
		3	58.84	60.82	60.59	62.33
		4	63.57	64.05	64.44	64.83
		5	60.22	63.27	60.81	63.79
		6	63.90	65.06	64.49	65.32
Magic	T-FRI	2	63.86	67.86	65.46	69.16
		3	67.79	67.97	67.53	67.95
		4	67.37	68.90	67.30	68.95
		5	68.91	69.02	68.28	68.90
		6	68.78	**69.61**	68.23	69.45
	KH	2	64.39	66.29	65.01	67.17
		3	66.79	66.78	67.00	67.05
		4	66.95	66.91	66.89	67.00
		5	67.45	**67.47**	67.31	67.44
		6	67.45	67.43	67.18	67.43
	CCL	2	65.71	67.62	66.48	68.43
		3	67.58	67.78	67.57	67.81
		4	67.41	68.03	67.16	67.79
		5	**68.48**	68.29	68.17	67.91
		6	67.83	68.22	67.31	67.73
Haberman	T-FRI	2	72.49	74.44	75.36	**77.19**
		3	72.94	73.39	73.53	74.56
		4	74.38	74.77	74.19	74.25
		5	74.32	74.83	74.12	74.71
		6	74.25	74.58	74.58	74.44
	KH	2	69.21	70.26	72.28	**73.39**
		3	72.29	71.69	72.88	71.30
		4	72.21	72.35	72.35	72.42
		5	72.16	71.64	72.55	71.36
		6	72.69	71.90	72.29	71.95
	CCL	2	70.91	72.36	73.13	**74.77**
		3	71.77	72.62	72.15	72.68
		4	72.48	72.88	71.83	72.29
		5	71.97	72.30	72.09	72.29
		6	72.09	71.97	71.90	71.90

Examining the results of Tables V-VII more closely, as highlighted in bold for each of the ten datasets, the best performance of each FRI across the four implementations (namely, $S_{\bar{w}}I_{\bar{w}}$, $S_{\bar{w}}I_w$, $S_wI_{\bar{w}}$, and S_wI_w) over the entire range

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

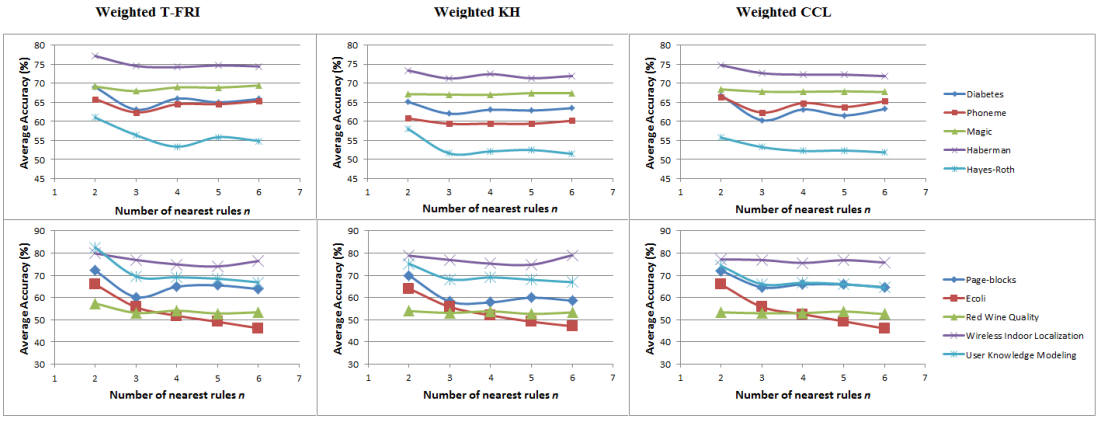

Fig. 2. Accuracy variation with number of neighbouring rules.

of n studied, is generally achieved using $S_w I_w$ with $n = 2$. However, for the conventional $S_{\bar{w}} I_{\bar{w}}$ FRI methods where no weighting scheme is employed, the accuracy increases with n. This forms a sharp contract between the weighted and unweighted approaches.

There are exceptional cases to observe. Particularly, the results show that the $S_{\bar{w}} I_w$ FRI method can do better than the rest if a large number (e.g., $n = 5$ or $n = 6$) of rules are used. Such situations occur mostly when the KH weighted interpolation method is employed with rules taken by unweighted selection. Nonetheless, the interpolation procedure is still weighted in these cases; this again demonstrates the effectiveness of weighting upon rule antecedent attributes. Besides, there is little win of $S_{\bar{w}} I_w$ over $S_w I_w$. Yet, such minor win is obtained at the expense of much more computational overheads as more rules are involved in the interpolation procedure, as shown below. This finding is of great importance in practical application of FRI since it empirically confirms that weighted FRI methods only require two (i.e., the least number of) nearest neighbouring rules to perform rule interpolation, significantly enhancing the algorithm efficiency.

3) Further analysis – Confusion matrix: The analysis of the confusion matrices has also been conducted for each of the three weighted FRI methods regarding the use of two or three nearest neighbouring rules. To save space, Tables VIII-X present the outcomes for the *Diabetes* dataset as an example case study, since the general trends for the others are similar. The comparison in each of these tables helps explain why the overall classification accuracy may dramatically decrease as n increases from 2 to 3. As reflected by these results, the adverse variation of the overall accuracy when $n = 3$ appears to be caused by the significant increase of false positives and the considerable reduction of true negatives. Of course, such situations must be minimised in any realistic application, especially for instance in medical diagnosis as is indeed the case concerning this dataset. Both increase in false positives and reduction in true negatives will usually cause undue anxiety of the patient, and in worse scenarios, may even cause missing the correct diagnosis of other disease(s) that the patient may be suffering from the given symptoms.

4) Further analysis – Run time: Results so far have demonstrated that weighted FRI methods (that involve additional computation in both rule selection and rule interpolation procedures) generally outperform their originals. However, a question may be raised as to how much extra computation effort is required to attain such improved performance, despite the recognition that learning the weights themselves is an offline task. This final experimental study therefore, addresses the natural concern regarding the run time performance of the weighted methods.

Table XI lists the average testing times recorded for all three weighted FRI methods (i.e., in the form of $S_w I_w$) and their originals (namely, $S_{\bar{w}} I_{\bar{w}}$), when dealing with the final five problems given in Table I. The tests are carried out in relation to the increase of the number of the nearest neighbouring rules employed. Note that these five cases are selected because they each involve more classes and hence, are more difficult to classify (whilst saving the space otherwise required to present similar results for the other five). As expected, there is indeed an increase in time consumption when exploiting more nearest neighbouring rules for all FRI methods (weighted or not). The use of fewer rules will thus be more efficient. However, as can be seen from this table, there is no significant increase in the time cost by a weighted FRI as compared to that by its original where no weights are involved, while using the same number of rules for interpolation. This once again demonstrates the efficacy of the proposed weighted FRI techniques and supports the outcome that least neighbouring rules do better with attribute weighted FRI.

V. Conclusion

This short paper has further developed the state-of-the-art work on fuzzy rule interpolation (FRI) [7], by extending the weighted transformation-based FRI to two other classical FRI methods, namely the KH [1] and CCL [4] algorithms. The work introduces weights into rule antecedent attributes within these FRI procedures. The extensions have been systematically evaluated on ten benchmark classification problems, demonstrating the superior performance of these extended methods over their originals. Very importantly, as illustrated by the experimental analysis, the weighted FRI methods only require

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

TABLE VI
AVERAGE CLASSIFICATION ACCURACIES (%) VS. NUMBER OF NEAREST NEIGHBOURING RULES USED FOR DIFFERENT FRI (CONTINUED)

Dataset	FRI	n	$S_{\bar{w}}I_{\bar{w}}$	$S_{\bar{w}}I_w$	$S_wI_{\bar{w}}$	S_wI_w
Hayes-Roth	T-FRI	2	46.87	47.00	60.37	**61.00**
		3	48.75	48.75	55.75	56.37
		4	50.50	50.50	53.99	53.37
		5	52.50	52.37	55.75	55.87
		6	51.87	52.37	54.75	54.87
	KH	2	47.75	48.50	57.12	**58.00**
		3	49.75	49.62	52.37	51.62
		4	51.37	52.00	51.87	52.12
		5	51.00	51.37	51.87	52.50
		6	51.12	51.12	51.50	51.50
	CCL	2	46.37	47.12	55.25	**55.75**
		3	48.25	49.00	51.87	53.25
		4	50.62	50.37	53.00	52.24
		5	52.37	51.62	53.00	52.37
		6	50.87	51.12	51.50	51.87
Page-blocks	T-FRI	2	66.77	66.76	72.12	**72.13**
		3	60.54	59.26	61.25	60.23
		4	65.45	65.37	64.99	64.91
		5	66.81	65.66	66.57	65.49
		6	64.97	64.28	64.39	63.85
	KH	2	65.18	65.12	69.86	**69.93**
		3	57.83	57.83	58.22	58.23
		4	59.39	58.02	59.11	57.82
		5	60.30	60.14	60.11	59.93
		6	59.44	58.83	59.00	58.58
	CCL	2	66.71	66.68	72.03	72.07
		3	**74.67**	63.77	74.63	64.63
		4	71.80	66.05	72.27	65.92
		5	68.08	65.86	68.31	65.86
		6	70.02	65.06	69.12	64.71
Ecoli	T-FRI	2	59.52	61.38	64.34	**65.96**
		3	56.56	56.50	55.77	55.71
		4	52.87	52.51	52.15	51.85
		5	49.04	48.93	48.99	49.11
		6	47.38	47.85	46.06	46.24
	KH	2	59.89	59.82	63.92	**64.04**
		3	56.50	56.50	55.77	55.77
		4	53.40	53.11	52.09	51.97
		5	49.52	49.53	49.17	49.12
		6	48.44	48.50	47.07	47.08
	CCL	2	61.56	61.68	65.60	**65.90**
		3	56.50	56.50	55.71	55.77
		4	52.87	52.81	52.34	52.46
		5	49.10	49.22	49.17	49.29
		6	47.37	47.37	46.06	46.07
Red Wine Quality	T-FRI	2	52.98	55.64	54.55	**57.37**
		3	53.29	53.28	53.17	53.14
		4	53.23	54.24	53.27	54.11
		5	53.33	53.39	52.69	52.85
		6	53.68	54.19	53.10	53.42
	KH	2	52.52	52.83	53.22	53.92
		3	53.26	53.23	53.14	53.07
		4	53.43	53.84	53.48	53.78
		5	53.29	53.29	52.68	52.62
		6	53.81	**54.13**	53.29	53.28
	CCL	2	52.73	52.74	53.24	53.33
		3	52.62	53.14	52.14	52.89
		4	53.56	53.14	53.37	52.99
		5	53.84	**54.54**	53.29	53.68
		6	53.87	53.87	52.95	52.55

TABLE VII
AVERAGE CLASSIFICATION ACCURACIES (%) VS. NUMBER OF NEAREST NEIGHBOURING RULES USED FOR DIFFERENT FRI (CONTINUED)

Dataset	FRI	n	$S_{\bar{w}}I_{\bar{w}}$	$S_{\bar{w}}I_w$	$S_wI_{\bar{w}}$	S_wI_w
Wireless Indoor Localisation	T-FRI	2	76.36	78.32	77.94	**79.89**
		3	77.22	77.22	76.93	76.93
		4	75.12	75.45	74.56	74.91
		5	74.41	74.50	73.92	74.02
		6	76.47	77.00	75.92	76.44
	KH	2	77.50	78.02	78.37	78.85
		3	77.22	77.22	76.93	76.94
		4	75.50	75.89	75.03	75.25
		5	74.52	75.10	74.06	74.79
		6	79.33	**79.62**	79.06	79.06
	CCL	2	75.59	76.24	76.12	**77.18**
		3	76.82	76.95	76.53	76.89
		4	75.37	76.12	74.66	75.58
		5	76.36	77.12	75.89	76.82
		6	76.11	76.33	75.47	75.76
User Knowledge Modeling	T-FRI	2	74.85	78.91	77.03	**82.54**
		3	74.44	75.54	63.84	69.55
		4	76.18	78.91	66.17	69.15
		5	76.29	79.16	61.55	68.50
		6	77.92	79.95	61.10	66.85
	KH	2	69.63	71.51	72.68	75.24
		3	74.69	74.69	64.89	68.21
		4	74.05	74.39	65.73	69.05
		5	76.09	**76.68**	62.94	68.01
		6	75.78	76.02	62.04	66.90
	CCL	2	70.53	69.73	70.64	74.22
		3	73.64	71.91	63.84	66.02
		4	73.99	72.05	63.19	66.65
		5	74.80	74.34	60.60	66.06
		6	**75.64**	74.54	59.50	64.42

TABLE VIII
CONFUSION MATRIX OF WEIGHTED T-FRI WITH $n = 2$ AND $n = 3$ NEAREST NEIGHBOURING RULES

		Classified ($n = 2$)		Classified ($n = 3$)	
		Positive	Negative	Positive	Negative
Actual	Positive	23.12%	11.76%	21.09%	13.79%
	Negative	19.26%	45.82%	24.34%	40.75%

TABLE IX
CONFUSION MATRIX OF WEIGHTED KH WITH $n = 2$ AND $n = 3$ NEAREST NEIGHBOURING RULES

		Classified ($n = 2$)		Classified ($n = 3$)	
		Positive	Negative	Positive	Negative
Actual	Positive	20.70%	14.19%	19.99%	14.89%
	Negative	20.74%	44.34%	23.57%	41.51%

TABLE X
CONFUSION MATRIX OF WEIGHTED CCL WITH $n = 2$ AND $n = 3$ NEAREST NEIGHBOURING RULES

		Classified ($n = 2$)		Classified ($n = 3$)	
		Positive	Negative	Positive	Negative
Actual	Positive	18.14%	16.73%	18.92%	15.95%
	Negative	16.49%	48.60%	23.65%	41.44%

the least number (i.e., 2) of the nearest neighbouring rules to perform interpolation, thereby ensuring their efficacy in practical applications.

Such improved performances of the extended methods are attainable owing to the use of the relative significance degrees, or weights, of the individual rule antecedents to guide the selection of the nearest neighbouring rules for interpolation. These weights are derived from ranking attributes using the given sparse rule base only, and the weighting scheme has proven to be effective [7]. The interpolation processes are modified by the weights as well, thereby reflecting different contributions made by different attributes in deriving the interpolated consequents. This differs from the existing approaches

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2019.2928496, IEEE Transactions on Fuzzy Systems

TABLE XI
AVERAGE TESTING TIME (SEC) VS. NUMBER OF NEAREST NEIGHBOURING RULES

Dataset	Methods		Number of Closest Rules (n)				
			2	*3*	*4*	*5*	*6*
Page-blocks	T-FRI	$S_{\bar{w}}I_{\bar{w}}$	0.1881	0.1930	0.1953	0.2013	0.2049
		S_wI_w	0.1876	0.1924	0.1961	0.2045	0.2063
	KH	$S_{\bar{w}}I_{\bar{w}}$	0.1794	0.1813	0.1915	0.1940	0.1971
		S_wI_w	0.1829	0.1839	0.1910	0.1941	0.1972
	CCL	$S_{\bar{w}}I_{\bar{w}}$	0.1789	0.1784	0.1833	0.1881	0.1887
		S_wI_w	0.1814	0.1813	0.1854	0.1922	0.1925
Ecoli	T-FRI	$S_{\bar{w}}I_{\bar{w}}$	0.0165	0.0203	0.0192	0.0207	0.0206
		S_wI_w	0.0162	0.0199	0.0192	0.0217	0.0207
	KH	$S_{\bar{w}}I_{\bar{w}}$	0.0166	0.0200	0.0175	0.0201	0.0198
		S_wI_w	0.0164	0.0197	0.0177	0.0199	0.0184
	CCL	$S_{\bar{w}}I_{\bar{w}}$	0.0181	0.0182	0.0188	0.0209	0.0199
		S_wI_w	0.0184	0.0198	0.0184	0.0213	0.0201
Red Wine Quality	T-FRI	$S_{\bar{w}}I_{\bar{w}}$	0.1673	0.1735	0.1733	0.1782	0.1819
		S_wI_w	0.1649	0.1722	0.1729	0.1773	0.1799
	KH	$S_{\bar{w}}I_{\bar{w}}$	0.1687	0.1695	0.1736	0.1754	0.1784
		S_wI_w	0.1692	0.1713	0.1753	0.1747	0.1777
	CCL	$S_{\bar{w}}I_{\bar{w}}$	0.1609	0.1612	0.1632	0.1647	0.1684
		S_wI_w	0.1625	0.1625	0.1634	0.1685	0.1688
Wireless Indoor Localization	T-FRI	$S_{\bar{w}}I_{\bar{w}}$	0.1594	0.1704	0.1694	0.1775	0.1735
		S_wI_w	0.1592	0.1723	0.1709	0.1788	0.1742
	KH	$S_{\bar{w}}I_{\bar{w}}$	0.1695	0.1692	0.1706	0.1755	0.1741
		S_wI_w	0.1660	0.1695	0.1682	0.1762	0.1723
	CCL	$S_{\bar{w}}I_{\bar{w}}$	0.1639	0.1629	0.1610	0.1655	0.1668
		S_wI_w	0.1643	0.1640	0.1639	0.1699	0.1691
User Knowledge Modeling	T-FRI	$S_{\bar{w}}I_{\bar{w}}$	0.0265	0.0351	0.0300	0.0323	0.0309
		S_wI_w	0.0264	0.0346	0.0301	0.0338	0.0313
	KH	$S_{\bar{w}}I_{\bar{w}}$	0.0268	0.0305	0.0294	0.0322	0.0301
		S_wI_w	0.0268	0.0304	0.0292	0.0319	0.0302
	CCL	$S_{\bar{w}}I_{\bar{w}}$	0.0256	0.0312	0.0284	0.0306	0.0289
		S_wI_w	0.0256	0.0317	0.0281	0.0309	0.0292

where all attributes are treated equally.

The conjecture that "least number of neighbouring rules do better with weighted FRI" has been demonstrated with substantial and consistent empirical results. It would be beneficial to further verify this notion through theoretical analysis also. This forms an important next step to reinforce the current research. Also, the experimental investigations carried out have not used any sophisticated representation means for fuzzy values, nor any complicated fuzzy rule generation mechanism. However, any optimisation of these aspects will help further improve the classification accuracy. This is omitted herein, as what has been concerned with is the relative performance given common settings for the running of both the extended and their original methods. Naturally, it would be interesting to examine how the extensions may perform with optimised fuzzy quantities and rules. In addition, it would be useful to apply the resultant weighted interpolative inference mechanisms to real-world problems, such as medical risk analysis [14] and serious crime investigation [15] where only sparse knowledge is available.

REFERENCES

[1] L.T. Kóczy, K. Hirota. "Approximate reasoning by linear rule interpolation and general approximation." *International Journal of Approximate Reasoning* **9**.3, 197-225 (1993).

[2] K. Uehara, K. Hirota. "Inference with Fuzzy Rule Interpolation at an Infinite Number of Activating Points." *Journal of Advanced Computational Intelligence and Intelligent Informatics* **19**.1, 74-90 (2015).

[3] Z. Huang, Q. Shen. "Fuzzy interpolative reasoning via scale and move transformations." *IEEE Transactions on Fuzzy Systems* **14**.2, 40-359 (2006).

[4] Y.C. Chang, S.M. Chen, C.J. Liau. "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets." *IEEE Transactions on Fuzzy Systems* **16**.5, 1285-1301 (2008).

[5] R. Diao, S. Jin, Q. Shen. "Antecedent selection in fuzzy rule interpolation using feature selection techniques." *Proceedings of the 23rd International Conference on Fuzzy Systems*, 2014.

[6] S.M. Chen, S.I. Adam. "Weighted fuzzy interpolated reasoning based on ranking values of polygonal fuzzy sets and new scale and move transformation techniques." *Information Sciences* **435**, 184-202 (2018).

[7] F. Li, C. Shang, Y. Li, J. Yang, and Q. Shen. "Fuzzy Rule-Based Interpolative Reasoning Supported by Attribute Ranking." *IEEE Transactions on Fuzzy Systems* **26**.5, 2758-2773 (2018).

[8] D. Tikk, I. Joó, L.T. Kóczy, P. Várlaki, B. Moser, T.D. Gedeon. "Stability of interpolative fuzzy KH controllers." *Fuzzy Sets and Systems* **125**.1, 105-119 (2002).

[9] K.W. Wong, D. Tikk, T.D. Gedeon, L.T. Kóczy. "Fuzzy rule interpolation for multidimensional input spaces with applications: A case study." *IEEE Transactions on Fuzzy Systems* **13**.6, 809-819 (2005).

[10] J. AlcaláFdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera. "KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework." *Journal of Multiple-Valued Logic and Soft Computing* **17**:2-3, 255-287 (2011). Available: http://sci2s.ugr.es/keel/datasets.php

[11] C.L. Blake, C.J. Merz. (1998) UCI Repository of machine learning databases, University of California, Irvine. Available: https://archive.ics.uci.edu/ml/datasets.html

[12] L. Wang, J.M. Mendel. "Generating fuzzy rules by learning from examples." *IEEE Transactions on Systems, Man, and Cybernetics* **22**.6, 1414-1427 (1992).

[13] Z. Huang, Q. Shen. "Fuzzy interpolation and extrapolation: A practical approach." *IEEE Transactions on Fuzzy Systems* **16**.1, 13-28 (2008).

[14] F. Li, C. Shang, Y. Li and Q. Shen. "Feature ranking-guided fuzzy rule interpolation for mammographic mass shape classification". *Proceedings of the 27th International Conference on Fuzzy Systems*, pp. 1039-1045, 2018.

[15] T. Boongoen, Q. Shen and C. Price. "Disclosing false identity through hybrid link analysis". *Artificial Intelligence and Law*, **18**.1, 77-102, 2010.