Aberystwyth University

A Novel Framework of Fuzzy Rule Interpolation for Takagi-Sugeno-Kang Inference Systems
Zhang, Pu; Shen, Qiang

A Novel Framework of Fuzzy Rule Interpolation for Takagi-Sugeno-Kang Inference Systems

Pu Zhang and Qiang Shen

Department of Computer Science, Faculty of Business and Physical Sciences
Aberystwyth University, Aberystwyth SY23 3DB, UK
{puz, qqs}@aber.ac.uk

Abstract—Fuzzy rule interpolation (FRI) technique has been proposed to infer conclusions for unmatched instances when a fuzzy rule-based system is presented with a sparse rule base. Most existing FRI methodologies are not developed for Takagi-Sugeno-Kang (TSK) inference models. TSK inference extension (TSK+) is one of the methodologies proposed for TSK models with sparse rule bases. It works by replacing matching degrees with similarity measures across all the given rules, instead of just the matched ones, to generate the final conclusion. However, those rules with low similarities bring bias to the final result, which is mainly determined by the closest rules. To significantly strengthen the efficacy of this, a novel framework is presented here through the use of just a small number of closest rules to derive the final outcome. Compared with TSK+, the proposed method reduces the computational overheads of the inference process while avoiding the adverse impact caused by the rules of low similarities with the new observation. More importantly, to deal with large sized sparse rule bases, where neighbourhood rules may be similar with each other, a rule-clustering approach is proposed. That is, a clustering algorithm (say, fuzzy c-means) is first employed to cluster rules into different groups and then, the final interpolated conclusion is computed by the use of the closest rules selected from a small number of closest rule clusters. This approach helps further decrease the time complexity. The efficacy of these two modified methods is demonstrated via systematic experimental comparisons against the performance of the original TSK+.

Index Terms—Fuzzy inference system, TSK, Fuzzy rule interpolation, Rule clustering, K closest rules, Closest rule clusters.

I. INTRODUCTION

Fuzzy rule based inference systems are one successful representative of knowledge-based systems, the basic idea of which is representing domain knowledge in the form of "if-then" production rules. These rules are generally applied such that if the input observations match rule antecedents, then the outputs are derived from the corresponding rule consequents [1] [2]. However, in traditional rule based systems, uncertain and linguistic terms are hard to be described precisely, such as fast, slow, young, old. Fortunately, with the support of fuzzy logic and fuzzy set theory, fuzzy rule-based inference systems allow all such terms to be represented by fuzzy sets, enabling the inference process to resemble human reasoning.

There are several types of fuzzy rule inference system that have been developed in the literature. Mamdani models [3] and Takagi-Sugeno-Kang (TSK) models [4] are two conventional and most widely used ones. The antecedents and consequents of Mamdani models are both represented by fuzzy sets. Thus, a defuzzification process is usually required to obtain crisp results in practice. On the contrary, TSK models use polynomials as rule consequents, resulting in crisp conclusions directly and more applicable for solving regression problems.

In fuzzy rule based systems, when the input domain is not fully covered, it is possible that an observation does not match any rule in the given rule base, thereby no conclusion can be produced using traditional rule-firing mechanisms. This is independent of what rule models are employed. Rule bases in this situation are named as sparse rule bases. Fuzzy rule interpolation (FRI) has been introduced to deal with this issue. When an observation does not overlap with any rule antecedent, FRI helps generate an intermediate rule by the approximation of neighbour rules to the observation in order to obtain a potentially relevant conclusion. Although a number of FRI methodologies have been established over the last decades, such as linear interpolation [5], transformation-based interpolation [6], adaptive interpolation [7] and GA-aided dynamic interpolation [8], most of them are developed for Mamdani models rather than for TSK models.

TSK inference extension (TSK+) [9] is a novel fuzzy inference approach based on the TSK model which extends its capability of handling sparse fuzzy rule bases. Instead of exploiting matching degrees, a similarity measure based on a certain distance metric is utilised to perform interpolative inference, with all rules in the rule base being involved in the interpolation. As such, even if an observation matches no rule, a certain conclusion can be generated. Whilst being a useful approach, TSK+ has its own shortcomings. In particular, far away rules are usually not relevant to the observation but may still bring (often counter-productive) biases to the final interpolated outcome. This is in addition to the artificial introduction of unnecessary computation, increasing computing overheads in vain.

To address these limitations, a modified approach is proposed in this paper, which has two forms of implementation. The underlying principle is to perform interpolation with just K closest rules (KCR) where K is normally a small number, that is, only K rules closest to the observation will contribute to the final conclusion. In so doing, the adverse impact caused by rules with low similarities can be avoided while incurring less computation. Furthermore, in cases where large sized

sparse rule bases are present, the K closest rules may be very similar with each other, which leads to lack of diversity for interpolation. Therefore, another implementation is to carry out interpolation with K closest rule clusters (CRC), i.e., one closest rule is selected from each of a small number of closest rule clusters for interpolation. According to systematic experimental comparisons, KCR has led to improved results over TSK+ and CRC in small sized sparse rule bases, while for systems involving large sized sparse rule bases, CRC outperforms TSK+ and KCR.

The rest of this paper is structured as follows. For completeness, Section II briefly outlines the inference process of the conventional TSK model. Section III reviews the TSK inference extension (TSK+). Section IV and Section V detail the two aforementioned improved implementations, for interpolation with K closest rules (KCR) and that with K closest rule clusters (CRC), respectively. Section VI describes the setting of the experiments carried out and discusses the results of comparative experimental evaluations. Finally, Section VII concludes the paper with future research pointed out.

II. TSK Fuzzy Inference Models

The TSK fuzzy inference model was originally developed by Takagi, Sugeno, and Kang in 1985 [4]. In general, suppose that each rule is of n antecedent variables. Within a TSK fuzzy rule base, a rule is then defined by

$$
\begin{aligned}
R_i : \ &\textbf{if } x_1 \text{ is } A_{i1}, ..., x_n \text{ is } A_{in}, \\
&\textbf{then } f_i(x_1, ..., x_n) = a_{i0} + a_{i1}x_1 + ... + a_{in}x_n
\end{aligned}
\tag{1}
$$

where $A_{i1}, ..., A_{in}$ are the fuzzy sets taken by the rule antecedent variables and $a_{i0}, a_{i1}, ..., a_{in}$ are the parameters specifying the polynomials of the rule's consequent.

Given an observation $O(B_1, ..., B_n)$, the TSK inference process can be briefly described as follows:

1) Calculate the **matching degrees** between the antecedent variables of the observation O and their counterparts in each rule R_i:
   $D(A_{i1}, B_1), ..., D(A_{in}, B_n)$
2) Determine the weight of R_i by integrating all matching degrees:
   $\alpha_i = D(A_{i1}, B_1) \wedge ... \wedge D(A_{in}, B_n)$
   where $\wedge$ is usually implemented by a minimum operator.
3) Take the observation O as the input to compute the rule consequent polynomial for each of the k matched rules, resulting in sub-conclusions:
   $f_i(B_1, ..., B_n) = a_{i0} + a_{i1}B_1 + ... + a_{in}B_n$
4) Integrate all sub-conclusions to obtain the final outcome for the consequent by weighted average:

$$
f(O) = \frac{\Sigma_{i=1}^{k} \alpha_i f_i(B_1, ..., B_n)}{\Sigma_{i=1}^{k} \alpha_i}
\tag{2}
$$

If the given observation matches no rule, the weight of each rule α_i will be 0. Thus, neither sub-conclusion nor final result can be generated. In this case, the conventional TSK model will fail. The fuzzy rule bases that suffer from this limitation

are known as sparse fuzzy rule bases, namely they do not cover such observations. As mentioned previously, FRI has been developed to generate conclusions for unmatched observations by exploiting the approximation of their neighbouring rules.

III. TSK Inference Extension (TSK+)

TSK inference extension (TSK+) [9] offers a novel fuzzy reasoning approach for extending TSK inference, making it applicable to handle sparse rule bases. Instead of using matching or overlapping degrees that the conventional TSK model utilises, TSK+ employs a modified similarity measure based on Euclidean distance [10] to evaluate relationships between an observation and the given rules. In this procedure, the similarities between the observation and all rules are always greater than zero. Thus, all rules will be involved in the derivation of the final consequent outcome. In so doing, even if an observation matches no rule, a conclusion can still be approximately derived.

A. TSK+ Procedure

Suppose that an observation $O(B_1, ..., B_n)$ and a sparse rule base are given, and that the rule base comprises m rules with n antecedent variables, with each rule being specified as per Eqn. (1). The inference procedure of TSK+ can be summarised as the following:

1) Calculate the **similarities** between the observation and rule R_i: $S(A_{i1}, B_1), ..., S(A_{in}, B_n)$
2) Determine the weight of rule R_i:
   $\alpha_i = S(A_{i1}, B_1) \wedge ... \wedge S(A_{in}, B_n)$
3) Integrate all similarity measures to obtain an interpolated rule, with the parameters of its consequent being:

$$
a_0 = \frac{\Sigma_{i=1}^{m} \alpha_i a_{i0}}{\Sigma_{i=1}^{m} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{m} \alpha_i a_{in}}{\Sigma_{i=1}^{m} \alpha_i}
\tag{3}
$$

4) Take the observation O as the input of the interpolated rule and compute the interpolated outcome as
   $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$

It is clear that TSK+ has similar inference steps as the standard TSK model (as outlined in Section II), except when no rules match a given observation, the matching degrees are replaced with similarity measures and all m rules are used to compute the final result (rather than the otherwise k matched ones). Note that the time complexity of TSK+ inference process is $O(mn)$.

B. Similarity Measure

The similarity measure applied in TSK+ is revised from the one proposed in [10]. In particular, a distance factor (DF) is utilised to increase the sensitivity of the similarity measure to the distance. As empirically proven, if the membership functions can be appropriately fine-tuned, the use of different types of membership function has little impact upon the fuzzy rule-based inference results [9] [11]. Based on this empirical observation, and for computational simplicity, only triangular membership functions will be considered in this paper.

Suppose that there are two normalized fuzzy sets, represented by triangular membership function $A = (a_1, a_2, a_3)$ and

$A' = (a'_1, a'_2, a'_3)$ respectively, their similarity degree $S(A, A')$ can be defined as follows:

$$S(A, A') = (1 - \frac{\Sigma_{i=1}^3 |a_i - a'_i|}{3}) \cdot DF$$

$$DF = 1 - \frac{1}{1 + e^{-sd+5}}$$

(4)

where d is the Euclidean distance between the gravity centres (or alternatively, representative values [6]) of the two fuzzy sets and s represents a sensitivity factor (a smaller value of s makes DF more sensitive to the distance measure). The constant 5 in this definition ensures that DF is normalized as 1 when d is 0. According to the definition, the greater the value of $S(A, A')$, the closer and more similar the two fuzzy sets A and A'. $S(A, A') = 1$ if and only if A and A' are identical [9].

The effectiveness and applicability of the modified similarity measure has been validated in [12] by comparing several most commonly used similarity measures, such as $S = 1 - d$ and $S = 1/(1+d)$ and the original similarity measure as given in [10].

IV. INTERPOLATION WITH K CLOSEST RULES (KCR)

When applying TSK+, it is observed that several closest rules have much higher similarity degrees than others. This may indicate that the final results are mainly determined by these closest rules. Moreover, a basic presumption generally assumed in FRI is that the interpolated consequent is estimated by the neighbouring rules to the observation [3] [4], the closest rules contain most relevant information while far away rules may introduce bias into the results, often counter-productively. Although such biases do not impose much influence upon the interpolated results due to their relatively smaller similarity measures, they do incur significant computational overheads and hence, should be minimised.

In light of this observation, this work introduces a revised inference procedure, termed *interpolation with K closest rules (KCR)*, based on the same similarity measure (4) applied in TSK+. In particular, only K closest neighbouring rules to the observation are involved in the interpolated rule generation, rather than involving all the rules in the sparse rule base.

A. KCR Procedure

Suppose that a sparse rule base is given, containing m rules with n antecedent variables, together with an observation $O(B_1, ..., B_n)$, where each rule is specified in the format of Eqn. (1). Then, the process of KCR can be detailed as follows:
1) Calculate the overall Euclidean distance between the representative values of the individual variables within the observation and those of the antecedent variables for each given rule.
2) Select K closest rules by the Quickselect algorithm [13] (which is utilised purely for efficiency while any alternative selection mechanism may be employed if preferred).
3) Calculate the similarity between the observation O and each of R_i that belongs to the set of the selected K

closest rules:
$S(A_{i1}, B_1), ..., S(A_{in}, B_n)$
4) Determine the weight of rule R_i:
$\alpha_i = S(A_{i1}, B_1) \wedge ... \wedge S(A_{in}, B_n)$
5) Integrate all K similarities to obtain a working interpolated rule with the following parameters for its consequent:

$$a_0 = \frac{\Sigma_{i=1}^K \alpha_i a_{i0}}{\Sigma_{i=1}^K \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^K \alpha_i a_{in}}{\Sigma_{i=1}^K \alpha_i}$$

(5)

6) Take the observation O as the input to fire the interpolated rule such that the consequent is computed by $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$

B. KCR Complexity

From the above process and Eqn. (4), it can be seen that whilst the Euclidean distance forms only a small part in the calculation of similarity measures, it captures the essential relationships between an observation and the rules. It is appropriately utilised for the purpose of efficient determination of closest rules, without resorting to the more complicated similarity measurement. Thus, the similarity measure is only applied K times for K selected rules rather than all rules. In this case, the modified approach can significantly reduce the running time. Additionally, the Quickselect algorithm helps decrease the computation of closest rule selection.

In summary, the time complexity of the proposed implementation for KCR is $O(mK+nK)$, where $O(mK)$ stands for the time complexity of K rules selection. In comparison, the time complexity of TSK+ is $O(mn)$ as mentioned previously. Note that generally, K is much smaller than m and n. Thus, the proposed approach has significantly lower time complexity.

V. INTERPOLATION WITH K CLOSEST RULE CLUSTERS (CRC)

When applying KCR in large sized sparse rule bases (e.g., for a rule base consisting of more than 200 rules), it is observed that the K closest rules with the greatest similarity degrees may appear to be very similar. The information reflecting the approximate relationships holding between the antecedent variables and the consequent may therefore be very similar also. If only just these K rules are taken into account, the interpolated rule will be also similar to them regardless the actual similarity measures. In TSK+, despite that all rules are involved in rule interpolation, this problem remains because the similarities of the K rules are much larger than the rest and the final result is therefore, still mainly determined by these closest ones.

To extend the diversity of rules used for interpolation without involving far too many similar rules, a clustering-aided inference process is proposed, termed *interpolation with K closest rule clusters (CRC)* hereafter. Rules in sparse rule bases are firstly clustered into different groups based on their representative values by a clustering method. Here, the popular fuzzy c-means algorithm [14] is adopted to implement this. Rules in the same clusters are deemed to contain similar

information. As such, K closest clusters are selected so that only one rule which is the nearest to the observation within each cluster is selected for use as an element of the set of K closest rules. The conclusion will be interpolated by such resulting K closest rules. In so doing, other rules measured without necessarily having the higher similarity measures will be able to participate in the generation of the final interpolated consequent. Yet, this approach always ensures that the closest rule with the largest similarity measure is selected to participates in rule interpolation, since it always is the representative of a certain cluster of rules given its highest similarity measure.

A. CRC Procedure

Suppose that a sparse rule base which contains m rules with n antecedents and an observation $O(B_1, ..., B_n)$ are given, with each rule being specified as per Eqn. (1). The detail of the proposed procedure for CRC is described in the following:

1) Cluster all rules into C different groups by their representative values, using fuzzy c-means.
2) Calculate the Euclidean distance between the observation and all cores of the C clusters and select K ($K \leq C$) closest clusters.
3) Choose one of the K clusters and compute the distance between the observation O and each and every rule in it.
4) Find the closest rule R_i in the selected cluster as the representative of this cluster.
5) Determine the weight of the rule R_i:
   $\alpha_i = S(A_{i1}, B_1) \wedge ... \wedge S(A_{in}, B_n)$
6) Repeat Steps 3,4 and 5 for all K selected clusters, obtaining K rules and corresponding similarities.
7) Integrate all K similarities to obtain the interpolated rule, the parameters of the consequent will be:

$$a_0 = \frac{\Sigma_{i=1}^{K}\alpha_i a_{i0}}{\Sigma_{i=1}^{K}\alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{K}\alpha_i a_{in}}{\Sigma_{i=1}^{K}\alpha_i} \quad (6)$$

8) Take the observation O as the input to fire the interpolated rule and compute the final consequent outcome with respect to the observation:
   $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$

B. CRC Complexity

The above SRS process has a time complexity of $O(KC + KG + nK)$, where $O(KC)$ stands for the complexity to conduct K clusters selection, and $O(KG)$ denotes that for K rules selection with one from each cluster, G being the largest number of rules contained within any cluster. Compared with KCR, of which time complexity is $O(mK + nK)$ as previously analysed, CRC can also decrease the computation effort required to perform similarity measurement. In addition, CRC does not need to compute the distances between the observation and all rules but only those of the cores of clusters and the rules in the K selected clusters. Therefore, $O(KC + KG)$ is generally smaller than $O(mK)$. In other words, the time complexity of the inference process can be further reduced by CRC.

VI. Experimental Evaluation

In this section, the performance of the proposed novel framework of FRI for the TSK model, which is implemented with two inference methods (namely KCR and CRC), is experimentally compared against TSK+ over three benchmark datasets. The datasets run include a nonlinear mathematical model and two real-world datasets (Stock and Quake [15]). In particular, the Stock dataset is adopted to evaluate their performance regarding small sized sparse rule bases, while the nonlinear model and Quake dataset represent sparse rule bases in large sizes.

A. Generation of Sparse Fuzzy Rule Bases

In the present experimental study a sparse fuzzy rule base is created artificially from a dense fuzzy rule base that is induced from the original datasets at first. This will offer an opportunity for the revealing of the potential of FRI should part of the underlying rules be unavailable.

Here, a simple data-driven fuzzy rule base generation method is employed: The instances in a given dataset are firstly clustered into different categorises through a classical clustering algorithm, fuzzy c-means [14]. Since fuzzy c-means allows a data point to belong to more than one cluster with different membership values, in this work, if an instance has larger than 0.2 membership value to a cluster, it is deemed as belonging to this category. As mentioned earlier, rule antecedent variables take fuzzy values represented by triangular membership functions. The three parameters of a triangular membership function are implemented by the infimum, centre and supremum of the corresponding cluster. The consequent of a rule, which is a polynomial, is then derived by the popular linear regression approach as per the work of [16].

The sparse fuzzy rule bases can be generated by randomly removing a certain number of rules from the resulting dense fuzzy rule bases. Specifically, in each of the following experiments, to emphasise on rule base sparsity, only 80% and 60% rules are retained to form the sparse fuzzy rule bases for inference.

B. Evaluation Methodology

To enable fair comparison, 10 times 10-fold cross-validation is employed. Training sets are used to generate sparse fuzzy rule bases by the above process while testing sets to evaluate the performance described by RMSE (root-mean-square error, in relation to the ground truth). The results are demonstrated via the following three criteria: the amount of the best results among 100 folds, Gaussian fitting and boxplots.

C. On Stock Dataset

The stock dataset investigated provides stock prices for ten aerospace companies. The task is to predict the price for the 10th company given the prices for the rest [15]. The dataset consists of 950 instances and 9 features (i.e., antecedent variables). The output domain is [34, 62]. Fifty rules extracted from the training sets constitute the dense fuzzy rule base. For the other parameters in this experiment, regarding KCR

the number of closest rules K is empirically set to 3, and regarding CRC, the number of clusters C is set to 5 with the number of closest rule clusters K set to 3.

Table I shows the parameters of Gaussian fitting and the amounts of the best results in 80% and 60% sparse fuzzy rule bases, while Fig. 1 and Fig. 2 illustrate the boxplots of the results. As can be seen from these results, KCR has the best and the most robust results, and TSK+ has slightly worse results due to all rules are involved and then bring forward adverse biases. CRC does not work well in small sized sparse fuzzy rule bases because the high relevant rules are clustered into one cluster and does not contribute to the final results generation. Note that although TSK+ does not produce any overall best result, it outperforms CRC in this particular case where the total number of rules within the rule base is rather small, apart from the rule base being sparse.

D. On Nonlinear Function Model

In this experiment, a dataset randomly sampled from a 3-dimensional nonlinear function is used as a benchmark dataset. Note that the random sampling method has been employed by a number of projects (e.g., those reported in [12] and [17]), and that the nonlinear function applied herein has been used in [9] and [18], which is given below:

$$F(x,y) = \frac{sin(x)sin(y)}{xy} \quad (7)$$

Two thousand points are randomly sampled as the original dataset. Each dense fuzzy rule base comprises 200 rules. The output domain is [-0.217, 1]. For KCR the number of closest rules K is set to 3, and for CRC, the number of rule clusters C is set to 10 with the number of closest rule clusters K set to 3. The means and standard deviations (SD) of Gaussian fitting and the amounts of the best results for TSK+, KCR and CRC while running on 80% and 60% sparse fuzzy rule bases are displayed in Table II, and the boxplots are shown in Fig. 3 and Fig. 4, respectively.

These results show that for this nonlinear model, in both cases of running the two sparse fuzzy rule bases, CRC is the overall winner. It obtains most of the best results, the best values of mean in Gaussian fitting and the best value of the median and interquartile range in boxplots. However, the SD of CRC is larger than the other two methods (TSK+ and KCR). One possible reason for this is that the number of rule clusters C is manually decided rather than automatically generated, in this specific dataset, such a manual decision does not ensure to lead to the best rule clusters for all training sets. In addition, when comparing TSK+ with KCR, KCR performs much better than TSK+, which confirms that although all (the sparse) rules are involved in deriving the final outcome, TSK+ does not cope with the problem well, which is caused by running many similar closest rules in such a large sized sparse rule base, as indicated in Section V.

E. On Quake Dataset

The quake dataset contains 2178 instances and 3 antecedent variables. The regression task is to estimate the strength of an

TABLE I
GAUSSIAN FITTING AND AMOUNT OF BEST ON STOCK DATASET

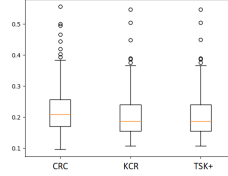|		Gaussian fitting	Number of best
	TSK+	0.21938, 0.09890	0
80%	KCR	**0.21932, 0.09888**	**88**
	CRC	0.24320, 0.11515	12
	TSK+	0.31831, 0.15113	0
60%	KCR	**0.31812, 0.15121**	**89**
	CRC	0.33991, 0.16307	11

Fig. 1. Boxplot of results running with 80% sparse rule base on stock dataset
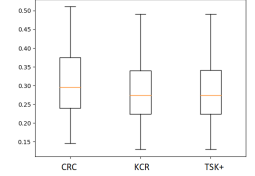

Fig. 2. Boxplot of results running with 60% sparse rule base on stock dataset

earthquake based on the depth of its focal point, its latitude and its longitude [15]. The output domain is [5.8, 6.9].

Two hundred rules are generated from the training sets as the dense fuzzy rule base. In this experiment, the number of closest rules K in KCR is set to 3, and the number of rule clusters and that of closest rule clusters are set to 10 and 3, respectively.

Table III lists the results of Gaussian fitting and the amounts of the best results, and Fig. 5 and Fig. 6 describe the distribution of the results obtained by running 80% and 60% sparse rule bases in boxplots.

From these results, once again, it can be seen that CRC outperforms TSK+ and KCR overall. In the comparison of TSK+ and KCR, KCR also performs slightly better than TSK+ in terms of accuracy, whilst taking much less computation

TABLE II
GAUSSIAN FITTING AND AMOUNT OF BEST FOR FUNCTION MODEL

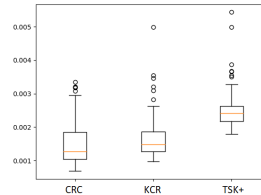|		Gaussian fitting	Number of best
	TSK+	0.00251, 0.00056	0
80%	KCR	0.00167, 0.00061	30
	CRC	**0.00150, 0.00065**	**70**
	TSK+	0.00342, 0.00063	0
60%	KCR	0.00307, 0.00069	25
	CRC	**0.00266, 0.00084**	**75**

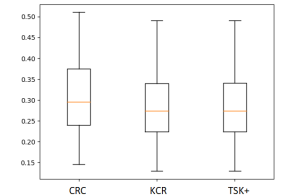Fig. 3. Boxplot of results with 80% sparse rule base on function model

Fig. 4. Boxplot of results with 60% sparse rule base on function model

time. This, together with the results of the experiments on the nonlinear and quake datasets, demonstrates that for large sized sparse rule bases, CRC can obtain much better results than KCR and TSK+, showing the great potential of clustering-aided fuzzy rule interpolation.

VII. CONCLUSION

This paper has presented a novel framework with two implementations suitable for performing fuzzy rule interpolation with TSK fuzzy inference models. The work has been motivated by the observation that the existing method, TSK+ involves the use of all given rules, including redundant or even possibly irrelevant rules in an attempt to compute the final conclusion. The framework entails the generation of more accurate interpolated results. In particular, for small sized sparse rule bases, the corresponding implementation (KCR) only requires the use of a small number of closest rules. When applied for systems with large sized sparse fuzzy rule bases, fuzzy c-means has been applied to clustering the rules first so that only one closest rule from each of a small number of the resulting rule clusters is utilised to perform interpolation (CRC). Systematic comparative experimental studies have demonstrated the effectiveness of both implementations.

The proposed work offers many opportunities for further development. For instance, CRC directly employs the original fuzzy c-means algorithm in rule clustering, but it may not generate the most appropriate categories since the rule bases are sparse in the first place. Modified fuzzy c-means algorithms, e.g., the kernel fuzzy c-means [19] and suppressed fuzzy c-means [20] may be adopted as the alternative to strengthen the performance. Also, the parameters required to carry out interpolation, such as the number of closest rules and that of the clusters are herein set manually. Introducing an automated way to decide on these parameters from the training data remains a challenge. Furthermore, all antecedent variables are treated equally in the present implementations, how weighted representations as per the most recent work of [21] may be extended to accommodating interpolation with TSK models forms another interesting piece of active research.

REFERENCES

[1] S.-H. Liao, "Expert system methodologies and applicationsa decade review from 1995 to 2004," *Expert systems with applications*, vol. 28, no. 1, pp. 93–103, 2005.
[2] C. Grosan and A. Abraham, "Rule-based expert systems," in *Intelligent Systems*, pp. 149–185, Springer, 2011.
[3] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," in *Proceedings of the sixth international symposium on Multiple-valued logic*, pp. 196–202, IEEE Computer Society Press, 1976.
[4] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE transactions on systems, man, and cybernetics*, no. 1, pp. 116–132, 1985.
[5] L. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *International Journal of Approximate Reasoning*, vol. 9, no. 3, pp. 197–225, 1993.
[6] Z. Huang, Q. Shen, *et al.*, "Fuzzy interpolation and extrapolation: A practical approach," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 13–28, 2008.
[7] L. Yang and Q. Shen, "Adaptive fuzzy interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 6, pp. 1107–1126, 2011.
[8] N. Naik, R. Diao, and Q. Shen, "Genetic algorithm-aided dynamic fuzzy rule interpolation," in *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pp. 2198–2205, IEEE, 2014.
[9] J. Li, L. Yang, Y. Qu, and G. Sexton, "An extended takagi–sugeno–kang inference system (tsk+) with fuzzy interpolation and its rule base generation," *Soft Computing*, vol. 22, no. 10, pp. 3155–3170, 2018.
[10] S.-J. Chen and S.-M. Chen, "Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers," *IEEE Transactions on fuzzy systems*, vol. 11, no. 1, pp. 45–56, 2003.
[11] P.-C. Chang and C.-Y. Fan, "A hybrid system integrating a wavelet and tsk fuzzy rules for stock price forecasting," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 6, pp. 802–815, 2008.
[12] J. Li, Y. Qu, H. P. Shum, and L. Yang, "Tsk inference with sparse rule bases," in *Advances in Computational Intelligence Systems*, pp. 107–123, Springer, 2017.
[13] C. A. Hoare, "Algorithm 65: find," *Communications of the ACM*, vol. 4, no. 7, pp. 321–322, 1961.
[14] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.
[15] H. A. Guvenir, I. Uysal, and F. A. Repositor, "Function approximation repository," *Bilkent University. URL http://funapp. cs. bilkent. edu. tr/DataSets*, 2000.
[16] D. A. Freedman, *Statistical models: theory and practice.* cambridge university press, 2009.
[17] H. Bellaaj, R. Ketata, and M. Chtourou, "A new method for fuzzy rule base reduction," *Journal of Intelligent & Fuzzy Systems*, vol. 25, no. 3, pp. 605–613, 2013.
[18] B. Rezaee and M. F. Zarandi, "Data-driven fuzzy modeling for takagi–sugeno–kang fuzzy system," *Information Sciences*, vol. 180, no. 2, pp. 241–255, 2010.
[19] D.-Q. Zhang and S.-C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation," *Artificial intelligence in medicine*, vol. 32, no. 1, pp. 37–50, 2004.
[20] J.-L. Fan, W.-Z. Zhen, and W.-X. Xie, "Suppressed fuzzy c-means clustering algorithm," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1607–1612, 2003.
[21] 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*, 2018.

TABLE III
GAUSSIAN FITTING AND AMOUNT OF BEST ON QUAKE DATASET

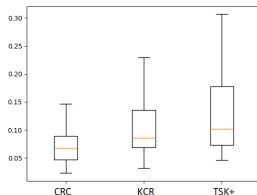|		Gaussian fitting	Number of best
80%	TSK+	0.18795, 0.20215	0
	KCR	0.16468, 0.19636	30
	CRC	**0.10440, 0.14123**	**70**
60%	TSK+	0.26869, 0.41907	1
	KCR	0.26275, 0.40029	36
	CRC	**0.20246, 0.28600**	**63**

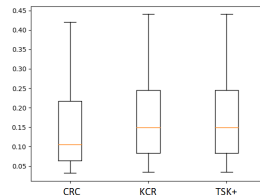Fig. 5. Boxplot of results with 80% sparse rule base on quake dataset

Fig. 6. Boxplot of results with 60% sparse rule base on quake dataset