Rule Base Structure and Influence of Existing Rules

Jiang, Changhong; Shang, Changjing; Shen, Qiang

Publication date:
2021

Citation for published version (APA):
Jiang, C., Shang, C., & Shen, Q. (2021). *Rule Base Structure and Influence of Existing Rules: An Initial Investigation*.

Rule Base Structure and Influence of Existing Rules: An Initial Investigation

Changhong Jiang, Changjing Shang and Qiang Shen

Department of Computer Science, Aberystwyth University, UK
{chj40, cns, qqs}@aber.ac.uk

Abstract. Fuzzy rule interpolation (FRI) provides an interpretable decision while making inference with a sparse rule base. The common FRI theories postulate that more existing rules involved can derive better reasoning outcomes. However, certain empirical results show that a large number of rules involved may adversely lead to worsening outcomes. The objective of this work is to establish a mathematical mapping of the structural pattern of a given fuzzy rule base for use in FRI, such that the original rule base can be effectively described and analysed. The resulting mathematically mapped pattern helps to produce a theorem which determines the upper limit of the number of rules required to perform FRI. The experimental investigations carried out so far demonstrate that the number of required rules, which can result in the best interpolative outcomes, obeys the theorem proposed by this work.

Keywords: Fuzzy Rule Interpolation (FRI), Rule Base Structure, Mapping, Upper Limit of Rule Number.

1 Introduction

Fuzzy reasoning was first introduced by Zadeh [1] in 1965. Then, in 1973, an innovative approximate reasoning method, termed Compositional Rule of Inference (CRI) was proposed by Zadeh [2], which has been used wildly since and which is still the most popular principled way to conduct approximate inference. In 1975, Mamdani and Assilian [3] applied CRI to the control of boilers and steam engines for the first time and obtained impressive results. Following that, Tsukamoto [4] improved the approach and presented a mathematical formulation for reasoning with rules involving multidimensional antecedents in 1979. In order to generalise Fuzzy Rule-Based Systems (FBRS) capable of addressing multidimensional problems, Sugeno and Takagi [5] proposed a new tool via linear functional regression between antecedents and consequents in 1983, and subsequently extended it in 1985 [6].

In 1993, Koczy and Hirota [7] [8] proposed a linear fuzzy rule interpolation inference mechanism to solve the problem of fuzzy inference under the condition of having a sparse fuzzy rule base. By introducing the use of Euclidean distance measures to describe the similarity degree of the fuzzy sets, when the inference premise falls on the gap between two rule antecedents in the rule base, a fuzzy conclusion can be

computed using the linear interpolation principle. The notion of a sparse fuzzy rule base means that a given observation may not match any existing rule in the given rule base, that is, it is an incomplete knowledge base without covering the full problem domain concerned. Koczy and Hirota's method (KH) plays a very important role in performing approximate inference for such situations. Since then, many significant developments have been made, presenting a diverse range of potentially effective fuzzy rule interpolation (FRI) tools.

Particularly, Huang and Shen [9] proposed an FRI method (HS) based on the exploitation of the centre of gravity of fuzzy sets, resolving the convexity problem that previous techniques fail to deal with (to ensure that convex inputs lead to convex outputs). In order to address interpolation with fuzzy rules that involve complex polygon, Gaussian or other bell-shaped fuzzy membership functions, they also proposed a method via scale and move transformations [10], which was then refined for practical use in 2008 [11]. This approach works by first creating a new rule from two nearest neighbouring rules, and then using scale and move transformations to derive the conclusion. However, these and indeed almost all common FRI methods suppose that the antecedent attributes of the rules are equally significant in the implementation of interpolation, which can lead to inaccurate or even incorrect interpolative results.

By combining feature selection techniques [12]-[14] a range of FRI methods have been put forward [15]-[17] to reflect the different levels of significance in different domain attributes or rules. For instance, Li et al. [18] proposed a feature ranking-guided FRI mechanism using the HS method as its basis. This work has been generalised through enhancing two alternative representative FRI methods also. 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 weighted FRI algorithms work effectively and efficiently, using just two nearest neighbouring rules [19].

However, the aforementioned approaches to FRI work by always focusing on the existing rules individually. There does not exist any reasonable theory to describe and analyse the underlying FRI techniques, from the perspective of rule space structure and the distribution of the rules in the given rule base. Besides, there is an interesting observation, on the contradiction between the theories and experimental results, that involving more rules for interpolation may make the interpolated outcomes worse. Therefore, it is desirable to develop an appropriate mathematical model to describe the process of FRI, and to analyse the existing approaches in a mathematically rigorous and systematic manner. This paper serves an initial attempt to establish such a generic theoretical FRI approach.

The remainder of this paper is organised as follows. Section II reviews the relevant background of FRI in general and the transformation-based FRI (T-FRI) in particular, upon which to develop the proposed work. Section III describes the mathematical representation from the viewpoint of rule space structure. Section IV shows the results of experimental evaluation. Section V concludes the paper and points out issues for further research.

2 Background

For academic completeness, an outline of the key concepts that form the foundations for the present development is given here, covering the generic task of any FRI method and a specification of the Transformation-Based Fuzzy Rule Interpolation (T-FRI) mechanism.

2.1 Generic Task of FRI

Fuzzy Rule-Based Systems (FBRS) can each be essentially represented by two key elements：

- A nonempty finite set of domain attributes $D = A \cup C$, where $A = \{A_j | j = 1,2,\dots,m\}$ represents the set of antecedent attributes and $C = \{C_j | j = 1,2,\dots,m\}$ stands for the consequent attribute.
- A nonempty set of finite fuzzy rules $R = \bigcup_{i=1}^{n}\{r^i: A_i \rightarrow C_i\}$, where n is the number of the fuzzy rules.

In a conventional FRI method, a given rule $r^i \in R$ can be described as follows:

r^i: If a_1 is A_i^1 and a_2 is A_i^2 and … and a_m is A_i^m
Then c is C_i

where A_i^j represents the value of the antecedent attribute a_j in the rule r^i, and C_i denotes the value of the consequent attribute c in r^i.

FRI is a method for computing the consequent given a novel observation that does not match the antecedent of any rule within the rule base. This is generally implemented through linear interpolative manipulation of certain rules in the rule base, guided by the observation. In other words, if there is an observation o^* which can be generally expressed as:

o^*: a_1 is A_*^1 and a_2 is A_*^2 and … and a_m is A_*^m

and which does not match any rule $r^i \in R$, the aim of FRI is to construct a suitable consequent C_* by linearly interpolating the consequents $C_i \in C$ of a certain subset of the existing rules $r^i \in R$ whose antecedents are closest to the observation, in an effort to derive the consequence of the given observation o^*.

2.2 Transformation-Based FRI

T-FRI [10] is arguably the most advanced and popular TFR method [20] that implements the generic task of FRI by four steps:

1. Selection of Closest Rules

This procedure searches for a certain number of rules that are the closest to the given (unmatched) observation. The distance between any two rules $r^p, r^q \in R$ (or observa-

tion o^*) is determined by computing the aggregated distances between every corresponding values of the shared attributes between them.

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

$$d\left(A_p^j, A_q^j\right) = \frac{|Rep\left(A_p^j\right) - Rep(A_q^j)|}{max_{A_j} - min_{A_j}} \quad (2)$$

where max_{A_j} and min_{A_j} denote the maximal and minimal value of the attribute a_j respectively, and the notation $Rep(A)$ is termed the representative value of the fuzzy set A. $Rep(A)$ captures the key information reflected by the overall location of A in its domain and the geometric shape of A. Normally, for an arbitrary polygonal fuzzy set $A = (a_1, a_2, \dots, a_n)$, the representative value $Rep(A)$ is defined as

$$Rep(A) = \sum_{i=1}^{n} \omega_i \, a_i \quad (3)$$

where ω_i denotes the weight assigned to the vertex a_i per i. Therefore, the n closest rules to o^* can be identified as those rules leading to the n smallest values of this distance measure between the representative value of the observation and that of the antecedent of a certain rule taken from the rule base.

2. Construction of Intermediate Fuzzy Rule

From the preceding procedure, n closest rules to a given observation o^* can be chosen which have the shorter distances than the rest of all the rules to o^*. From this, an intermediate fuzzy rule r' can be constructed as the beginning of the transformation process.

The constructed intermediate rule is described as follows:

r': If a_1 is A_r^1 and a_2 is A_r^2 and … and a_m is A_r^m

Then c is C_r

which is a weighted aggregation of the n closest rules. Normally, the weight for every attribute is set as:

$$\omega_i^j = \frac{1}{1 + d(A_i^j, A_*^j)} \quad (4)$$

3. Computation of Scale and Move Factors

The goal of this step is to obtain the scale rate s_{A_j} and move ratio m_{A_j}, so that the transformed shape and representative value of the intermediate fuzzy set A_r^j coincide with those of the observed value A_*^j.

4. Scale and Move Transformation

After calculating the necessary scale and move factors, the final step is using them (through linear mapping) to derive the required consequent of C_*

Note that in this initial investigation of producing a mathematically rigorous model for T-FRI, only Steps 1 and 2 as described above are of direct relevance (and hence, Steps 3 and 4 are just roughly described). More technical details regarding T-FRI can be checked in [10].

3 Rule Space Structure in T-FRI

During this process of performing interpolative reasoning, all involved rules $r^i \in R$ are decided by Step 1 shown in Section 2.2, which are then treated individually. As each existing rule includes certain information in support of the construction of the intermediate rule, it seems that the more rules involved, the better result may be obtained. In the present work, such hypothesis and observation are investigated theoretically. However, as experimental results already shown in [18][19], this is often not necessarily the case.

3.1 Rule Space Mapping

To analyse the structure of the rule space R, mapping R onto an isomorphic data space R^D offers a natural approach. Motivated with this recognition of the potential of having such a mapping, let each rule r^i or observation o^* be mapped onto a hyper point with m dimensions, i.e.,

$$r^i\{a_1, a_2, \dots, a_m\} \in R \rightarrow r_{A_i}(A_i^1, A_i^2, \dots, A_i^m) \in R^D$$
$$o^*\{a_1, a_2, \dots, a_m\} \rightarrow o_{A_*}(A_*^1, A_*^2, \dots, A_*^m)$$

where a_j represents the antecedent attribute in the rule r^i and A_i^j represents the value of a_j. Then, the mapping $r^i \rightarrow r^{iD}$ and $o^* \rightarrow o^{*\mathrm{D}}$ can be shown correspondingly, as

r^{iD}: If $r^i\{a_1, a_2, \dots, a_m\} \rightarrow r_{A_i}(A_i^1, A_i^2, \dots, A_i^m)$
Then c is C_i
$o^{*\mathrm{D}}$: $o^*\{a_1, a_2, \dots, a_m\} \rightarrow o_{A_*}(A_*^1, A_*^2, \dots, A_*^m)$

Obviously, the two mappings are both bijections. Thus, the data space $R^D = \{r_{A_i} | \, i = 1,2, \dots, n\}$ is an isomorphic space of the original rule space R.Consider a pair of rules $r^p, r^q \in R$ ($r^{pD}, r^{qD} \in R^D$),

$$d(r^{pD}, r^{qD}) = \gamma d\left(r_{A_p}, r_{A_q}\right) = \gamma \sqrt{\sum_{j=1}^{m} \left|Rep\left(A_p^j\right) - Rep\left(A_q^j\right)\right|^2}$$
$$= \gamma \sqrt{\sum_{j=1}^{m} d(A_p^j, A_q^j)^2} = \gamma \, d(r^p, r^q) \quad (5)$$

where γ is a constant defined by m, max_{A_j} and min_{A_j}, according to Eqn. (1) and Eqn. (2), such that

$$\gamma = \frac{1}{\sqrt{m}(max_{A_j} - min_{A_j})} \quad (6)$$

Therefore, the distance measure between the two representations of an individual rule in the two spaces remains the same. This means that the n rules to be selected for use in T-FRI within Step 1 are to be the same in both spaces.

3.2 Rules with One Antecedent Attribute

Consider the simplest circumstance first, where every hyper point r_{A_i} is only of one dimension. In this case, all points $r_{A_i} \in R^D$ are distributed in a one-dimensional axis and can be represented as:

$$r_{A_i}(A_i)$$

For simplicity, the hyper points $r_{A_i} \in R^D$ are also interchangeably called rules below, because the data space R^D is an isomorphic space of the original rule space R as shown previously.

As the intermediate rule is a weighted aggregation of the n closest rules (as per Step 2 in Section 2.2), the intermediate rule $r_{A,}(A,)$ can be denoted by

$$r_{A,}(A,) = \sum_{i=1}^{n} \omega_i r_{A_i}(A_i) \tag{7}$$

where ω_i is the weight of the attribute and $0 \leq \omega_i \leq 1$. It is straightforward to envision that the intermediate rule $r_{A,}$ always lies between the maximal selected rule $\max r_{A_i}$ (i.e., the selected one that is farthest to the observation) and the minimal selected rule $\min r_{A_i}$ (i.e., the one nearest to the observation). Let

$$r_{A,,} = \sum_{i=1}^{n+1} \omega_i r_{A_i} \tag{8}$$

$$\begin{aligned} d\left(r_{A_{n+1}}, r_{A,,}\right) &= A_{n+1} - \sum_{i=1}^{n+1} \omega_i A_i = A_{n+1} - \sum_{i=1}^{n} \omega_i A_i - \omega_{n+1} A_{n+1} \\ &= d\left(r_{A_{n+1}}, r_{A,}\right) - \omega_{n+1} A_{n+1} < d\left(r_{A_{n+1}}, r_{A,}\right) \end{aligned} \tag{9}$$

Therefore, every selected rule r_{A_i} is to pull the intermediate rule $r_{A,}$ towards itself. Thus, there are just two situations to address:

- The distribution of hyper points is uniform.
- The distribution of hyper points is uneven. In other words, there exist certain clusters of such points.

These two different circumstances are discussed separately below.

For Uniform Space

For a uniformly distributed hyper data space R^D, the distances between every pair of neighbouring rules are approximately same. Formally, consider such a rule base together with the observation o^*, the relationship of the n closest rules r_{A_i}, the intermediate rule $r_{A,}$ and the observation o_{A_*} can now be shown as:

$$\min r_{A_i} < \cdots < r_{A,} < \cdots < o_{A_*} < \cdots < \max r_{A_i}$$

or

$$\min r_{A_i} < \cdots < o_{A_*} < \cdots < r_{A,} < \cdots < \max r_{A_i}$$

Thus, if the distribution of the rule base is uniform, two closest neighbouring rules are sufficient to compute an appropriate intermediate rule. From this, it can be readily derived that if there are the same number of rules chosen from both sides of the emerging intermediate rule $r_{A,}$, the result will be approximately the same to the result that only two rules are chosen. This means that in this situation, the rules (except the

two closest neighbouring rules) have limited contribution to the construction of the intermediate rule and hence, to the computation of the eventual interpolated result.

An alternative derivation to the above can be obtained when the rule base is uniformly distributed. That is, if more neighbouring rules are chosen from one side of the observation than from the other, the intermediate rule will deviate from the neutrally appropriate position, biasing towards the side where more rules are selected.

For Uneven Space

For an uneven data space R^D, where the distances between every pair of neighbouring rules in the same cluster are much closer than those between pairs of rules taken from different clusters. If the observation o^* falls into a cluster, the relationship of the n closest rules r_{A_i}, the (emerging) intermediate rule $r_{A,}$ and the observation o_{A_*} can then be dealt with in the same way as above, artificially treating that particular rule cluster as the given rule base.

If however, o^* dose not fall into any cluster, the relationship can be shown as:

$$\min r_{A_i} < \cdots < r_{A,} < \cdots < \max r_{A_i} < o_{A_*}$$

or

$$o_{A_*} < \min r_{A_i} < \cdots < r_{A,} < \cdots < \max r_{A_i}$$

Thus, if the distribution of hyper points is uneven, two nearest neighbouring rules are also sufficient to produce an appropriate intermediate rule (as extrapolation rather than interpolation, of course). If there are too many rules having been chosen, one from each of the many nearest clusters, the intermediate rule $r_{A,}$ may be too far away from $\max r_{A_i}$ (or $\min r_{A_i}$), causing it may be too far away from the observation $o_{A_*}$to implement reasonable extrapolation.

As a summary, for models involving rules that only have one antecedent attribute, given the hyper data space $R^D = \{r_{A_i}(A_i) | \ i = 1,2, \dots, n\}$, selection of two closest neighbouring rules is sufficient to produce an appropriate intermediate rule. Employment of more rules may have limited contribution in the construction of the intermediate rule, or even cause the quality of the intermediate rule to decay.

3.3 Rules with n Antecedent Attributes

Consider a specific case under this situation to start with: where all hyper points $r_{A_i} \in R^D$ have two dimensions. Thus, the hyper points can be denoted as:

$$r_{A_i}(A_i^1, A_i^2)$$

Further, consider the most special circumstance, where every hyper point r_{A_i} is only concerned with one dimension, such that

$$r_{A_i} \in \{r_{A_i} | (A_i^1, 0) \ or \ (0, A_i^2)\}$$

Under such assumptions, based on the summary presented in the preceding section, to create the intermediate rule $r_{A,}(X, Y)$, only two closet rules $r_{A_\alpha}(A_\alpha^1, 0)$ and $r_{A_\beta}(A_\beta^1, 0)$ are necessary in order to construct the dimension A_r^1 of the intermediate rule $r_{A,}$ thereby obtaining $r_{A_r^1}(A_r^1, 0)$. Then, there should be one more closet rule $r_{A_\gamma}(0, A_\gamma^2)$ together with $r_{A_r^1}(A_r^1, 0)$ to construct the other dimension A_r^2 of $r_{A,}$, obtaining $r_{A,}(A_r^1, A_r^2)$.

Generalising the above, as illustrated in Fig. 1, it can be derived that three hyper points are sufficient to construct an appropriate intermediate rule with rules whose antecedent part is of two dimensions. Thus, expanding on this generalization further by induction, the following theorem holds:

Theorem 1: For hyper points $r_{A_i} \in R^D$ that have n dimensions, no more than $(n+1)$ hyper points are required to construct an appropriate intermediate rule.

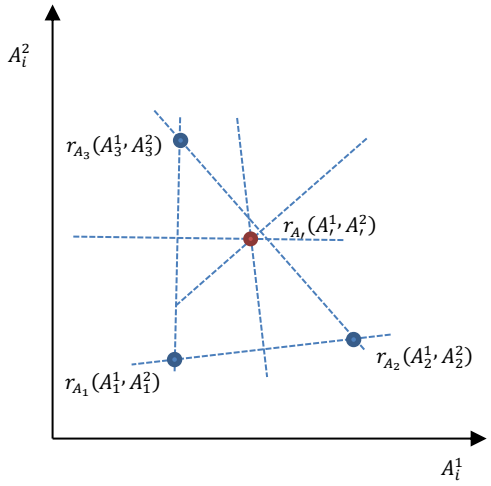

Fig. 1. Creation of an intermediate rule in a two-dimensional space.

4 Experimental Evaluation

This section presents a systematic experimental evaluation of the feature ranking-guided T-FRI approach over a benchmark dataset, for the task of performing pattern classification, using rules in the mapped hyper point structure. Note that as weighted T-FRI assigns a weight to each antecedent attribute of the original (unweighted) rules, it makes the rule space structure different from the one otherwise resulting from the mapping of the original rule base.

4.1 Experimental Setup

The benchmark dataset is taken from Knowledge Extraction based on Evolutionary Learning (KEEL) dataset repository [21], whose details are summarised in Table I.

Table 1. Dataset Used for Experimental Investigation

Dataset	Attributes #	Classes #	Instances #
Phoneme	5	2	5404

Triangular membership functions are used to represent the fuzzy sets of the antecedent variables due to their popularity and simplicity. As the task at hand is to perform classification, the consequent variable always adopts a singleton fuzzy set (i.e., a crisp value) for its value. Generally, different variables have their own underlying domains. For simplicity, these domains are normalised to take a value from the common range of 0 to 1 (without optimisation), as illustrated in Fig. 2.

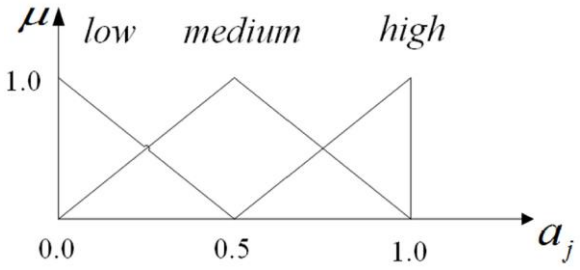

Fig. 2. Membership functions defining the linguistic terms.

Experiments are validated by 10-fold cross validation, repeated for 10 times. The feature scores for attribute weighting are derived from the five methods that are utilised in the existing literature [19], where the rules are learned from data after fuzzification [22]. Note that 20% of the learned rules are deliberately removed randomly, in order to make the resultant rule base sparse whilst having a ground-truth rule base that covers the entire problem domain to facilitate performance comparison. The classification performance is assessed in terms of the average accuracy calculated over the process of 10x10-fold cross validation.

4.2 Results and Discussion

Table 2 shows the average classification accuracies with different numbers of closest rules involved, where the method Ori stands for the conventional unweighted T-FRI algorithm and the other five represent those weighted T-FRI methods with each implemented using a different feature selection tool. The position of every highlighted figure indicates that by the use of what number (n) of the closest neighbouring rules the best result is reached. From these results, it can be seen that increasing the number of the closest rules does not improve the performance significantly for any of the five weighted T-FRI methods. In fact, the performance even deteriorates as n increases. In particular, according to Theorem 1, for the present problem, there are five antecedent attributes, thus six closest rules are all that is required to perform the interpolation, at most. Obviously, for weighted methods other than LS, all the results clearly obey Theorem 1.

Table 2. Average Accuracy (%) vs. Number of Closest Rules Used for Interpolation

Dataset	Method	Number of Closest Rules (n)				
		2	3	4	5	6
Phoneme	Ori	57.10	54.16	57.54	58.91	**59.19**
	IG	**67.33**	64.93	63.45	64.63	65.08
	Relief-F	**64.78**	62.89	62.91	63.71	63.82
	LLCFS	**64.59**	61.56	60.99	60.65	61.02
	LS	60.47	61.28	60.28	61.47	**62.43**
	RSFS	61.67	61.34	61.76	**61.82**	60.39

To have a closer examination of the results, Table 3 lists the individual weights that are learned by the different feature selection methods. Note that one or two antecedent attributes may be given a weight of 0 when using certain methods, meaning that the number of informative attributes of a rule after feature-weighting may be less than that of its (unweighted) original. From this viewpoint, looking at Table 3 again, the results of the original T-FRI and four weighted T-FRI methods still obey Theorem 1, but the one with LS used for feature weighting does not. This is summarised in Table 4.

Table 3. Attribute Weights Using Different Ranking Schemes

Method	Antecedent Weights				
IG	0.2852	0.0792	0.0125	0.5724	0.0507
Relief-F	0.1326	0.0414	0	0.7286	0.0973
LLCFS	0.0001	0	0	0.7416	0.2583
LS	0	0.4541	0.0988	0.1995	0.2476
RSFS	0.0016	0.0016	0.0016	0.9938	0.0016

A question may therefore, arise for the case where attribute weights are computed using LS; that is, why such a weighted method requires one more closest rule to perform interpolation (by ignoring the fact that the theorem is ultimately correct when considering the number of the original antecedent attributes involved). This is likely caused by a further approximation incurred during the process of computing the attribute weights. Unlike the other feature selection methods employed, where the symbolic linguistic terms representing the attribute values are directly used in feature evaluation and ranking. However, for LS, it is the representative value $Rep(A)$ of a fuzzy A that is used. As such, this method inevitably leads to a certain amount of information loss, producing a zero weight for the first attribute incorrectly. Nonetheless, as stated earlier, compared to the real number of the original rule antecedent attributes, Theorem 1 stands.

Table 4. Number of Closest Rules Required vs. Theoretical Upper Limit

Method	Informative Attribute #	Theoretical Upper Limit #	# Leading to Best Result	Original Attribute #
Ori	5	6	6	5
IG	5	6	2	5
Relief-F	4	5	2	5
LLCFS	3	4	2	5
LS	4	5	6	5
RSFS	5	6	5	5

5 Conclusion

This paper has presented a completely new view to describe FRI and T-FRI. As the very first to such a structural approach to theoretically verifying fuzzy rule interpolation, the paper has analysed the relationship between the n closest rules r_{A_i}, the intermediate rule r_{A}, and the (unmatched) observation o_{A_*}. The work has illustrated how these key concepts within an FRI algorithm may influence one another. It shows that the maximum number of the closest rules required is one more than that of the rule antecedent attributes. This offers a mathematically sound conclusion that helps explain the empirical observation that the use of more than two neighbouring rules does not necessarily lead to more accurate interpolated outcomes but instead, often produces less accurate results.

In this initial investigation, only one dataset is used for the present experimental verification, more systematic experimental studies would help to reinforce the promising results achieved. Further investigations also include: i) devising a potentially more efficient T-FRI method that is to be guided by the conclusion drawn above; ii) extending Theorem 1 from the static rule base to the dynamic rule conditions since many practical applications require a dynamic rule base to deal with the changing environment; iii) introducing new evaluation functions to assess the contribution of each attribute to the eventual interpolation outcome, so that the influence of every individual rule upon the consequent can be better identified and therefrom, the decision on which neighbouring rules to use can be more appropriately made.

References

1. L. A. Zadeh. Fuzzy sets. Information Control, vol. 8, no. 1, 338-353, 1965.
2. L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Fuzzy Systems, Man, and Cybernetics, vol. SMC-3, no. 1, 28-44, 1973.

3. E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, vol. 7, 1-13, 1975.
4. Y. Tsukamoto. An approach to fuzzy reasoning method. Advances in Fuzzy Set Theory and Applications, 137-149, 1979.
5. M. Sugeno and T. Takagi, Multi-dimensional fuzzy reasoning. Fuzzy Sets Systems, vol. 9, 313-325, 1983.
6. T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Fuzzy Systems, man, and cybernetics, vol. SMC-15, no. 1, 116-132, 1985.
7. L. Kóczy and K. Hirota. Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. Information Sciences, vol. 71, 169-201, 1993.
8. L. Kóczy and K. Hirota. Approximate reasoning by linear rule interpolation and general approximation. International Journal of Approximate Reasoning, vol. 9, no. 3, 197-225, 1993.
9. Z. Huang and Q. Shen. A new fuzzy interpolative reasoning method based on center of gravity. The 12th IEEE International Conference on Fuzzy Systems vol. 1, 25-30, 2003.
10. Z. Huang and Q. Shen. Fuzzy interpolative reasoning via scale and move transformations. IEEE Transactions on Fuzzy Systems, vol. 14, no. 2, 340-359, 2006.
11. Z. Huang and Q. Shen. Fuzzy interpolation and extrapolation: a practical approach. IEEE Transactions on Fuzzy Systems, vol. 16, no. 1, 13-28, 2008.
12. M. Dash and H. Liu. Feature selection for classification. Intelligent data analysis, vol. 1, 131-156, 1997.
13. R. Jensen and Q. Shen. New approaches to fuzzy-rough feature selection. IEEE Transactions on Fuzzy Systems, vol. 17, no. 4, 824-838, 2009.
14. Y. Cui, J. Jin, S. Zhang, S. Luo and Q. Tian. Correlation-based feature selection and regression. Pacific-Rim Conference on Multimedia, Lecture Notes in Computer Science, vol. 6297, 25-35, 2010.
15. Y. Li, D. Huang, Tsang and L. Zhang. Weighted fuzzy interpolative reasoning method. Machine Learning and Cybernetics. Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol.5, 3104-3108, 2005.
16. S. Chen and Y. Chang. Weighted fuzzy rule interpolation based on GA-based weight-learning techniques. IEEE Transactions on Fuzzy Systems, vol. 19, no. 4, 729-744, 2011.
17. R. Diao, S. Jin and Q. Shen. Antecedent selection in fuzzy rule interpolation using feature selection techniques. 2014 IEEE International Conference on Fuzzy Systems, 2206-2213, 2014.
18. F. Li, Y. Li, C. Shang and Q. Shen. Fuzzy knowledge-based prediction through weighted rule interpolation. IEEE Transactions on Cybernetics, vol. 50, no. 10, 4508-4517, 2020.
19. F. Li, C. Shang, Y. Li, J. Yang and Q. Shen. Interpolation with Just Two Nearest Neighbouring Weighted Fuzzy Rules. IEEE Transactions on Fuzzy Systems, vol. 28, no. 9, 2255-2262, 2020.
20. F. Li, C. Shang, Y. Li, J. Yang and Q. Shen. Approximate reasoning with fuzzy rule interpolation: Background and recent advances. Artificial Intelligence Review, DOI: 10.1007/s10462-021-10005-3, 2021.
21. J. Alcal-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garca, L. Snchez and 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.
22. L. Wang and J. Mendel. Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, vol. 22, no. 6, 1414-1427, 1992.