TSK Inference with Sparse Rule Bases

Jie Li, Yanpeng Qu, Hubert P. H. Shum, Longzhi Yang

Abstract The Mamdani and TSK fuzzy models are fuzzy inference engines which have been most widely applied in real-world problems. Compared to the Mamdani approach, the TSK approach is more convenient when the crisp outputs are required. Common to both approaches, when a given observation does not overlap with any rule antecedent in the rule base (which usually termed as a sparse rule base), no rule can be fired, and thus no result can be generated. Fuzzy rule interpolation was proposed to address such issue. Although a number of important fuzzy rule interpolation approaches have been proposed in the literature, all of them were developed for Mamdani inference approach, which leads to the fuzzy outputs. This paper extends the traditional TSK fuzzy inference approach to allow inferences on sparse TSK fuzzy rule bases with crisp outputs directly generated. This extension firstly calculates the similarity degrees between a given observation and every individual rule in the rule base, such that the similarity degrees between the observation and all rule antecedents are greater than 0 even when they do not overlap. Then the TSK fuzzy model is extended using the generated matching degrees to derive crisp inference results. The experimentation shows the promising of the approach in enhancing the TSK inference engine when the knowledge represented in the rule base is not complete.

Jie Li, Hubert P. H. Shum, Longzhi Yang
Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK, e-mail: {jie2.li, hubert.shum, longzhi.yang}@northumbria.ac.uk

Yanpeng Qu
Information Science and Technology College, Dalian Maritime University, Dalian, 116026, China, e-mail: yanpengqu@dlmu.edu.cn

1 Introduction

Fuzzy inference system is a mechanism that uses fuzzy logic and fuzzy set theory to map inputs to outputs. Due to the simplicity and effectiveness in representing and reasoning on human natural language, it has become to one of the most advanced technologies in control field. A typical fuzzy inference system consists of mainly two parts, a rule base (or knowledge base) and an inference engine. A number of inference engines have been developed, with the Mamdani method [1] and the TSK method [2] being the most widely used. Mamdani fuzzy inference method is more intuitive and suitable for handling human natural language inputs, which is an implementation of the extension principle [3]. As fuzzy outputs are usually led by the Mamdani approach, a defuzzification approach, such as the centre of gravity method [4], has to be employed to map fuzzy outputs to crisp values for general system use. The TSK approach however uses polynomials to generate the inference consequence, and it therefore is able to directly produce crisp values as outputs, which is often more convenient to be employed when the crisp values are required. Both of these traditional fuzzy inference approaches require a dense rule base by which the entire input domain need to be fully covered; otherwise, no rule will be fired when a given observation does not overlap with any rule antecedent.

Fuzzy rule interpolation (FRI), firstly proposed in [5], not only addresses the above issue, but also helps in complexity reduction for complex fuzzy models. When a given observation does not overlap with any rule antecedent value, fuzzy rule interpolation is still able to obtain certain conclusion, and thus improves the applicability of fuzzy models. FRI can also be used to reduce the complexity of fuzzy models by removing those rules that can be approximated by their neighbouring ones. A number of fuzzy rule interpolation methods have been developed in the literature, including [6, 7, 8, 9, 10], and have been successfully employed to deal with real world application, such as [11, 12]. However, all of existing FRI methods were developed on (sparse) Mamdani rule bases which lead to fuzzy outputs.

This paper proposes a novel extension of traditional TSK fuzzy model, which is not only able to deal with sparse TSK fuzzy rule bases, but also able to directly generate crisp outputs. To enable such extension, a new similarity degree measurement is proposed first to calculate the similarity degrees between given observations and each individual rule in the rule base. Dissimilar with the similarity measure used in the existing TSK approach, the introduced one leads to similarity degrees between the observation and others rule antecedents always greater than 0 even when they do not overlap at all. Then the TSK fuzzy model is extended using this new matching degree to obtain crisp inference results from sparse TSK fuzzy rule bases. The experiments show comparable result, which demonstrates the promising of the approach in enhancing the traditional TSK model when the knowledge represented in the rule base is not complete.

The rest of the paper is structured as follows. Section 2 introduces the theoretical underpinnings of TSK fuzzy inference model and measurement of similarity degrees. Section 3 presents the proposed approach. Section 4 details a set of experiments for comparison and validation. Section 5 concludes the paper and suggests probable future developments.

2 Background

In this section, the original TSK approach is briefly introduced, and the existing similarity measures are briefly reviewed.

2.1 TSK Fuzzy Model

The TSK fuzzy model was proposed by Takagi, Sugeno, and Kang in 1985 [2], and a typical fuzzy rule for the TSK model is of the following form:

$$\textbf{IF } u \; is \; A \; and \; v \; is \; B \textbf{ then } w = f(u,v) \;, \tag{1}$$

where A and B are fuzzy sets regarding antecedent variables x and y respectively, and $f(u,v)$ is a crisp function (usually polynomial), which determines the crisp value of the consequent. For instance, assume that a rule base for TSK model is comprised of two rules:

$$\begin{array}{ll} R_i : \textbf{IF} & x \; is \; A_i \; and \; y \; is \; B_i \textbf{ THEN } z = f_i(x,y) = a_i x + b_i y + c_i \\ R_j : \textbf{IF} & x \; is \; A_j \; and \; y \; is \; B_j \textbf{ THEN } z = f_j(x,y) = a_j x + b_j y + c_j , \end{array} \tag{2}$$

where a_i, a_j, b_i, b_j, c_i, and c_j are constants in the polynomial equation in the consequent part of the rule. The consequences of rules R_i and R_j deteriorate to constants $c_i$and c_j when $a_i = a_j = b_i = b_j = 0$. TSK model is usually employed to crisp inputs and outputs problems. Given an observation with singleton values as input (x_0, y_0), the working progress of this approach is demonstrated in Fig. 1. The inferred output from the given observation from rules R_i and R_j are $f_i(x_0, y_0)$ and $f_j(x_0, y_0)$ respectively. The overall output is then taken as the weighted average of outputs from all rules, where the values of weights are the firing strengths of corresponding rules. Suppose that $\mu_{A_i}(x_0)$ and $\mu_{B_i}(y_0)$ represent the matching degrees between input (x_0, y_0) and rules R_i and R_j, respectively. The firing strength (weight) of rule R_i, α_i, is calculated as:

$$\alpha_i = \mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0), \tag{3}$$

where $\wedge$ is a t-norm, which usually implemented as a minimum operator. Obviously, if a given input (x_1, y_1) does not overlap with any rule antecedent, the matching degree between this input and rules R_i and R_j are $\mu_{A_i}(x_1)$ and $\mu_{B_i}(y_1)$ are equal to 0. Then no rule will be fired. Then, no consequence can be derived for such case. As the final result of the consequent variable z is a crisp value, the defuzzification progress then can be saved, which in turn reduces the overall computational efforts.

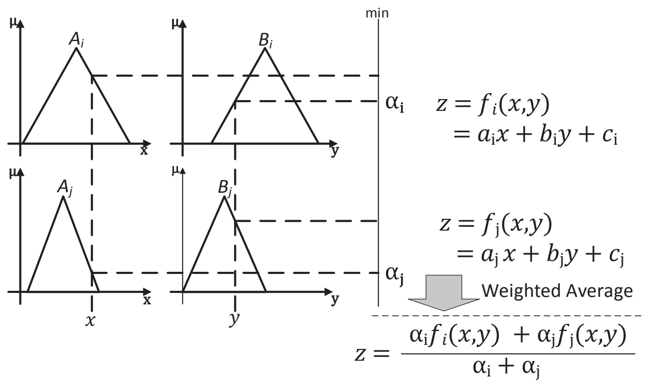


Fig. 1 Representation of TSK approach

2.2 Similarity Degree Measurement

Based on different measurement standards, various similarity measures have been proposed in literature to calculate the degree of similarity between two fuzzy sets, such as[13, 14, 15, 16]. Note that, in order to generate reasonable measurement of similarity, the corresponding variable domain is required to be normalised first. Given two triangle fuzzy sets on the variable with normalised domain, $A = (a_1, a_2, a_3)$ and $A' = (a'_1, a'_2, a'_3)$, where $0 \leqslant a_1 \leqslant a_2 \leqslant a_3 \leqslant 1$, and $0 \leqslant a'_1 \leqslant a'_2 \leqslant a'_3 \leqslant 1$, the degree of similarity $S(A, A')$ between fuzzy sets A and A' can be calculated as follows [13]:

$$S(A, A') = 1 - \frac{\sum_{i=1}^{3} |a_i - a'_i|}{3} \,. \tag{4}$$

The larger value of $S(A, A')$ means that is the more similar between fuzzy sets A and A'. This method is also the most wildly applied.

The above approach requires a normalisation progress for the concerned variable domain. A graded mean integration representation distance-based similarity degree measurement does not need such normalisation. This similarity measure is summarised as [17]:

$$S(A,A') = \frac{1}{1+d(A,A')} , \tag{5}$$

where $d(A,A') = |P(A) - P(A')|$, $P(A)$ and $P(A')$ are the graded mean integration representation of A and A', respectively [17]. In particular, the graded mean integration representation $P(A)$ and $P(A')$ can be defined as:

$$\begin{aligned} P(A) &= \frac{a_1 + 4a_2 + a_3}{6} , \\ P(A') &= \frac{a'_1 + 4a'_2 + a'_3}{6} . \end{aligned} \tag{6}$$

In this approach, the larger value of $S(A,A')$ means higher degree of similarity between fuzzy sets A and A'.

The above two approaches may not provide correct similarity degrees in certain situations, such as two generalised fuzzy sets (which are fuzzy sets may not be normal), although they are usually able to produce acceptable results and widely applied. A generalised triangle fuzzy set regarding variable x can be represented as $A = (a_1, a_2, a_3, \mu(a_2))$, where $\mu(a_2)$ ($0 < \mu(a_2) \leqslant 1$) is the membership of element a_2, and $\mu(a_2) \geqslant \mu(a), \forall a \in D_x$, D_x is the domain of variable x, as illustrated in Fig 2. If $\mu(a_2) = 1$, the generalised triangle fuzzy set deteriorates to normal a fuzzy set which is usually denoted as $A = (a_1, a_2, a_3)$. A centre of gravity method (COG) has been proposed to work with generalised fuzzy sets [15]. The process to calculate the COG-based similarity degree measure is summarised as below.

Step 1: Determine the point of centre of gravity for each triangle fuzzy set. Given a generalised triangle fuzzy set A, its COG $G(a^*, \mu(a^*))$ is shown in Fig. 2, which can be calculated by:

$$a^* = \frac{a_1 + a_2 + a_3}{3} , \tag{7}$$

$$\mu(a^*) = \frac{\mu(a_1) + \mu(a_2) + \mu(a_3)}{3} . \tag{8}$$

As $\mu(a_1) = \mu(a_3) = 0$, $\mu(a^*)$ can be simplified to:

$$\mu(a^*) = \frac{\mu(a_2)}{3} . \tag{9}$$

Step 2: Calculate the similarity degree $S(A,A')$ between fuzzy sets A and A' by:

$$\begin{aligned} S(A,A') = \left(1 - \frac{\sum_{i=1}^{3} |a_i - a'_i|}{3}\right) \cdot \left(1 - |a^*_A - a^*_{A'}|\right)^{B(Supp_A, Supp_{A'})} \\ \cdot \frac{min(\mu(a^*_A), \mu(a^*_{A'})}{max(\mu(a^*_A), \mu(a^*_{A'})} , \end{aligned} \tag{10}$$

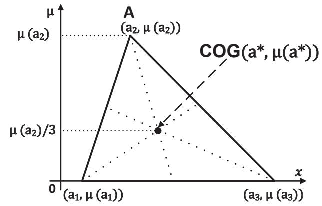


Fig. 2 A example triangular fuzzy set and its COG

where a^*_A and $a^*_{A'}$ are calculated by Equation 7, $\mu(a^*_A)$ and $\mu(a^*_{A'})$ are obtained from Equation 9, and $B(Supp_A, Supp_{A'})$ is defined as follow:

$$B(Supp_A, Supp_{A'}) = \begin{cases} 1,\ if\ Supp_A + Supp_{A'} \neq 0 \\ 0,\ if\ Supp_A + Supp_{A'} = 0\ , \end{cases} \tag{11}$$

where $Supp_A$ and $Supp_{A'}$ are the supports of the fuzzy sets A and A' respectively, which in turn are calculated as:

$$\begin{aligned} Supp_A &= a_3 - a_1\ , \\ Supp_{A'} &= a'_3 - a'_1\ . \end{aligned} \tag{12}$$

In the above equation, $B(Supp_A, Supp_{A'})$ is used to determine whether COG distance $(1 - |a^*_A - a^*_{A'}|)$ needs to be considered. For instance, if fuzzy sets A and A' both are the crisp values, (i.e., $Supp_A = Supp_{A'} = Supp_A + Supp_{A'} = 0$), the COG distance will not be considered for the degree of similarity measure; otherwise, the COG distance will be considered. In this measure, again the larger value of $S(A, A')$ means that the two fuzzy sets A and A' are more similar.

3 The Proposed Approach

The proposed fuzzy rule interpolation approach regarding TSK style of inference is introduced in this section. In order to enable the extension, the existing measure of similarity degree between two fuzzy sets as shown in Equation 10 is modified first by introducing an extra monotone decreasing function of the geometric distances between the two fuzzy sets. Given an observation, the similarity degree between the given observation and each rule antecedent can then be calculated based on this modified similarity degree measure, which always results a similarity degree greater than 0 even when the two fuzzy sets are not overlapped. Then, a crisp inference result can be obtained by considering all the rules associated with their corresponding matching degrees, based on underpinning principle of the original TSK inference.

As the similarity degrees between the given observation and all the rule antecedents are greater then 0, all the rules have firing strengths greater than 0, that is all rules are used for interpolation. Consequently, a crisp result can still be generated even when a given observation does not overlap with any rule antecedent.

3.1 A Modified Similarity Measure

The similarity measure expressed by Equation 10 may fail in certain situations, despite of its simplicity. For instance, if a large distance between two fuzzy sets presents, those two fuzzy sets should not similar at all. However, by applying this similarity measure, a big value of similarity degree, representing high similarity, may still be generated, which will lead to an unexpected result. In order to address this, the distance between fuzzy sets has been considered to extend this similarity measure [15]. However, the introduction of linear distance parameter may still not be sufficiently flexible to support various fuzzy models, as the sensitivity of similarity degree to distance is fixed. In order to provide a similarity measure whose sensitivity to distance is flexible and configurable to support fuzzy interpolation for TSK model, the similarity measure introduced in [15] is extended. In particular, the $distance\ factor$ (DF), which is an monotone decreasing function with an adjustable parameter, is proposed to replace the linear distance function of the existing approach ([15]).

Suppose the variable domain has been normalised, and assume that there are two generalised triangle fuzzy sets A and $A^{'}$ regarding this variable, where $A = (a_1, a_2, a_3)$, and $A' = (a'_1, a'_2, a'_3)$. The degree of similarity $S(A, A')$ between fuzzy sets A and A' can be calculated as follows:

$$S(A,A') = \left(1 - \frac{\sum_{i=1}^{3} |a_i - a'_i|}{3}\right) \cdot (DF)^{\tilde{B}(Supp_A, Supp_{A'})} \cdot \frac{min(\mu(a^*_A), \mu(a^*_{A'}))}{max(\mu(a^*_A), \mu(a^*_{A'}))} , \tag{13}$$

where DF, termed as $distance\ factor$, is a function of the distance between the two concerned fuzzy sets. DF is in turn defined as:

$$DF = 1 - \frac{1}{1 + e^{-nd+5}} , \tag{14}$$

where n ($n > 0$) is a sensitivity factor, and d represents the distance between the two fuzzy sets usually defined as the distance of their COGs. Smaller value of n leads to a similarity degree which is more sensitive to the distance of two fuzzy sets, and vice versa. The value of this factor needs to be determined based on the specific problems. However, some early stage experimentation generally suggests that $20 \leqslant n \leqslant 60$. A further study of the automatic determination of DF remains for the future work. It is worth to note that, there are two special situations where the modified similarity measure and Equation 10 lead to the same result: 1) when fuzzy sets A and A' have the same COG, and 2) when A and A' are two boundary crisp values and the distance between them is 1.

Compared to the approach proposed in [15], the modified similarity measure between two given fuzzy sets preserves the same set of good properties, including 1) The lager value is $S(A, A')$, the more similar are between fuzzy sets A and A', and 2) fuzzy set A and A' are identical if and only if $S(A, A') = 1$. The proposed approach also introduces one more important property, which is the similarity degree between any two fuzzy sets (excluding the two boundary crisp values that distance between them is 1) in the input domain will be always greater than 0. Without lose generality, given two fuzzy sets $A = (a_1, a_2, a_3)$ and $A' = (a'_1, a'_2, a'_3)$ within a normalised input domain. Suppose that fuzzy sets A and A' are not the boundary sets, where $0 < a_1 \leqslant a_2 \leqslant a_3 < 1$, and $0 < a'_1 \leqslant a'_2 \leqslant a'_3 < 1$. Then,

$$\begin{aligned} &|a_1 - a'_1| < 1 \ , \\ &|a_2 - a'_2| < 1 \ , \\ &|a_3 - a'_3| < 1 \ . \end{aligned} \tag{15}$$

This is followed by:

$$1 - \frac{\sum_{i=1}^{3} |a_i - a'_i|}{3} > 0 \ . \tag{16}$$

Also, $0 < DF < 1$ based on Equation 14, and $min(\mu(a^*_A), \mu(a^*_{A'})) > 0$. According to Equation 13, the value of $S(A, A')$ must be greater than 0.

3.2 Extending the TSK Model

For simplicity, this work only considers problems with two inputs and one output. A typical fuzzy rule for the original TSK fuzzy model is of the following form:

$$R_i: \ \textbf{IF} \ x \ is \ A_i \ and \ y \ is \ B_i \ \textbf{THEN} \ z = f_i(x, y), \tag{17}$$

where A_i, and B_i are fuzzy sets regarding input variable x and y, and $f_i(x, y)$ is a crisp function which determines the consequence. For a given observation, the original TSK approach first determines those rules whose antecedents overlap with the given observation, and then obtains the firing strength (α_i) of the overlapped rules by integrating the matching degrees between observation terms and rule antecedent terms. From this, the sub-consequence from each fired rule is computed using the consequent function. And finally, a crisp value of output ($f_i(x, y)$) is aggregated by calculating the weighted average of the sub-consequences, as introduced in Section 2.1. If a given observation does not overlap with any rule antecedent, no rule will be fired, and thus no inference can be made.

The above issue can be addressed by extending the original TSK approach using the similarity measure proposed in Section 3.1. Assume that a sparse rule base is comprised of n rules, which is:

$$\begin{aligned}
R_1 : \textbf{IF} \quad & x \ is \ A_1 \ and \ y \ is \ B_1 \ \textbf{THEN} \ z = f_1(x,y) = a_1 x + b_1 y + c_1, \\
& \dots \dots \\
R_i : \textbf{IF} \quad & x \ is \ A_i \ and \ y \ is \ B_i \ \textbf{THEN} \ z = f_i(x,y) = a_i x + b_i y + c_i, \\
& \dots \dots \\
R_n : \textbf{IF} \quad & x \ is \ A_n \ and \ y \ is \ B_n \ \textbf{THEN} \ z = f_n(x,y) = a_n x + b_n y + c_n,
\end{aligned} \tag{18}$$

where a_i, b_i, and c_i ($1 \leqslant i \leqslant n$) are constants of polynomials in rule consequences. When a input $O(A^*, B^*)$, alternatively termed as observation, is given, a crisp output can be generated by the following steps.

Step 1: Determine the matching degrees $S(A^*, A_i)$ and $S(B^*, B_i)$ between the input values (A^* and B^*) and rule antecedents (A_i and B_i) for each rule using Equation 13.

Step 2: Calculate the firing degree of each rule by integrating the matching degrees of its antecedents and the given inputs:

$$\alpha_i = S(A^*, A_i) \wedge S(B^*, B_i) \ , \tag{19}$$

where $\wedge$ is a t-norm, usually implemented by minimum operator in TSK inference model.

Step 3: Calculate the sub-consequence of the final result from each rule based on the given input $O(A^*, B^*)$ and the polynomial in consequent.

$$f_i(A^*, B^*) = a_i \cdot COG(A^*) + b_i \cdot COG(B^*) + c_i \ . \tag{20}$$

Step 4: Integrate the sub-consequences to get the final output:

$$z = \frac{\sum_{i=1}^{n} \alpha_i f_i(A^*, B^*)}{\sum_{i=1}^{n} \alpha_i} \ . \tag{21}$$

As discussed earlier, the similarity degree between any two fuzzy sets (excludes the two boundary sets) in the input domain is always greater than 0. Therefore, different from traditional TSK method, which only considers those rules overlapped with the given observation, the proposed approach takes account all rules in the rule base to aggregate a crisp result. As a result, even if the given observation does not overlap with any rule antecedent in the rule base, certain inference result is still able to be generated, which significantly improves the applicability of the original TSK approach.

4 Experimentation

In order to validate and evaluate the proposed approach, a non-linear function, which has been considered in [18], is re-considered in this section to demonstrate the functionality of proposed system. The problem is to model the non-linear function given below:

$$f(x, y) = sin\left(\frac{x}{\pi}\right) sin\left(\frac{y}{\pi}\right). \tag{22}$$

The fuzzy model takes two inputs, x ($x \in [-10, 10]$) and y ($y \in [10, 10]$), and produces a single output z ($z \in [-1, 1]$), as illustrated in Fig. 3. In order to enable the employment of the revised TSK style fuzzy rule interpolation, the input domains are normalised first. The normalisation maps any value x_0 of variable x to x_0' by:

$$x_0' = \frac{x_0 - \max_x}{\max_x - \min_x}, \tag{23}$$

where $\min_x$ is the minimum value in the domain of variable x, and $\max_x$ is the maximum value in the domain of variable x.

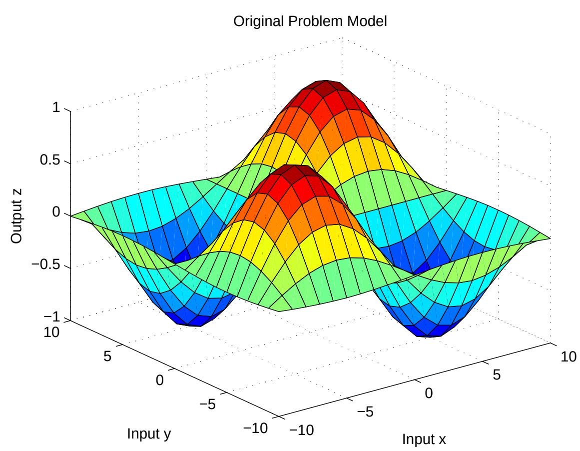
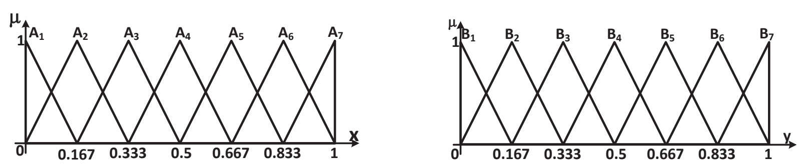


Fig. 3 Surface view of the model

In order to generate a optimal sparse TSK rule base for the model, a dense rule base is generated first. Then some of the less important rules are removed manually to demonstrate the working of the proposed TSK style fuzzy rule interpolation approach. The evaluation of the proposed approach based on incomplete data remains as active work.

4.1 TSK Rule Base Generation

A dense TSK fuzzy rule base was generated based on the given model first, by which the entire input domain is fully covered. In order to do so, a training data set comprised of 500 data points have been randomly generated from Equation 22. Then, a linear regression-based Matlab TSK rule base generation approach [19] was employed to derive a normal TSK fuzzy rule base that partitions each antecedent variable domain by 7 fuzzy sets. The surface view of fuzzy partition of TSK model is also illustrated in Fig. 5. As there are two input variables, this leads to 49 fuzzy rules in total, as listed in Table 1 and shown in Fig. 4. Briefly, the employed data-driven approach first grid partitions the given input domain into sub-regions. Then, for each sub-region, a linear regression approach, the least-squares approach, is employed to represent the data in an initial fuzzy rule. After that, linear quadratic estimation (Kalman Filter) algorithm is used to fine tune the rules' parameters until the satisfactory solution is found. The data-driven approach for TSK rule base generation is beyond the scope of this paper, and thus details are omitted here, however, more information can be found in [20].



(a) Normalised Input x (b) Normalised Input y

Fig. 4 Fuzzy partition of domain of input

4.2 Sparse TSK Rule Base Generation

The TSK rule base was then simplified to sparse rule base, by which some observations may not be covered by any rule antecedents in the rule base, to enable the evaluation of the proposed system. In this initial work, this progress was performed manually by removing some less important one, however, the study on sparse rule base generation or rule base simplification was left as future work. In particular, the size of the TSK rule base has been manually reduced 67%, which is comprised of

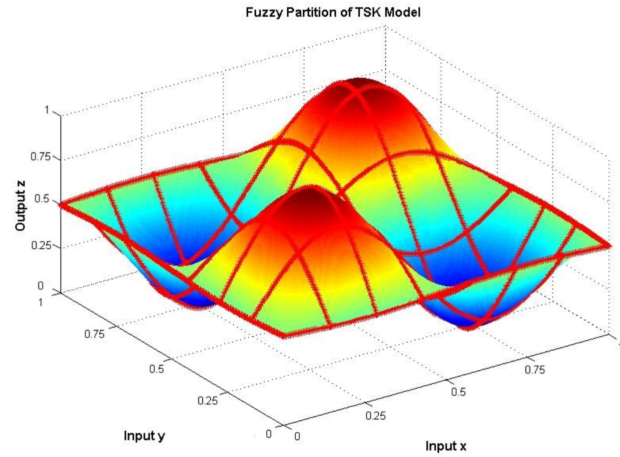


Fig. 5 Fuzzy partition for TSK modelling

Table 1 Generated TSK Rule Base

i	IF x	IF y	THEN z	i	IF x	IF y	THEN z
1	A_1	B_1	$0.315x + 0.249y + 0.501$	26	A_4	B_5	$1.967x + 0.164y - 0.472$
2	A_1	B_2	$1.589x + 0.112y + 0.494$	27	A_4	B_6	$1.165x - 0.098y + 0.087$
3	A_1	B_3	$1.366x - 0.075y + 0.543$	28	A_4	B_7	$-0.270x + 0.220y + 0.414$
4	A_1	B_4	$-0.296x - 0.139y + 0.566$	29	A_5	B_1	$0.064x - 1.463y + 0.442$
5	A_1	B_5	$-1.181x - 0.058y + 0.524$	30	A_5	B_2	$-0.605x - 0.437y + 0.526$
6	A_1	B_6	$-0.727x + 0.373y + 0.180$	31	A_5	B_3	$-0.486 + 1.347y - 0.010$
7	A_1	B_7	$0.491x + 0.551y - 0.033$	32	A_5	B_4	$0.327x + 2.060y - 0.629$
8	A_2	B_1	$0.188x + 1.693y + 0.485$	33	A_5	B_5	$0.720x + 0.930y - 0.116$
9	A_2	B_2	$0.568x + 0.757y + 0.710$	34	A_5	B_6	$0.607x - 0.492y + 0.841$
10	A_2	B_3	$0.571x - 0.859y + 1.052$	35	A_5	B_7	$0.374x - 0.750y + 0.802$
11	A_2	B_4	$-0.044x - 1.379y + 1.099$	36	A_6	B_1	$0.283x - 1.098y + 0.260$
12	A_2	B_5	$-0.252x - 0.400y + 0.337$	37	A_6	B_2	$0.879x - 0.361y - 0.468$
13	A_2	B_6	$-0.283x + 0.630y - 0.305$	38	A_6	B_3	$0.723x + 0.840y - 0.634$
14	A_2	B_7	$0.237x + 0.595y + 0.048$	39	A_6	B_4	$0.066x + 1.217y - 0.088$
15	A_3	B_1	$0.020x + 1.385y + 0.508$	40	A_6	B_5	$-0.832x + 0.785y + 1.012$
16	A_3	B_2	$-1.100x + 0.531y + 1.136$	41	A_6	B_6	$-0.115x - 0.073y + 0.889$
17	A_3	B_3	$-0.849x - 0.848y + 1.373$	42	A_6	B_7	$0.408x - 0.100y + 0.127$
18	A_3	B_4	$0.361x - 1.323y + 0.956$	43	A_7	B_1	$0.333x + 0.342y + 0.179$
19	A_3	B_5	$1.409x - 0.460y - 0.049$	44	A_7	B_2	$1.093x + 0.063y - 0.404$
20	A_3	B_6	$0.654x + 0.663y - 0.529$	45	A_7	B_3	$0.697x - 0.396y + 0.102$
21	A_3	B_7	$-0.375x + 0.741y + 0.037$	46	A_7	B_4	$-0.308x + 0.351y + 0.554$
22	A_4	B_1	$-0.009x - 0.280y + 0.504$	47	A_7	B_5	$-0.586x + 0.322y + 0.676$
23	A_4	B_2	$-1.736x - 0.004y + 1.262$	48	A_7	B_6	$-0.021x + 0.264y + 0.163$
24	A_4	B_3	$-1.341x + 0.359y + 0.958$	49	A_7	B_7	$0.232x + 0.112y + 0.229$
25	A_4	B_4	$0.502x + 0.384y + 0.084$				

23 rules: R_i, ($i = \{1, 3, 5, 7, 10, 11, 14, 17, 19, 21, 23, 25, 27, 30, 32, 34, 35, 39, 41, 43, 45, 47, 49\}$).

4.3 TSK Inference with Sparse Rule Base

To facilitate the comparison between the proposed approach and the approach proposed in [18], 36 testing data points were randomly generated by Equation 22 for testing and evaluation purpose. Note that although the considered problem in [18] was solved by Mamdani fuzzy model, it does not affect the result of the comparison as crisp results have been derived in this work using the defuzzification process.

To better illustrate the proposed approach, one randomly generated testing data point $O(A^* = 0.299, B^* = 0.441)$ was used as an example below to demonstrate the working progress of the proposed approach. The $distance\ factor$ in this experimentation is implemented as:

$$DF = 1 - \frac{1}{1 + e^{-20d+5}} \ . \tag{24}$$

From the given observation $O(0.299, 0.441)$, and the sparse rule base generated in Section 4.2, the proposed approach first calculated the similarity degree between the given input and rule antecedents ($S(A^*, A_i),\ S(B^*, B_i)$) ($i = \{1, 3, 5, 7, 10, 11, 14, 17, 19, 21, 23, 25, 27, 30, 32, 34, 35, 39, 41, 43, 45, 47, 49\}$) using Equation 13, with the results shown in the second and third columns of Table 2.

Based on the calculated similarity degree, the firing strength (FS) of each rule was calculated according to Equation 19, as shown in the fourth column of Table 2. From this, the sub consequence of the given observation from each rule was calculated by applying the observation to the linear function of rule consequence, as shown in the fifth column. The final result of variable (z) was then calculated by Equation 20, 21, which is $z = 0.566$ in this demonstration. Note that the ground truth of the consequence for the given observation is 0.478, then the error is 0.088. Using the same approach, the errors for other 35 testing points were also calculated. The reconstructed model based on the sparse TSK rule base with 23 rules (and $DF = 40$) is demonstrated in Fig. 6 for comparison.

4.4 Result Analysis

By following the testing design and error representation of work [18], the sum of errors for the 36 randomly generated testing data points with different parameters have been summarised in Table 3. Also, to enable comparison, experiments based on sparse rule bases with 41,39,36,23 rules have also been conducted, with the re-

Table 2 The Calculation of Similarity Degree

i	$S(A*, A_i)$	$S(B*, B_i)$	$FS(A*, B*)$	Consequence
1	0.404	0.038	0.038	0.027
3	0.404	0.806	0.404	0.370
5	0.404	0.480	0.404	0.059
7	0.404	0.003	0.003	0.001
10	0.772	0.806	0.772	0.652
11	0.772	0.850	0.772	0.369
14	0.772	0.003	0.003	0.001
17	0.866	0.806	0.806	0.601
19	0.866	0.480	0.480	0.082
21	0.866	0.003	0.003	0.082
23	0.581	0.246	0.246	0.0008
25	0.581	0.850	0.581	0.205
27	0.581	0.033	0.033	0.013
30	0.055	0.276	0.055	0.008
32	0.055	0.850	0.055	0.021
34	0.055	0.033	0.055	0.027
35	0.055	0.003	0.055	0.002
39	0.002	0.850	0.002	0.0007
41	0.002	0.033	0.002	0.001
43	0.001	0.038	0.001	0.0005
45	0.001	0.806	0.001	0.0002
47	0.001	0.480	0.001	0.0008
49	0.001	0.003	0.001	0.0005

sults also shown in Table 3. From this table, it is clear that the proposed system outperforms the system proposed in [18].

The experimentation results suggest that sparser rule bases always lead to large errors, which is consistent with the intuitive expectation. It also can be seen from the result table that the sensitivity factor (n) in $distance\ factor$ indeed affects the accuracy of the inference results. Based on the initial investigation through this experimentation, the system performs the best when the sensitivity factor is set to 40.

Table 3 Experimentation Results for Comparison

Numbers of Rules	**Proposed Approach**			**Approach in [18]**
	n=20	n=40	n=60	
41	3.27	**2.25**	2.41	2.1
39	3.24	**2.28**	2.41	3.1
36	3.29	**2.29**	2.42	5.5
23	3.36	**2.96**	2.99	6.0

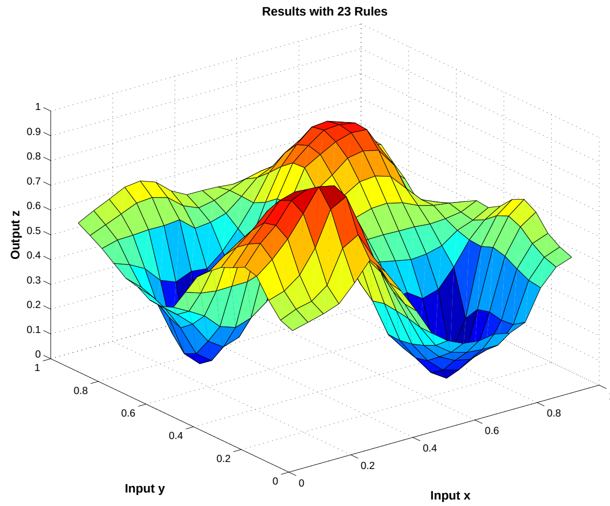


Fig. 6 Surface view of results based on 23 rules

4.5 Discussion

Although many FRI approaches have been proposed to enable fuzzy inference with sparse rule bases, they were all developed on the Mamdani fuzzy model. The proposed approach is the first attempt to extend this idea to TSK fuzzy inference such that inference can be performed based on sparse rule bases. This will therefore provide an additional alternative solution for those existing applications of FRI, such as [21] and in the same time to enjoy the advantages of TSK fuzzy inference. This will also enables extensions of existing FRI, such as the experience-based rule base generation and adaptation approach [21] to work with TSK inferences targeting a wider range of applications.

The rule base for traditional TSK fuzzy model, which used in this initial work, was generated by the linear regression algorithm, based on a randomly generated data set. Note that a recent development on sparse rule base updating and generating has been reported[22]. Although this approach was implemented on the Mamdani inference, the underpinning principle can be used to generate sparse TSK rule base. In particular, given a training data set, a sparse TSK rule base can be generated directly from data by strategically locating the important regions for fuzzy modelling [22], thus to boost the applicability of the proposed approach.

5 Conclusion

This paper presented a novel approach to extend TSK inference to work with sparse rule bases. This is enabled by generating a crisp inference result based on all the rules in the rule base rather than only those whose antecedents overlap with observations. In particular, the paper firstly proposed a new similarity degree measure by considering an extra $distance\ factor$ to obtain the similarity degree between the given observation and the corresponding rule antecedent of each rule. Then, based on the calculated degrees of similarity, all rules in the rule base will be considered with different firing strengths to generate a final crisp result. The experimentation shows that the proposed approach is not only able to deal with sparse TSK fuzzy rule bases, but is also able to generate competitive results in reference to the existing approach.

Although promising, the work can be further extended in the following areas. Firstly, the value of sensitivity factor n in $distance\ factor$ was arbitrarily given in this work based on some initial experimentation, and thus it would be worthwhile to further study how this parameter can be automatically determined or learned. Secondly, it is interesting to study if the curvature-based sparse rule base generation approach [22] can be used to support TSK rule base generation. Finally, it may be worthwhile, in further research, to investigate how the proposed approach can work with experience-based rule base generation [21].

References

1. E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis. *Computers, IEEE Transactions on*, C-26(12):1182–1191, 1977.
2. T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-15(1):116–132, 1985.
3. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - i. *Information Sciences*, 8(3):199 – 249, 1975.
4. Chuen Chien Lee. Fuzzy logic in control systems: fuzzy logic controller. ii. *Systems, Man and Cybernetics, IEEE Transactions on*, 20(2):419–435, 1990.
5. L. Kóczy and K. Hirota. Approximate reasoning by linear rule interpolation and general approximation. *International Journal of Approximate Reasoning*, 9(3):197–225, 1993.
6. Z. Huang and Q. Shen. Fuzzy interpolative reasoning via scale and move transformations. *Fuzzy Systems, IEEE Transactions on*, 14(2):340–359, 2006.
7. Z. Huang and Q. Shen. Fuzzy interpolation and extrapolation: A practical approach. *Fuzzy Systems, IEEE Transactions on*, 16(1):13–28, 2008.
8. L.Yang and Q. Shen. Adaptive fuzzy interpolation and extrapolation with multiple-antecedent rules. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–8, 2010.
9. L.Yang and Q. Shen. Adaptive fuzzy interpolation. *Fuzzy Systems, IEEE Transactions on*, 19(6):1107–1126, Dec 2011.
10. L.Yang and Q. Shen. Closed form fuzzy interpolation. *Fuzzy Sets and Systems*, 225:1 – 22, 2013. Theme: Fuzzy Systems.
11. J. Li, L. Yang, H. P. H. Shum, G. Sexton, and Y. Tan. Intelligent home heating controller using fuzzy rule interpolation. In *UK Workshop on Computational Intelligence*, 2015.

12. G.I. Molnarka, S. Kovacs, and L.T. Kóczy. Fuzzy rule interpolation based fuzzy signature structure in building condition evaluation. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 2214–2221, 2014.
13. S. M. Chen. New methods for subjective mental workload assessment and fuzzy risk analysis. *Cybernetics and Systems*, 27(5):449–472, 1996.
14. B Sridevi and R Nadarajan. Fuzzy similarity measure for generalized fuzzy numbers. *Int. J. Open Problems Compt. Math*, 2(2):242–253, 2009.
15. S. J. Chen and S. M. Chen. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on Fuzzy Systems*, 11(1):45–56, 2003.
16. L. Niyigena, P. Luukka, and M. Collan. Supplier evaluation with fuzzy similarity based fuzzy topsis with new fuzzy similarity measure. In *Computational Intelligence and Informatics (CINTI), 2012 IEEE 13th International Symposium on*, pages 237–244, 2012.
17. Shan Huo Chen and Chih Hsun Hsieh. Ranking generalized fuzzy number with graded mean integration representation. In *Proceedings of the Eighth International Conference of Fuzzy Sets and Systems Association World Congress*, volume 2, pages 551–555, 1999.
18. H. Bellaaj, R. Ketata, and M. Chtourou. A new method for fuzzy rule base reduction. *Journal of Intelligent & Fuzzy Systems*, 25(3):605–613, 2013.
19. S. Konstantin. *Sugeno-type FIS output tuning*, 2010. `http://www.mathworks.com/matlabcentral/fileexchange/28458-sugeno-type-fis-output-tuning`.
20. Babak Rezaee and MH Fazel Zarandi. Data-driven fuzzy modeling for takagi–sugeno–kang fuzzy system. *Information Sciences*, 180(2):241–255, 2010.
21. J. Li, H. P. H. Shum, X. Fu, G. Sexton, and L. Yang. Experience-based rule base generation and adaptation for fuzzy interpolationn. In *IEEE World Congress on Computation Intelligence Internation Conference*, 2016.
22. Y. Tan, J. Li, M. Wonders, F. Chao, H. P. H. Shum, and L. Yang. Towards sparse rule base generation for fuzzy rule interpolation. In *IEEE World Congress on Computation Intelligence Internation Conference*, 2016.