Rough-Fuzzy Rule Interpolation for Data-Driven Decision Making

Chen, Chengyuan; Shen, Qiang

Published in:
Proceedings of the 20th Workshop on Computational Intelligence

Publication date:
2021

Citation for published version (APA):
Chen, C., & Shen, Q. (in press). Rough-Fuzzy Rule Interpolation for Data-Driven Decision Making. In *Proceedings of the 20th Workshop on Computational Intelligence*

Rough-Fuzzy Rule Interpolation for Data-Driven Decision Making

Chengyuan Chen[1] and Qiang Shen[2]

[1] School of Intelligent Technology and Engineering
Chongqing University of Science and Technology, Chongqing 401331, China
ccy@cqust.edu.cn
[2] Department of Computer Science
Faculty of Business and Physical Sciences
Aberystwyth University, Aberystwyth SY23 3DB, UK
qqs@aber.ac.uk

Abstract. Any practical decision making strategy is required to ensure that the best decision is made with respect to the information available and the knowledge possessed by experts. A rule-based fuzzy decision making system typically works on the fuzzy rules generated from numerical data acquired in the problem domain. However, different expert opinions on fuzzy partitions may result in a range of uncertainties in representing the domain knowledge. The invention of rough-fuzzy sets offers a great potential in the representation, handling and utilisation of different levels of uncertainty in knowledge. Inspired by this observation, a rough-fuzzy rule interpolation method is introduced in this paper to enable decision making systems modelling and harnessing additional uncertain information, in order to implement a fuzzy reasoning system that can work with incomplete rule base. An initial experimental investigation is carried out and the results are presented to demonstrate the effectiveness of the proposed method in aiding the development of an intelligent decision making system.

Keywords: Decision making, rough-fuzzy sets, fuzzy rule interpolation

1 Introduction

Decision making is one of the most important activities for real-world applications of intelligent systems [9]. With given domain knowledge, the task of decision making is to obtain an optimal, or at least a near optimal solution, from input information using an inference procedure. That is, the subject of decision making is the study of how decisions can be made and how they may be made most effectively. Expert systems have been widely used in decision making [14], and fuzzy set theory is frequently used in such expert systems, because of their simplicity for implementation and their similarity to human reasoning. A fuzzy rule-based expert system contains fuzzy rules in its knowledge base upon which to derive conclusions from given inputs via a fuzzy reasoning process.

A core of producing a fuzzy system from data available in the problem domain is the generation of fuzzy rules, often referred to as data-drive learning. Such a learning procedure from numerical data consists of two key phases: the partition of the input spaces into fuzzy subspaces and the determination of the shapes of membership functions. For a fuzzy partition, an element does not need to be associated with a single region, but has a set of membership functions that indicate the extent to which it is regarded as belonging to each of the regions [5]. Usually, the partition can be generated from the advice of human experts. Experts can define a number of fuzzy sets for each variable, which are interpreted as linguistic variables and shared for use in all of the rules to learn [6]. However, this type of approach relies heavily on the opinions of experts, who must have a comprehensive and detailed understanding of the problem at hand. Such opinions are often subjective and/or inconsistent between different individuals. In this case, individuals may have a different understanding for the same information and different experiences in the area of a current problem, so that different sorts of expertise may be obtained from different experts.

In addressing real problems, the situation can become even more complicated. The same rule in terms of superficial or linguistic representation may be derived from the partitions that are provided by all experts, yet this rule may have different underlying interpretations for the linguistic terms used by different experts. Such inconsistency reveals the uncertainty involved in the decision process. Words can mean different things to different people, so that a concept may have an uncertain profile for human opinions. In addition, when a phenomenon or an event is too complex or too ill to be expressed, experts would be forced to make unclear judgements. Consequently, the decision process is usually accompanied by imprecision and uncertainty that characterise expert judgements or opinions.

As different partitions of the same set of elements are usually provided, it is relevant to consider obtaining a single consensus partition which summarises the relevant (and imprecisely described) uncertain information contained within the separate partitions. Such a consensus partition provides a way of simplifying this information and obtaining an overall view of the relationships within the set of elements. The reason for doing this is that each partition would otherwise lead to a single decision result, resulting in a significant challenge to reach a consensual (and successful) decision. Furthermore, the knowledge learned may not fully cover the entire problem space, resulting in situations where an incomplete (which is often in the literature, somewhat misnamed as sparse) rule base is all that is available to perform inference. Thus, techniques that facilitate the exploitation of such sparse knowledge or rule interpolation are needed to support the implementation of the required decision-making systems [13].

Inspired by the aforementioned observations, a method of rough-fuzzy rule interpolation for decision making is introduced in this paper. This is in order to better address the underlying different types of uncertainty [4], thereby determining appropriate decisions. The reason for choosing interpolative technique is that if the views of all individuals only cover part of the problem space, conventional fuzzy inference methods cannot derive a conclusion. The remainder of this paper is organised as follows. Section 2 reviews the general concepts of rough-fuzzy sets and rough-fuzzy rule interpolation. Section 3 illustrates the proposed approach for decision making. Section 4 provides an application that demonstrates the procedures of the proposed work, and verifies its effectiveness in comparison to possible alternative techniques. The paper is concluded in Section 5, including suggestions for possible further work.

2 Background

2.1 Rough-Fuzzy Sets

Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, where $\mathbb{U}$ is a non-empty set (the universe) of finite objects and $\mathbb{A}$ is a nonempty finite set of attributes such that $a : \mathbb{U} \to V_a$ for every $a \in \mathbb{A}$ with V_a being the domain that attribute a takes values from. With any $P \subseteq \mathbb{A}$ there is a crisp equivalence relation $IND(P)$ [15]:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 \mid \forall a \in P,\ a(x) = a(y)\} \tag{1}$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence class with respect to such an indiscernibility relation defined on P is denoted by $[x]_P$, $x \in \mathbb{U}$.

Let $X \subseteq \mathbb{U}$, X can be approximated using only the information contained within P by constructing the P-*lower* and P-*upper* approximations of X [15]:

$$\begin{aligned} \underline{P}X &= \{x|[x]_P \subseteq X\} \\ \overline{P}X &= \{x|[x]_P \cap X \neq \emptyset\} \end{aligned} \tag{2}$$

The tuple $< \underline{P}X, \overline{P}X >$ is called a rough set.

Definition 1. *[3] With any $P \subseteq \mathbb{A}$, an alternative equivalence relation $IND(P)$ to the traditional one of Eq. (1) can be defined by*

$$IND(P) = \{(x, y) \in \mathbb{U}^2 \mid \forall F_g \in P,\ F_g(x) \in C,\ F_g(y) \in C\} \tag{3}$$

where F_g, $g \in \{1, \ldots, G\}$, are fuzzy sets that jointly define a particular concept C in X, $X \subseteq \mathbb{U}$.

Eq. (3) expresses the equivalence relation between the memberships of x and y to different fuzzy sets of a given concept. Using this equivalence relation, the lower and upper approximations for a single C in X can be redefined as follows.

Definition 2. *[3] Let $IND(P)$ be an equivalence relation on $\mathbb{U}$ and F_g, $g \in \{1, \ldots, G\}$, be fuzzy sets in C $(C \in X)$, the lower and upper approximations are a pair of fuzzy sets with membership functions defined by the following, respectively:*

$$\begin{aligned} \mu_{\underline{P}C}(x \in [x]_P) &= inf\{\mu_{F_g}(x), g \in \{1, \ldots, G\}|x \in [x]_P\} \\ \mu_{\overline{P}C}(x \in [x]_P) &= sup\{\mu_{F_g}(x), g \in \{1, \ldots, G\}|x \in [x]_P\} \end{aligned} \tag{4}$$

The tuple $< \underline{P}X, \overline{P}X >$ is called a rough-fuzzy (RF) set (which differs from the alternative use of this term in the literature [2] due to the parallel development of these related but different concepts).

2.2 Rough-Fuzzy Rule Interpolation

An RF set A can be represented by the lower membership function (LMF) A^L and the upper membership function (UMF) A^U, i.e., $A =< A^L, A^U >$. In particular, when triangular membership functions are used, such a set can be illustrated as shown in Fig. 1, where $A^L = (a_0^L, a_1^L, a_2^L; H_A^L)$, $A^U = (a_0^U, a_1^U, a_2^U; H_A^U)$, with (a_0^L, a_1^L, a_2^L) and (a_0^U, a_1^U, a_2^U) denoting the three limit points of the LMF and those of the UMF, respectively, and H_A^L and H_A^U denoting the maximum membership values of A^L and A^U, with $a_0^U \leq a_0^L$, $a_3^L \leq a_3^U$, $0 < H_A^L \leq H_A^U = 1$. Clearly, the closer the shapes of A^L and A^U are, the less uncertain the information contained within A is. When A^L coincides with A^U, the RF set degenerates to a conventional fuzzy set.

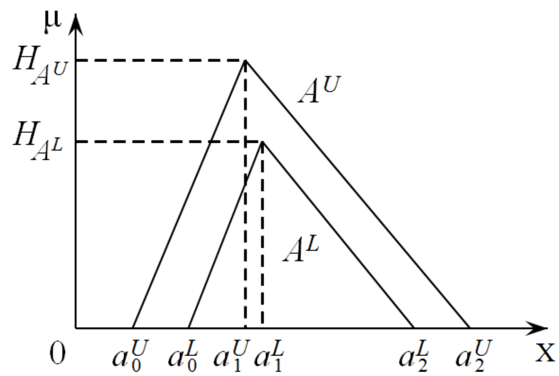

Fig. 1. Lower membership function A^L and upper membership function A^U of a triangular RF set A

Suppose that an RF set A as defined in Fig. 1 has the following six distinct coordinates: $(a_0^L, 0)$, (a_1^L, H_A^L), $(a_2^L, 0)$, $(a_0^U, 0)$, (a_1^U, H_A^U) and $(a_2^U, 0)$. The lower and upper representative values $\text{Rep}(A^L)$ and $\text{Rep}(A^U)$ of A can then be computed, such that:

$$\begin{cases} \text{Rep}(A^K)_x = \dfrac{1}{3}(a_0^K + a_1^K + a_2^K) \\ \text{Rep}(A^K)_y = \dfrac{1}{3}H_A^K \end{cases} \tag{5}$$

where x and y denote a certain variable dimension and the corresponding membership distribution, respectively, $K \in \{L, U\}$.

Definition 3. *[3] The lower and upper shape diversity factors f_A^L and f_A^U are defined by*

$$f_A^K = \sqrt{\frac{\sum_{i=0}^{2}(a_i^K - Rep(A^K)_x)^2}{3}}, \quad K = L, U \tag{6}$$

Definition 4. *[3] The lower and upper weight factors w_A^L and w_A^U are defined as the weights of the shape diversity factors, in terms of the areas of the LMF and UMF, such that*

$$w_A^K = \frac{f_A^K}{f_A^L + f_A^U}, \quad K = L, U \tag{7}$$

where $f_A^L + f_A^U \neq 0$. If however, $f_A^L + f_A^U = 0$, i.e., $f_A^L = 0$ and $f_A^U = 0$, the RF set degenerates to a singleton value, $w_A^L = w_A^U = 1/2$.

Definition 5. *[3] The overall representative value $Rep(A)$ of a given RF set A is defined by*

$$Rep(A) = \sum_{K \in \{L,U\}} (w_A^K \sum_{e \in \{x,y\}} Rep(A^K)_e) \tag{8}$$

where the lower (upper) shape diversity factor is regarded as the weight of the lower (upper) representative value of the LMF (UMF). This is necessary, as otherwise, the same value for representative value would be derived from different shapes of the RF sets.

Given the above definitions, the algorithm for deriving the interpolated conclusion with multiple multi-antecedent rules is summarised below. It follows the procedures as reported in the seminal, and now very popular, transformation-based approach (T-FRI) [7, 8].

Calculate Representative Values. The lower and upper representative values $\text{Rep}(A^K)_x$ and $\text{Rep}(A^K)_y$ of a given RF set A are calculated first using Eq. (5). The shape diversity factors f_A^K and weight factors w_A^K are computed according to Eqs. (6) and (7), respectively. The overall representative value $\text{Rep}(A)$ is then obtained by Eq. (8), $K = L, U$. The calculations for the antecedents of the observation and all rules follow the same procedure.

Choose N Closest Rules. The distances between the artificially created observation and all rules in the rule base are calculated. The N ($N \geq 2$) rules which have minimal distances are then chosen as the N closest rules to perform interpolation.

Construct Intermediate Rule. The normalised weight $w'_{A_{ij}}$ of the jth antecedent of the ith chosen rule, together with the parameter δ_{A_j}, are used to obtain the antecedent of the intermediate rule A'_j for each antecedent dimension x_j, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$. From this, two parameters w'_{B_i} and δ_B are computed and then used to construct B', resulting in the intermediate rule $A'_1 \wedge \cdots A'_j \wedge \cdots A'_M \Rightarrow B'$.

Carry out Scale, Move and Height Transformations. In conjunction with the given A_j^* for each antecedent dimension x_j, the rates s_j^K, m_j^K and h_j, $K \in \{L, U\}$, can then be calculated. Due to the uncertainty introduced in the membership functions, a further transformation on the height of the LMF is needed, while the height of the UMF remains the same owing to its normality. Since the LMFs of different RF sets may have different heights, the height transformation is therefore used to transform the heights of $A_j^{'L}$ to the heights of A_j^{*L}. The *height rate* h is calculated by:

$$h_j = \frac{H_{A_j}^{*L}}{H_{A_j}^{'L}} \tag{9}$$

where $0 < H_{A_j}^{*L} \leq H_{A_j}^{*U} = 1$ and $0 < H_{A_j}^{'L} \leq H_{A_j}^{'U} = 1$, as defined previously. This constraint applies to the interpolated conclusion as well. That is, if the height of B^{*L} is greater than the height of B^{*U} after the height transformation, then $H_B^{*L} = H_B^{*U}$.

Derive Interpolated Conclusion. The second intermediate term B'' and the final interpolated result B^* can then be estimated by the combined s_B^K, m_B^K and h_B, $K \in \{L, U\}$. Here, h_B is computed according to Eq. (9) such that

$$h_B = \frac{1}{M} \sum_{j=1}^{M} h_j \tag{10}$$

3 Rough-Fuzzy Rule Interpolation for Decision Making

3.1 Partition of Problem Space

In order to generate a set of fuzzy rules, each input/output space is divided into K ($K \geq 2$) subspaces. For simplicity, each variable is divided into K fuzzy regions with the corresponding fuzzy sets calculated by the triangular-shaped membership functions. This is illustrated in Fig. 2, where le and re denote the left and right extreme values of variable x_j, respectively. The vertex location of a triangle is determined by its position in the K partition. Any membership value of x_j in a new input below le or above re is set to 1.

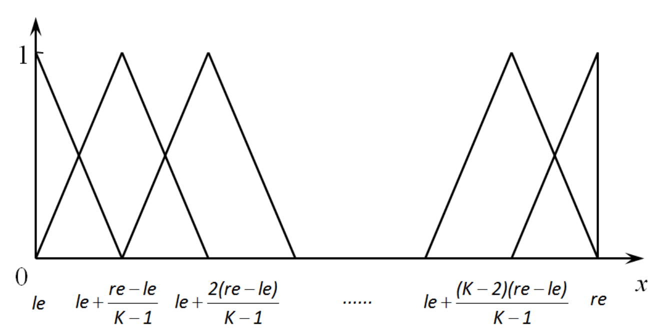

Fig. 2. Partitioning of each input/output space

Without losing generality, suppose that m experts are required to provide their opinions for partitioning the input/output space. In particular, Expert 1 is defined as a generator using $le = 0$ and $re = 1$, while the others are not predefined. Instead, they are randomly constructed from the baseline possessed by Expert 1, that is, $le \in (0, 1)$, $re = 1 - le \in (0, 1)$. This results in $(m - 1)$ other different fuzzy regions. These partitions are then used to determine a set of fuzzy rules that model the relationships between the input-output data pairs. Note that in this initial investigation, unless otherwise stated, m is set to 5.

In order to reflect the gradualness in improving the quality of fuzzy partition, different fuzzy partitions can be created and utilised. In this initial investigation, three partitions are generated and tested where the consequent is each divided into K ($K = 2, 4, 6$) fuzzy subsets, with $K = 2$ representing a rough partition and $K = 6$ representing a detailed partition. Note that the purpose of investigating the use of different dataset partitions is to examine the performance of the proposed approach for fine fuzzy partitions as well as coarse fuzzy partitions.

3.2 Generation of Fuzzy Rules

Given a collection of input-output data pairs, a fuzzy rule base can be formed by creating a rule that best covers a certain given input-output data pair, such that the region with maximum membership degree is assigned to each data pair [17]. This rule generation method is conceptually simple and implementation-wise trivial; it is therefore adopted in this work for the creation of the rule base.

As a result, based on the specification outlined in the preceding subsection, m (or five in the implementation) rule bases can be constructed from the opinions of each expert available. Next, an RF rule base can be built on top of these rule bases. That is, each quintuple fuzzy region is aggregated into an RF set, where the uncertainty is described by the lower and upper approximations. Based on the membership of lower approximation, a given data pair is then allocated to the region with maximum membership degree. Since the lower approximation characterises the grade of certainty, a higher degree indicates a higher certainty. Such a rule base includes different forms of uncertainty by representing the values of the underlying variables as RF sets. These can then be considered in the process of interpolation in order to obtain the required inference conclusions when given an unmatched novel input.

3.3 Implementation of Interpolation

As indicated earlier, the popular transformation-based approach (T-FRI) [7,8] is adopted as the basis for adaptation to perform the interpolative reasoning in the single expert rule base, while the RF rule interpolation is used for the RF rule base. This process is implemented repeatedly for each partition.

Whilst it is a very interesting discovery that empirically, only two closest rules are needed to perform accurate fuzzy rule interpolation if rules are represented as standard fuzzy version (without involving rough-fuzzy sets) [12], it is unclear whether the same conclusion could be drawn for rough-fuzzy rule interpolation. Thus, in this implementation, the number of closest rules with respect to the given unmatched input can be set to a different value to facilitate such a comparative study. In particular, N ($N = 2, 4, 6$) fuzzy rules are chosen as the closest rules to respectively interpolate the conclusions.

3.4 Evaluation of Accuracy

In order to make a comprehensive comparison, each interpolated result is defuzzified to a crisp value using its representative value. Root-mean-square error (RMSE) is adopted to calculate the accuracy (with regard to the underlying ground truth):

$$\epsilon_{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(O_k - G_k)^2} \tag{11}$$

where O_k and G_k denote the kth testing output value and its corresponding ground truth (the ideal consequent of the testing data), respectively.

Ten times 10-fold cross-validation [10, 16] is then employed to evaluate the generalisation ability of the proposed approach. One single subset is maintained as the validation data for testing, while the remaining 9 subsets are used for training. This is then repeated 10 times with each of the subsets used as the testing data, and the rest as the training data. The 10 results from the folds are averaged to produce a single accuracy measure. This process is repeated 10 times by initialising different, randomly assigned initial 10 subsets, and the average of the resultant ten accuracy values is recognised as the final accuracy measure.

4 Experiment and Evaluation

As an initial piece of work, to evaluate the proposed approach, an application to a common benchmark dataset is provided. Particularly, the performance of the approach is examined by running over the UCI *servo* dataset [1]. The RF sets and RF rule interpolation are utilised for problem solving.

Table 1 lists the results of the averaged RMSE in terms of 10 times 10-fold cross-validation in relation to K partitions and N closest rules. Paired t-test results with significance level of 0.05 are also identified with the achieved accuracies of the RF approach as reference, those significantly better, worse and no difference are marked with "(v)", "(*)" and "(-)", respectively.

As reflected in the table, the accuracies from five separate expert rule bases are unstable. That is, the opinions from a certain individual expert may perform well in certain partitions, and badly in others. Theoretically, this is acceptable as someone is only an expert in a particular field, namely, the necessary expertise may be only available for a certain concept. However, this leads to difficulty for making informed decisions in practical applications. Since different opinions result in a better or worse accuracy, it is difficult to conclude which one should be chosen.

Table 1. Accuracies (RMSE×100) over servo dataset

	Partition = 2					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	20.75(*)	20.41(*)	20.10(*)	20.58(*)	20.25(*)	19.88
4	22.38(*)	22.07(*)	21.78(*)	22.22(*)	21.92(*)	21.22
6	22.08(*)	21.75(*)	21.44(*)	21.92(*)	21.59(*)	20.75
	Partition = 4					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	10.72(*)	10.50(-)	10.71(*)	10.70(*)	10.65(-)	10.54
4	12.67(-)	12.95(*)	13.15(*)	12.82(*)	13.04(*)	12.71
6	11.88(-)	12.10(*)	12.35(*)	12.02(*)	12.25(*)	11.91
	Partition = 6					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	10.00(*)	9.90(*)	10.09(*)	10.12(*)	9.92(*)	9.74
4	11.86(-)	11.98(-)	12.37(*)	11.89(-)	12.08(*)	12.00
6	10.94(v)	11.05(-)	11.52(*)	10.96(v)	11.22(*)	11.14

Fortunately, it can be seen that over different domain partitions and different number of closest rules for use in the interpolation, the accuracies obtained by the RF approach are generally better or at least comparable to those achievable by the use of the conventional T-FRI (which is implemented on the expertise offered by one and only one of the five experts). This reflects an important advantage of the proposed approach in that more considered information produces better results, integrating different perspectives of the domain experts. This shows that the range of uncertainties provides useful information on depicting a concept.

Additionally, note that the effect of the number of N is also revealed in this experimental study. The choice and use of just two closest neighbouring rules leads to a performance superior to that of four or six rules. This reflects that more than two neighbouring rules does not necessarily result in more accurate interpolated outcomes, but can be counter-productive, possibly due to the introduction of noise caused by the use of further rules that are farther away from the unmatched input. This finding conforms to what has been empirically established in the literature for the use of standard T-FRI method [12].

5 Conclusion

A data-driven rule-based fuzzy system works on the generation of fuzzy rules from numerical data. However, different expert opinions on the fuzzy partitions of the problem domain may result in a range of uncertainties in the representation of the rule base. The approach introduced in this paper helps provide an effective solution to this practical challenge, by including such uncertainty information into the inference process. Experimental results have shown that the exploitation of uncertain knowledge across multiple opinions offered by different experts generates better results than the use of just the expertise offered by a single expert.

Although promising, there is still room to improve the current work. For instance, all attributes in the rules are of equal weighting in deriving the conclusions, it is interesting to investigate the relative significance of individual attributes and differentiate the contributions of the rule antecedents [11]. Of course, this initial investigation only applies the proposed work to a single benchmark dataset, further evaluation of its performance over many other datasets forms another piece of future research.

Acknowledgment

This research was partly supported by the Research Foundation of Chongqing University of Science and Technology, China (Grant No. CK2016B04).

References

1. K. Bache and M. Lichman, “UCI machine learning repository,” 2013. [Online]. Available: http://archive.ics.uci.edu/ml

2. M. Banerjee and S. Pal, "Roughness of a fuzzy set," *Information Sciences*, vol. 93, no. 3, pp. 235–246, 1996.
3. C. Chen, N. M. Parthalain, Y. Li, C. Price, C. Quek, and Q. Shen, "Rough-fuzzy rule interpolation," *Information Sciences*, vol. 351, pp. 1–17, 2016.
4. X. Fu and Q. Shen, "Fuzzy compositional modeling," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 4, pp. 823–840, 2010.
5. A. Gordon and M. Vichi, "Fuzzy partition models for fitting a set of partitions," *Psychometrika*, vol. 66, no. 2, pp. 229–247, 2001.
6. S. Guillaume, "Designing fuzzy inference systems from data: an interpretability-oriented review," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 426–443, 2001.
7. Z. Huang and Q. Shen, "Fuzzy interpolative reasoning via scale and move transformations," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 340–359, 2006.
8. ——, "Fuzzy interpolation and extrapolation: a practical approach," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 13–28, 2008.
9. V.-N. Huynh, Y. Nakamori, and J. Lawry, "A probability-based approach to comparison of fuzzy numbers and applications to target-oriented decision making," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 2, pp. 371–387, 2008.
10. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the International Joint Conference on Artficial Intelligence*, 1995, pp. 1137–1143.
11. 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, pp. 4508–4517, 2020.
12. F. Li, C. Shang, Y. Li, J. Yang, and Q. Shen, "Interpolation with just two nearest neighboring weighted fuzzy rules," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2255–2262, 2020.
13. ——, "Approximate reasoning with fuzzy rule interpolation: Background and recent advances," *Artificial Intelligence Review*, 2021.
14. Y. Lin, G. A. Cunningham III, and S. V. Coggeshall, "Using fuzzy partitions to create fuzzy systems from input-output data and set the initial weights in a fuzzy neural network," *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 4, pp. 614–621, 1997.
15. Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.
16. J. Shao, "Linear model selection by cross-validation," *Journal of the American statistical Association*, vol. 88, no. 422, pp. 486–494, 1993.
17. L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.