Static and Dynamic TSK Inference Systems Supported by Fuzzy Rule Interpolation

Pu Zhang

Supervisors: Prof. Qiang Shen
Dr. Neil Mac Parthalin

Ph.D Thesis

Department of Computer Science

Faculty of Business and Physical Sciences

Aberystwyth University

02, September, 2021

Declaration and Statement

Word Count of thesis: 21583

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Candidate name: Pu Zhang

Signature:

Date: 02, September, 2021

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where [1]**correction services** have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signature:

Date: 02, September, 2021

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signature:

Date: 02, September, 2021

[1]This refers to the extent to which the text has been corrected by others.

ABSTRACT

Fuzzy rule-based inference systems are successful representatives of knowledge based systems. Takagi-Sugeno-Kang (TSK) systems are one of the conventional and most exploited of such systems, providing an effective approach for performing prediction and regression tasks. However, when the input domain is not fully covered, it is possible for an observation to match no rule in the given rule base. In such a case, no conclusion can be drawn when using traditional rule-firing mechanisms. Fuzzy rule interpolation (FRI) has been introduced to deal with this problem. Whilst offering a potentially powerful inference mechanism, most existing FRI methodologies in the current literature are not developed for TSK inference models. In addition, several that are relevant may introduce an undesirable bias into the results while incurring significant computational overheads.

Motivated by the above observation, this thesis presents a novel FRI approach through the use of comparatively few neighbour rules to derive interpolative results with TSK models. Compared with existing methods, the proposed approach helps reduce the computational overheads of the inference process while avoiding the adverse impact caused by rules with low similarity to the new, unmatched observation. More importantly, to deal with large-sized sparse rule bases, where neighbourhood rules may be similar to each other, a rule-clustering-based method is introduced. In particular, a clustering algorithm is first employed to cluster rules into different groups, and the final interpolated conclusion is computed using the closest rules selected from a small number of closest rule clusters. The efficacy of the proposed approach is verified via systematic experimental examinations in comparison with existing methods, over a range of benchmark regression problems, whilst utilising different clustering algorithms.

Particularly, to verify and demonstrate its potential, the proposed FRI approach is systematically adopted to conduct missing value imputation tasks for a real-world dataset, the water treatment plant dataset, in which features with missing values are regarded as the output of the system while others are used as inputs. The success of such a realistic application indicates the practicality of the proposed interpolation approach.

In more real-world applications, the inputs are usually time-dependent, thereby requiring dynamic management of the rule base to maintain and possibly improve the efficacy of such a system. Situations may become more complicated if the training data does not sufficiently cover the problem space. FRI systems may help, whilst most of them follow a static approach, tending to process a large number of interpolated rules which are generally discarded once the results have been derived. Nonetheless, the interpolated rules may contain potentially useful information. This thesis presents a dynamic TSK system by exploiting such rules to support subsequent inference and promote rule bases. The obtained intermediate rules supplement the initial rule base, allowing it to expand to a larger set size. Afterwards, a clustering algorithm is employed to categorise rules into different groups so that an interpolated conclusion can be computed using the closest rules selected from a small number of the nearest rule clusters. Through systematic experimental comparisons with the conventional static approach, the proposed dynamic TSK system not only improves the overall reasoning accuracy but also reduces interpolation overheads by avoiding the need to interpolate similar observations.

Acknowledgements

Firstly, I would like to give my sincere gratitude to my first supervisor, Prof. Qiang Shen, for his invaluable supervision and stimulating encouragement throughout this project. What I learned is not only his research skills, but also his rigorous attitude toward research.

Then, thanks must also go to my second supervisor, Dr. Neil Mac Parthalin, for his patient guidance during the research.

In particular, I would like to express my gratitude to Dr. Changjing Shang, for her advice and feedback, which has helped improve my work significantly.

I am also extremely grateful for my friends and all fellow researchers in the Advance Reasoning Group for their support in my overseas study and daily life.

Finally, my sincere gratitude go to my beloved parents, Hongjian Zhang and Qingying Li, for their love and the unwavering support through all these years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The field of fuzzy logic arose from Lotfi A. Zadehs publication (Zadeh, 1965), which introduced the fuzzy set theory in detail. In recent decades, fuzzy logic has developed rapidly and has been widely utilised to reinforce artificial intelligence systems, including clustering methodologies (e.g., (Bezdek *et al.*, 1984); (Zhang & Chen, 2004); (Fan *et al.*, 2003); (Graves & Pedrycz, 2010); (Jiang *et al.*, 2015)) and fuzzy expert systems (e.g., (Mamdani & Assilian, 1975); (Takagi & Sugeno, 1985); (Jang, 1993); (Kasabov & Song, 2002); (Castillo & Melin, 2012)). It has also been successfully applied to solving various real-world problems (e.g., (Ross, 2004); (Asai *et al.*, 1994); (Bo & Lai, 2011); (Alcala-Fdez & Alonso, 2016)).

One of the central topics of the fuzzy set theory is expressing uncertain information and linguistic terms that are difficult to describe using precise mathematical language, such as fast and slow, and young and old, which makes it well known and widely applied. Fuzzy expert systems are a type of popular inference systems developed on the basis of fuzzy theory, enabling the inference process to resemble human activity and reasoning intuition by exploiting their potential for imprecision, partial truth and approximations. Fuzzy rule-based systems are one of the most commonly used such inference mechanisms.

1.1 Fuzzy Rule-Based Systems

Rule-based systems provide the computational mechanisms that can be found in most expert or knowledge-based systems (Kurfess, 2003). The knowledge bases are represented as a collection of rules that are typically expressed as if-then clauses, and the inference method for deriving new conclusions from existing knowledge follows the manner of *modus ponens* (Kovarik, 2009) (Liao, 2005) (Grosan & Abraham, 2011). It can be summarised as that if the input observations match rule antecedents, then the outputs can be obtained as the corresponding rule consequents. However, conventional crisp rule-based systems may fail to generate satisfactory results while involving uncertain information and linguistic terms that are hard to describe precisely.

Fortunately, with the support of fuzzy logic and fuzzy set theory, fuzzy rule-based systems (FRBSs) allow all such terms to be represented by fuzzy sets, enabling the inference process to resemble human reasoning. The inference mechanism of most FRBSs follows the compositional rule of inference (CRI) principle (Zadeh, 1973). That is, if the system input coincides with the antecedent of a fuzzy rule, then the output should coincide with the consequent that corresponds to the antecedent of that fuzzy rule (Scherer, 2012). This is different from crisp rule-based systems, where in general, only one rule will be fired for each instance, as the FRBSs allow an instance to match with multiple rules at the same time, as long as the instance overlaps with corresponding rule antecedents. Various fuzzy rule inference systems have been developed in the literature based on the law of CRI. Mamdani models (Mamdani & Assilian, 1975) and TSK models (Takagi & Sugeno, 1985) are two conventional systems and are the most widely used. Mamdani models use fuzzy sets as rule antecedents and consequents. A defuzzification process is usually required to obtain crisp results in practice. In TSK models, fuzzy sets are used as rule antecedents and polynomials as the consequents, whose application directly results in crisp conclusions. As such, TSK models are particularly suitable for solving regression and prediction problems in continuously valued domains.

Although generally powerful, CRI-type FRBSs all suffer from an important limitation: the dense rule base is generally required. FRBSs may fail to draw a reasonable conclusion for an unknown observation when the rule base is not dense but sparse. The observation may not overlap with any of the given rules. Thus, no rule can be fired to produce the corresponding conclusion by directly applying CRI. In real-world tasks, sparse rule bases are a common issue whereas dense ones are difficult to acquire (Baranyi *et al.*, 1999).

1.2 Fuzzy Rule Interpolation

FRI has been developed to rectify or at least to reduce the adverse impact of the key limitation of CRI-type FRBSs. It provides an approximate reasoning mechanism when only sparse rule bases are available. As CRI-type FRBSs fundamentally require that the unknown observations must match at least one fuzzy rule in the rule base to draw reasonable and accurate conclusions, the input domain must be fully covered by the rule antecedents in the given rule base. Otherwise, it is possible that an observation may 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 termed sparse (although this is often taken to simply imply an incomplete rule base) in the literature. The dense and sparse rule bases considered in this thesis refer to the coverage of the input domain by the antecedents of rules rather than the quantity and density of rules in a given rule base.

The tomato classification problem (Mizumoto & Zimmermann, 1982) (Kóczy & Hirota, 1993) is a typical example in this field. The ripeness–colour relation rule base consists of two rules:

- *If a tomato is green, the tomato is unripe.*
- *If a tomato is red, then the tomato is ripe.*

For the new observation, a yellow tomato, which matches none of the given rules, the conventional fuzzy inference will collapse and can not provide any result.

The resolution of real-world cases is typically undertaken using incomplete rule bases, given that denser ones are difficult to acquire. (Baranyi *et al.*, 1999) and (Tikk & Baranyi, 2000a) introduced several reasons leading to the use of such incomplete rule bases. The two most common reasons are as follows:

1. The knowledge base for fuzzy rule generation is incomplete at first, because of the limitation of human expertise or machine learning techniques.
2. Simplification on fuzzy rule bases, to reduce computational overheads.

FRI introduces an underlying mechanism different from CRI and many of its derivatives to deal with these issues. When an observation does not overlap with any rule antecedent, FRI helps generate an intermediate rule by approximating neighbouring rules for the observation to obtain a potentially relevant conclusion (Kczy & Hirota, 1993) (Kczy & Hirota, 2015). FRI makes an essential breakthrough in FRBSs to address the key limitation, that is, the requirement of a dense fuzzy rule base that fully covers the entire input domain.

FRI contributes significantly to the development of FRBSs by enabling reasoning in situations where only incomplete rule bases are available. It also provides the potential for a reverse application: simplifying over dense fuzzy rule bases that suffer overlap redundancy (Koczy *et al.*, 1997). This can usually be achieved by two distinct implementations: (i) replacing two existing rules with the interpolated one, iteratively; (ii) eliminating those fuzzy rules that can be approximated from their neighbours (Li *et al.*, 2021). Although it is rarely found in the majority of FRI methodologies developed in the literature, over dense rule base simplification may play an important role in future research involving big data.

With the support of FRI, FRBSs have greatly improved its ability and efficacy in resolving practical problems. FRI has helped to reinforce the performance of FRBSs on practical pattern recognition problems, including classic classification and prediction problems (Li *et al.*, 2018a) (Li *et al.*, 2019) (Chen & Chen, 2016) (using weighted FRI techniques); computer vision and image super-resolution (Yang *et al.*, 2021); and disease diagnosis, especially mammographic mass risk analysis (Li *et al.*, 2020) and colorectal polyp detection (Nagy *et al.*, 2018). The work on dynamic FRI (Naik *et al.*, 2014) (Naik *et al.*, 2017b) provides promising solutions to cyber-security problems, such as network security analysis, intelligent intrusion detection (Naik *et al.*, 2017b), and firewall reinforcement (especially for Microsoft Windows firewall) (Naik *et al.*, 2017a). FRI also has impressive applications in systems control, such as the simulation of automated guided vehicles (Kovcs-Lszl & Kczy, 1999), surveillance navigation control of mobile robots (Vincze & Kovacs, 2008), and behaviour based control (Kovcs & Kczy, 2004).

1.2.1 Current Challenges and Motivations

A good number of FRI approaches have been established over the past few decades. Just considering the popular family of transformation-based fuzzy rule interpolation (T-FRI) techniques (Huang & Shen, 2008) that generally follow the seminal work on linear interpolation (Kóczy & Hirota, 1993), there have been many distinct FRI mechanisms reported in the literature, including adaptive interpolation (Yang & Shen, 2011), higher-order interpolation (Chen *et al.*, 2016), and weighted FRI techniques (Li *et al.*, 2018a). However, these exemplified techniques are all developed for Mamdani models rather than for TSK models. Therefore, a novel approach suitable for performing FRI for TSK models under different circumstances is worth investigating.

In real-world data mining applications, missing values are a common problem that may lead to lack of data integrity and reduce the reliability of subsequent developments. Missing value imputation has the capability to resolve such issues. However, existing imputation approaches have their own shortcomings. For example, statistic imputation, such as mean and median imputation, ignores the relationship between attributes, and linear regression imputation can deal only with linear relations, not nonlinear ones. TSK inference systems have the potential to simulate a nonlinear relationship between features, but how they may be adopted to resolve missing value imputation tasks remains a challenge.

In real-world applications of fuzzy systems, the inputs are usually time dependent, and the requirements may change over time. If the frequently appearing unmatched observations are highly similar, the use of a static rule base (one that does not change over time) will repeat similar work and affect the efficacy of fuzzy inference systems. Therefore, dynamically maintaining the rule base is required in an effort to greatly improve the efficacy of the system concerned through enhanced coverage of the rule base.

1.3 Main Contributions within Thesis

Motivated by the above observation, this thesis makes four major contributions to the FRI literature: (i) a novel FRI approach for TSK models, interpolating with a small number of closest neighbour rules, derives reasonable interpolative conclusions where a sparse rule base is present; (ii) the proposed approach provides a clustering-aided implementation to resolve the problem of lacking diversity of rules being involved in subsequent interpolation where a large-sized sparse rule base is present; (iii) a novel missing value imputation approach based on TSK inference systems, which provides more accurate estimated values and helps improve the accuracy of subsequent classification systems; (iv) a novel framework of dynamic TSK systems supported by the proposed FRI approaches provides a novel approach to make dynamically maintaining rule bases possible. The proposed systems significantly reduce computational overheads and enhance the coverage of rule bases.

1.3.1 Static TSK Models Supported by FRI with K-Neighbours

TSK inference extension (TSK+) (Li *et al.*, 2018b) is a recently proposed fuzzy interpolative reasoning method aimed at fuzzy systems employing TSK models. Instead of relying on computing the matching degrees, it uses a distance metric-based similarity measure to perform interpolative reasoning, by manipulating all rules that the rule base contains. In so doing, when an input matches no rule, a certain output is still obtained. Although TSK+ offers a useful means for innovative inference, it has its own shortcomings. Particularly, it is not sufficiently efficient for many practical applications because it fundamentally requires the use of all given rules, thus incurring significant computational overheads. Besides, redundant or even possibly irrelevant rules are also included in any attempt to compute the output. This may well introduce undesirable bias into the final interpolated outcomes, thereby reducing accuracy of the system.

Whilst running TSK+, it is easy to reveal that those rules nearest to an unmatched observation generally have a much higher similarity degree than others. Inspired by this observation, a novel approach, interpolation with (just) K closest neighbours, is proposed herein that provides two novel implementations: (i) interpolation with K closest rules (KCR) for sparse rule bases of a small size, and (ii) K closest rule clusters (CRC) for large-sized sparse rule bases. The underlying principle for both is to perform interpolation with K closest neighbouring rules where K is normally a small number. In so doing, the adverse impact caused by rules with low similarities can be avoided while reducing computation. Furthermore, the problem of lacking diversity of rules being involved in subsequent interpolation, which is caused by the situation where a large-sized sparse rule base is present but the K closest rules may be very similar with each other, can be resolved. To ensure the proposed approach does not rely on any specific clustering algorithm, the proposed implementation for the second method is systematically evaluated using five clustering techniques.

1.3.2 Missing Value Imputation by Fuzzy Inference Systems

This thesis presents a missing value imputation approach supported by TSK models for the water treatment plant dataset. The mechanism of this approach is to establish a TSK system, the rule base of which is derived by regarding the variable with missing values as the output variable and others as input variables. In a situation where no rules within the obtained rule base match, the interpolation technique is utilised to approximate the missing value. Systematic experiments demonstrate that the proposed imputation approach can produce accurate outcomes and help improve the accuracy of subsequent classification systems.

1.3.3 Dynamic TSK Systems Supported by FRI

To design a dynamic TSK system, additional information is necessary. Fortunately, FRI offers such potential due to the fact that most existing FRI systems tend to produce a large amount of interpolated rules over time, which are generally discarded once the results have been derived. Exploiting these interpolated rules may help update the original sparse rule base, thereby constructing a dynamic system.

Many approaches make the creation of a real-time rule base possible, including adaptive fuzzy control (Astrom & Witternmark, 1995) (Mohan & Bhanot, 2006) (Zhang & Bien, 2000) and optimisation-based fuzzy rule generation (Wu *et al.*, 2001) (Angelov & Buswell, 2003) (Angelov, 2003). Unfortunately, all these techniques are developed for dense (fully covered) rule bases. They can not be applied to sparse rule bases directly because of the inherent pattern-matching mechanisms they use, given that no conclusion can be drawn when an observation does not (partially or fully) match any of the rules in the rule base. Dynamic fuzzy rule interpolation (D-FRI) (Naik *et al.*, 2017b) provides a novel methodology to exploit the interpolated rules generated by FRI. However, due to it is specifically designed for Mamdani-type models (whose consequents are fuzzy sets), this approach can not be directly applied to TSK models (whose consequents are polynomials). To construct a dynamic TSK model, the corresponding interpolation method and rule promoting process for polynomial consequents are required.

In extending the underlying ideas of the aforementioned FRI approach and D-FRI, a dynamic TSK system is presented in this thesis. Particularly, when running on a sparse rule base of a small size, the outcomes with respect to unmatched observations are inferred by interpolating a small number of closest rules. The interpolated rule is then constructed to a new rule and directly added to the original sparse rule base. When the number of rules, be they original or interpolated, reaches a certain threshold, the rules are clustered first and the corresponding conclusions are computed using the closest rules selected from a small number of closest rule clusters. The interpolated rules are then, also integrated into the rule base.

1.4 Thesis Structure

The rest of this thesis is structured as follows (with an indication of directly relevant, peer-reviewed publications produced as a result of this research):

Chapter 2: Literature Review

This chapter reviews the preliminary knowledge that is relevant to this project, including a summary of several existing static and dynamic fuzzy rule inference systems supported by FRI methodologies; an overview of different types of clustering algorithms, especially fuzzy c-means and its extensions; and an outline of conventional missing value imputation methodologies.

Chapter 3: FRI with K-Neighbours for Static TSK Models

This chapter presents a novel FRI approach with two implementations suitable for performing fuzzy rule interpolation with static TSK fuzzy inference models: (i) KCR for sparse rule bases of a small size, and (ii) CRC for large-sized sparse rule bases. The key principle for both is to perform interpolation using a small number of distinctive rules closest to an unmatched input. In addition, rule weight determination approaches related to interpolative reasoning are also discussed in this chapter. The content of this chapter have been accepted for publication by *IEEE Transaction on Fuzzy Systems* (Zhang *et al.*, 2021).

Chapter 4: Experimental Investigation of FRI for Static TSK Models

This chapter describes the setting of the experiments carried out and discusses the results of comparative experimental evaluations. Several rule weight determination approaches are systemically evaluated over benchmark datasets at first, and the overall winner is applied in subsequent investigations. Afterwards, to ensure the proposed approach does not rely on any specific clustering algorithm, several popular clustering algorithms are used to perform rule clustering. Systematic comparative experimental studies demonstrate the effectiveness and robustness of both implementations.

Chapter 5: Missing Value Imputation for Water Treatment Plant Dataset

This chapter proposes a missing value imputation approach for the water treatment plant dataset. The TSK model is used to establish the relationship between the variables with and without missing values. The proposed imputation technique will be evaluated based on the accuracy of obtained outcomes and the accuracy of subsequent classification systems. The content of this chapter is under peer-review for journal publication.

Chapter 6: Dynamic TSK Systems Supported by Fuzzy Rule Interpolation

This chapter presents an initial investigation on a novel dynamic TSK inference system suitable for working with sparse rule bases. This system facilitates the dynamic maintenance of rule base and performing on real-world applications whose requirements may change over time. FRI approaches provide such potential because of the large number of interpolated rules produced in their reasoning process. Experimental results demonstrate that compared with the existing static TSK inference approaches, the proposed systems increase the overall reasoning accuracy and decrease interpolation overheads by avoiding the need for interpolating observations similar to those experienced. The content of this chapter has been published in (Zhang & Shen, 2020) and selected as one of the finalists in WCCI 2020's Best Student Paper Awards.

Chapter 7: Conclusion

This chapter summarises the work contained in this thesis, along with interesting suggestions for future research.

Appendices

Appendix A lists the publications arising from the work presented in this thesis, including three journal articles and three conference papers that have been published or are under review for publication. Appendix B summarises the abbreviations employed in this thesis.

The guidance for assisting the reading of this thesis for the following chapters (bar Appendices) is given in Fig. 1.1.

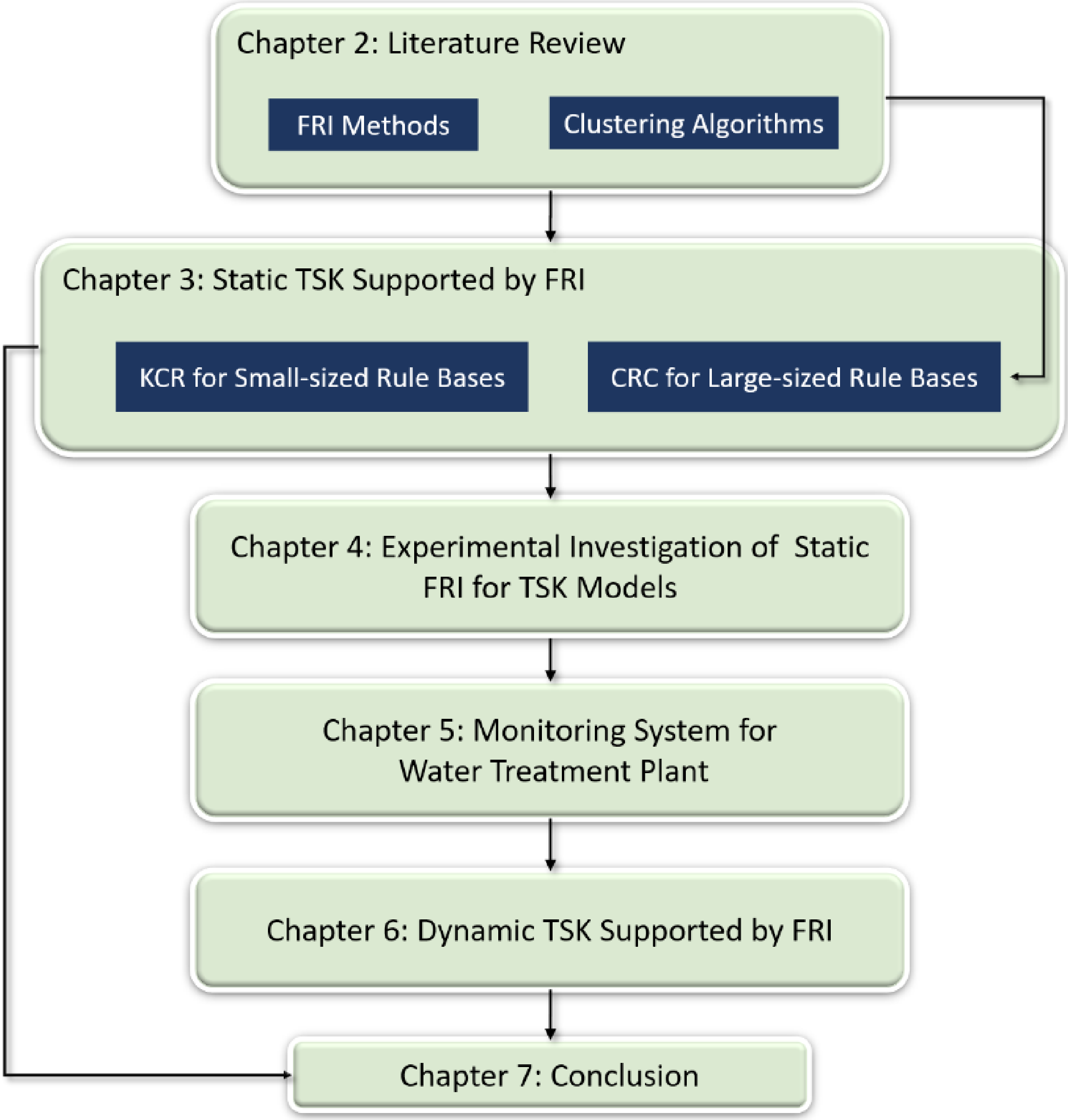

Figure 1.1: Guidance for thesis reading.

Chapter 2

Literature Review

This chapter reviews the preliminary background knowledge which is relevant to the research of the subsequent chapters, including TSK fuzzy models, static and dynamic fuzzy inference systems supported by FRI methodologies, clustering algorithms, and missing value imputation techniques. Fig. 2.1 provides a tree structure for assisting the reading of the literature review.

2.1 Fuzzy Rule-Based Inference Systems

The Mamdani (Mamdani & Assilian, 1975) and TSK (Takagi & Sugeno, 1985) fuzzy models are the two most widely applied fuzzy rule-based systems following the law of CRI.

TSK models use fuzzy sets as rule antecedents and polynomials as rule consequents while the antecedents and consequents of Mamdani models are both represented by fuzzy sets. The inference process of the Mamdani and TSK model are respectively exemplified in Fig. 2.2 and Fig. 2.3.

In these two figures, (A_i, B_i), (A_j, B_j) stand for the antecedents of rules R_i, R_j; C_i, C_j mean the consequents of R_i, R_j in Mamdani models; $f_i(x, y)$, $f_j(x, y)$ denote the consequents of R_i, R_j for TSK models; α_i, α_j are the weights

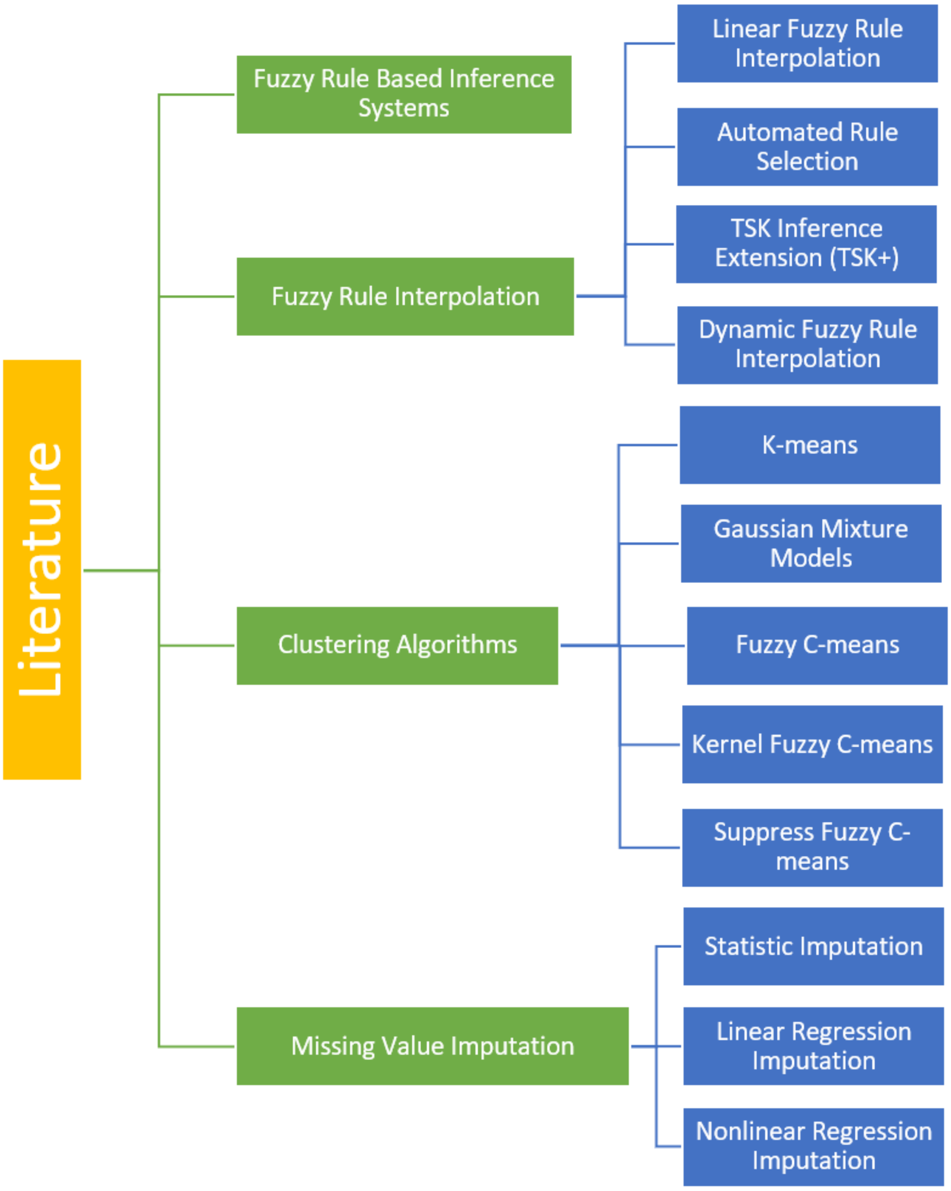

Figure 2.1: Structure of literature review.

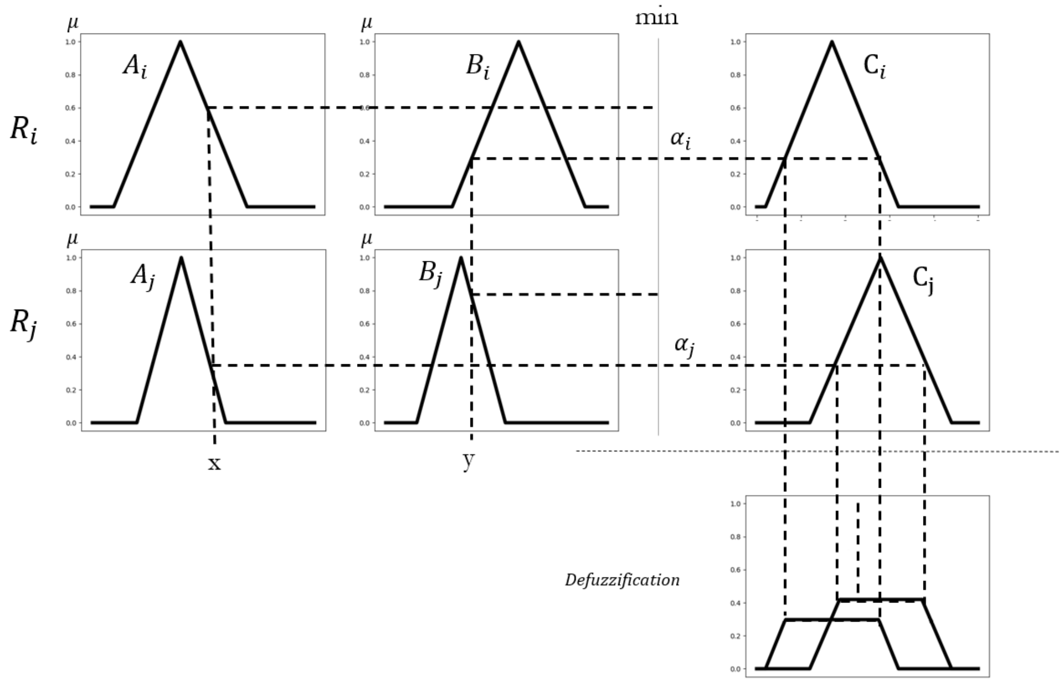

Figure 2.2: Example of inference process with Mamdani model.

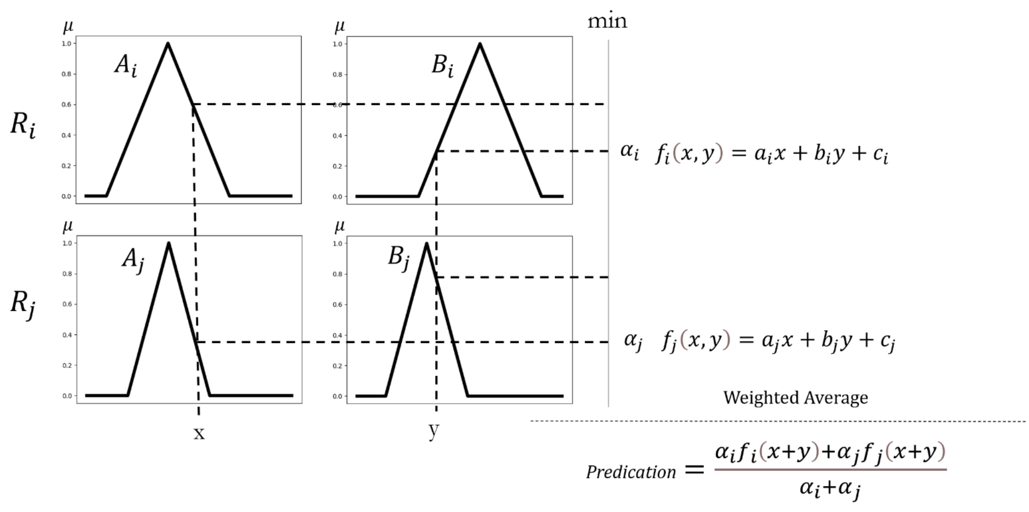

Figure 2.3: Example of inference process with TSK model.

of the corresponding rules; and (x, y) represents the observation. The observation (x, y) are firstly matched with the corresponding rule antecedents (A, B) with different membership degrees, the minimal membership will be regarded as the weight of the rule. In Mamdani model, a defuzzification process is usually required to obtain crisp results in practice. In TSK models, the weighted average method is usually employed to integrate sub-conclusions directly into crisp conclusions, which makes it more applicable for solving regression problems.

As presented in Fig. 2.3, TSK models perform inference for a novel observation by integrating conclusions generated by its overlapping rule, which is generally implemented by the weighted average. The weight of these overlapping rules is determined by an integrating operation (usually implemented by a minimum operator) on the matching degrees between the antecedent variables of the observation and their counterparts in each rule.

Particularly, rule antecedents in FRBSs are usually represented by the fuzzy sets. Distinguish with conventional Boolean sets which have crisp boundaries (an instance can only belong or not belong to a certain group), fuzzy sets use the membership function to describe the partial degrees to which an instance is considered to belong to a certain category.

Membership functions have two major categories in terms of their properties in practice (Li *et al.*, 2021), which are:

- Polygonal (piecewise linear) functions, including triangular-shaped, trapezoidal-shaped, hexagonal-shaped MF, etc. Generally, they are defined by their characteristic points in ascending order (the odd points of the corresponding mathematical function) (Huang & Shen, 2008).
- Nonlinear functions, typically including Gaussian, generalised bell-shaped, and Sigmoid function. They are usually represented by the parameters

that specify each nonlinear function (e.g., mean and variance for Gaussian function).

Theoretically, the universe of membership functions is infinite, as any convex and continuous function can be adopted as a membership function. It is a relief to note that only a few of them are typical and extensively used. Triangular and trapezoidal functions are the two most popular ones in the literature amongst the family of membership functions (Li *et al.*, 2021). Moreover, as empirically proven, if the membership functions can be appropriately fine-tuned, the use of different types of membership functions has little impact on the fuzzy rule-based inference results (Li *et al.*, 2018b) (Chang & Fan, 2008). Therefore, in this thesis, triangular membership functions are applied to describe fuzzy sets and construct all FRI methods unless otherwise stated for illustrative and demonstrative consistency and simplicity, as well as for their popularity in the literature.

Fig. 2.4 illustrates a normal and convex triangular membership function as an example. The fuzzy set A can be described by the three characteristic points as $A(a_1, a_2, a_3)$. $\mu(\mu \in [0, 1])$ represents that in which membership degree the instance belongs to the corresponding category. It should be noticed that although $\mu = 0$ denotes that the instance does not belong to the fuzzy set, the instance with $\mu = 1$ may still belong to other fuzzy sets with a different μ.

Suppose that a TSK sparse fuzzy rule base is given, containing m rules with n antecedent variables, each rule is then defined by

$$R_i: \textbf{ if } x_1 \text{ is } A_{i1}, ..., x_n \text{ is } A_{in}, \textbf{ then } f_i(x_1, ..., x_n) = a_{i0} + a_{i1}x_1 + ... + a_{in}x_n \tag{2.1}$$

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

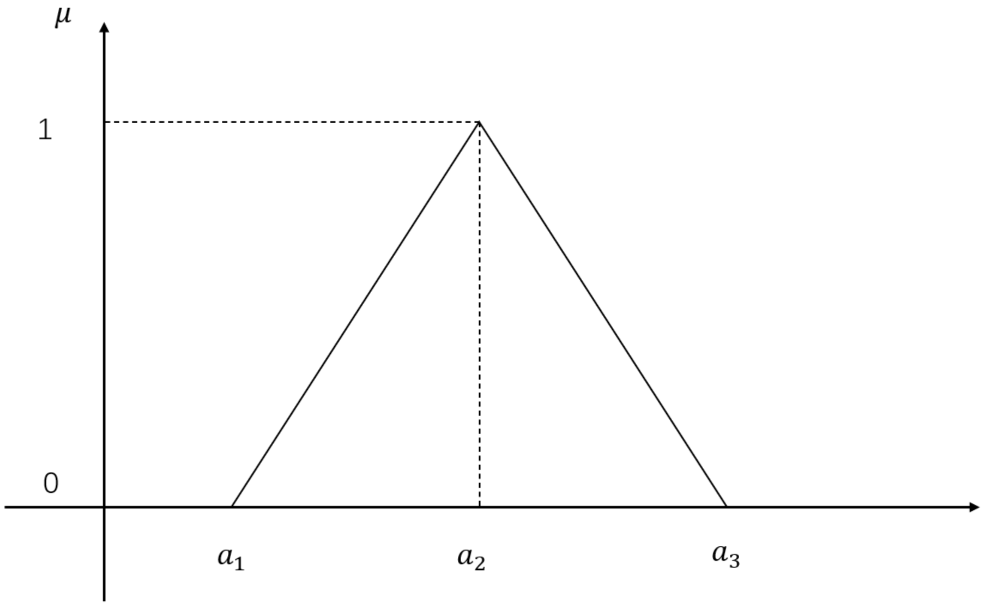

Figure 2.4: Fuzzy set represented by normal and convex triangular membership function.

consequent. Given an observation $O(B_1, ..., B_n)$, the detailed description of the TSK model can be found in Alg. 1.

2.2 Fuzzy Rule Interpolation

As indicated previously, CRI-type FRBSs require the input domain to be fully covered by the given rule base. While only sparse fuzzy rule bases are available, a given observation may match no rule and the weight of each rule α_i will be 0. Thus, neither sub-conclusion nor final result can be generated. FRI has been developed to produce conclusions for unmatched observations by exploiting the approximation of their neighbouring rules. What follows is a summary of the most common and popular interpolation technics, upon which this work is built.

Algorithm 1: Inference with TSK models

Input: Rule base $\{R_i\}$; Observation O

Output: Conclusion $f(O)$

1. Calculate matching degrees between antecedent variables of observation O and their counterparts in each rule R_i:

   $D(A_{i1}, B_1), ..., D(A_{in}, B_n)$

2. Determine 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 minimum.

3. Take observation O as input to compute rule consequent polynomial for each of 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 final outcome for consequent by weighted average:

$$f(B_1, ..., B_n) = \frac{\Sigma_{i=1}^{k} \alpha_i f_i(B_1, ..., B_n)}{\Sigma_{i=1}^{k} \alpha_i} \quad (2.2)$$

5. Return: $f(B_1, ..., B_n)$.

2.2.1 Linear Fuzzy Rule Interpolation

Linear fuzzy rule interpolation (Kóczy & Hirota, 1993), also known as the KH interpolation method, is the very first method to perform fuzzy interpolative reasoning on sparse rule bases through the manipulation of α-cut distances. It provides the initial principle for fuzzy rule interpolation reasoning. That is:

The closer a rule antecedent is to the observation, the closer the rule consequent is to the predicted conclusion corresponding to the observation.

Suppose that the two fuzzy rules is given:

- *If X is A_i then Y is B_i*
- *If X is A_j then Y is B_j*

and A^* is an observation, which have no overlap with A_1 and A_2, and B^* is the conclusion to be interpolated. Then, the notion of linear fuzzy rule interpolation to implement the above principle can be written as:

$$\frac{d(A^*, A_1)}{d(A^*, A_2)} = \frac{d(B^*, B_1)}{d(B^*, B_2)} \tag{2.3}$$

The fuzzy distance between the two fuzzy sets can be also interpreted by the pair-distance of the lower and upper boundary between their α-cut sets ($\alpha \in [0, 1]$), as denoted follows:

$$d_L(A_{i\alpha}, A_{j\alpha}) = d(\inf\{A_{i\alpha}\}, \inf\{A_{j\alpha}\}) \tag{2.4}$$

$$d_U(A_{i\alpha}, A_{j\alpha}) = d(\sup\{A_{i\alpha}\}, \sup\{A_{j\alpha}\}) \tag{2.5}$$

where d_L means the fuzzy distance between the lower boundary, and d_U means distance between the upper boundary.

Replacing the distance in equation 2.3 with fuzzy distance in equations 2.4 and 2.5, B^* can be calculated by:

$$\min\{B^*_\alpha\} = \frac{\frac{\inf\{B_{i\alpha}\}}{d_L(A^*_\alpha, A_{i\alpha})} + \frac{\inf\{B_{j\alpha}\}}{d_L(A^*_\alpha, A_{j\alpha})}}{\frac{1}{d_L(A^*_\alpha, A_{i\alpha})} + \frac{1}{d_L(A^*_\alpha, A_{j\alpha})}} \tag{2.6}$$

$$\max\{B^*_\alpha\} = \frac{\frac{\sup\{B_{i\alpha}\}}{d_U(A^*_\alpha, A_{i\alpha})} + \frac{\sup\{B_{j\alpha}\}}{d_U(A^*_\alpha, A_{j\alpha})}}{\frac{1}{d_U(A^*_\alpha, A_{i\alpha})} + \frac{1}{d_U(A^*_\alpha, A_{j\alpha})}} \tag{2.7}$$

It is also demonstrated that if the rule antecedents and the observations are normal triangular membership functions, the newly interpolated rules will also be normal triangular (Kóczy & Hirota, 1993). Therefore, two different α are sufficient to obtain the conclusion B^*. As for extending to multiple antecedent variables, which means that the antecedent consists of more than one variable, the distance will be calculated by follows:

$$d_i^\alpha = (\sqrt{(d_{i1}^\alpha)^2 + (d_{i2}^\alpha)^2 + ... + (d_{im}^\alpha)^2}))^{-1} \tag{2.8}$$

where d_{im}^α represents the distance of the mth antecedent between observation and rule R_i.

In last decades, various FRI approaches have been developed following this seminal work to perform fuzzy interpolative reasoning with different characteristics. For example, extensions of the KH FRI (e.g., (Koczy *et al.*, 1997); (Kovcs, 2014); (Koczy *et al.*, 2000); (Baranyi *et al.*, 1999); (Tikk & Baranyi, 2000b); (Yam *et al.*, 1999); (Joo *et al.*, 2002); (Au & Chan, 2002)), which establish the fundamental of FRI; FRI with multiple fuzzy rules (Chang *et al.*, 2008) (Chen *et al.*, 2013) (Yang & Sheng, 2013) (Kovcs, 2006), which expand the mechanism to multiple rules instead of only two rules; FRI with weighted fuzzy rules (Chen & Chen, 2016), which imports feature weights in FRI;, FRI with interval type-2 fuzzy sets (Chen & Lee, 2011), which extends FRI to type-2 fuzzy systems; transformation-based FRI (Huang & Shen, 2008), which solves the convex problem in FRI; and FRI with adaptivity (Cheng *et al.*, 2016) (Chen & Adam, 2017), which provide an approach to rectify FRI. The most recent work as reported in (Chen *et al.*, 2019) provides a novel approach that makes it possible to automatically select fuzzy rules for interpolation. Having recognised this, the approach presented in this thesis will also be compared with that approach.

2.2.2 Automated Rule Selection (AutoRS)

In many classic fuzzy rule interpolation methods, the number of closest rules K is a fixed number provided by a trial and error process, which is time and computation expended. Automated rule selection (AutoRS) (Chen *et al.*, 2019) proposed a novel approach that makes automatically selecting fuzzy rules useful for subsequent interpolation without human intervention possible. It searches for a set of closest fuzzy rules whose antecedent fuzzy sets lie on both sides of the unmatched observation, or preciously, the selected fuzzy rule set $U(R)$ must satisfy one of the following three scenarios:

1. $\exists R_i, R_j \in U(R)$, $Rep(A_{ik}) > Rep(B_k)$ and $Rep(A_{jk}) < Rep(B_k)$ (i < j)
2. $\exists R_i, R_j \in U(R)$, $Rep(A_{ik}) < Rep(B_k)$ and $Rep(A_{jk}) > Rep(B_k)$ (i < j)
3. $\exists R_i \in U(R), Rep(A_{ik}) = Rep(B_k)$

where $Rep(A_{ik})$ stands for the representative value of the k-th antecedent fuzzy set of i-th fuzzy rule, and $Rep(B_k)$ means the representative value of the k-th feature of the unmatched observation O.

The candidate of the required rule set for subsequent interpolation is obtained by adding rules iteratively from the nearest one to far away ones until the rule set satisfies the above scenarios. Then, the redundant rules in the candidate rule set are removed through a post pruning process in order to make the final rule set can keep the amount of rules to a minimum while still meeting the above scenarios. Its main procedure can be summarised as shown in Alg. 2.

Algorithm 2: AutoRS for interpolation

Input: Rule base $\{R_i\}$; Observation O

Output: Selected rule set for interpolation $U(R)$

1. Obtain a candidate rule set, $U'(R_k)$, for each feature by iteratively adding rules from nearest one to far away ones until rule set satisfies one of three scenarios.

2. Assign largest candidate rule set as initial $U(R)$.

3. Check if $U(R)$ satisfies either of scenarios for all other features. If so, stop; otherwise, update $U(R) = U(R) \cup U'(R_k)$

4. Post prune $U(R)$: For each $R_i \in U(R)$, if $U(R) - \{R_i\}$ still satisfies one of scenarios, update $U(R) = U(R) - R_i$; otherwise, retain R_i.

5. Repeat Step 4 with $i = i + 1$.

6. Return: $U(R)$.

Whilst having its own advantages, AutoRS still has two main limitations.

- It strongly restricts that the rules for interpolation have to flank the unmatched observation. It may fail in the certain situation where existing one of the variables whose values only lie on one side of the observation.
- It suffers the problem that in large size sparse rule bases, the chosen rules may be very similar with each other, which leads to lacking diversity for FRI.

2.2.3 TSK Inference Extension (TSK+)

TSK inference extension (TSK+) (Li *et al.*, 2018b) is a recently proposed fuzzy inference approach based on the TSK model, extending its capability to be able to handle sparse fuzzy rule bases. Instead of exploiting matching degrees, a similarity measure based on a certain distance metric (Chen & Chen, 2003) 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. The inference procedure of TSK+ is summarised in Alg. 3.

Whilst being a useful approach, TSK+ has its own shortcomings. Particularly, it is not sufficiently efficient for many practical applications given its nature of fundamentally requiring the use of all given rules, incurring significant computational overheads. Besides, redundant or even possibly irrelevant rules are also included in any attempt to compute the output. This may well introduce undesirable bias into the final interpolated outcomes, thereby reducing the system accuracy.

Algorithm 3: TSK extension

Input: Rule base $\{R_i\}$; Observation O

Output: Interpolated conclusion $f(O)$

1. Calculate similarities between observation and each rule R_i:
   $S(A_{i1}, B_1), ..., S(A_{in}, B_n)$
2. Determine 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 interpolated rule, with parameters of its consequent being:
   $$a_0 = \frac{\Sigma_{i=1}^{m} \alpha_i a_{i0}}{\Sigma_{i=1}^{m} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{m} \alpha_i a_{in}}{\Sigma_{i=1}^{m} \alpha_i} \quad (2.9)$$
4. Take observation O as input of interpolated rule and compute interpolated outcome as
   $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$
5. Return: $f(B_1, ..., B_n)$.

2.2.4 Dynamic Fuzzy Inference Systems Supported by Fuzzy Rule Interpolation

FRI methodologies for static systems provide the potential for performing fuzzy reasoning on sparse rule bases. However, in real-world applications of fuzzy systems, the inputs are usually time-dependent and the requirements of fuzzy systems may change over time. In such cases, the use of a static rule base (one that does not change over time) will affect the efficacy of fuzzy inference systems. Note that the goal of FRI is to produce an interpolative conclusion relevant to the input observation rather than to produce an interpolated rule through interpolative reasoning. The interpolated rules are usually discarded once the results are derived.

As such, if the frequently appearing unmatched observations are of high similarity, similar rule interpolation procedures will keep repeating, resulting in the increase of the computation overheads. Moreover, the abandoned interpolated rules may contain valuable potential information such as the uncovered input domain in original sparse rule bases. In order to collect and utilise such abandoned interpolated rules, a dynamic rule-based interpolation method is proposed in (Naik *et al.*, 2014).

Although there are many approaches which make the creation of a real-time rule base possible, in the areas of adaptive fuzzy control (Astrom & Witternmark, 1995) (Mohan & Bhanot, 2006) (Zhang & Bien, 2000) (Liu *et al.*, 2001) and optimization-based fuzzy rule generation (Wu *et al.*, 2001) (Angelov & Buswell, 2003) (Angelov, 2003). Unfortunately, all these techniques are developed for dense (fully covered) rule bases. They can not be applied to sparse rule bases directly due to the inherent pattern-matching mechanisms used by them, where no conclusion can be drawn when an observation does not (partially or fully) match any of the rules in the rule base. There are several methods have been proposed for dynamic models including (Åström & Wittenmark, 2013), most of them are developed on dense rule bases, which can not be employed to sparse rule bases directly.

Dynamic fuzzy rule interpolation (D-FRI) (Naik *et al.*, 2017b) presents a Genetic Algorithm (GA)-based dynamic FRI method for Mamdani models where a sparse rule bases is present. The overall inference process of D-FRI is summarised in Fig. 2.5.

In implementing this system, scale and move transformation-based fuzzy rule interpolation (T-FRI) (Huang & Shen, 2008) is employed to conducting interpolative reasoning and producing candidate interpolated rules. The new rules to be added into the original sparse rule base is generated by aggregating

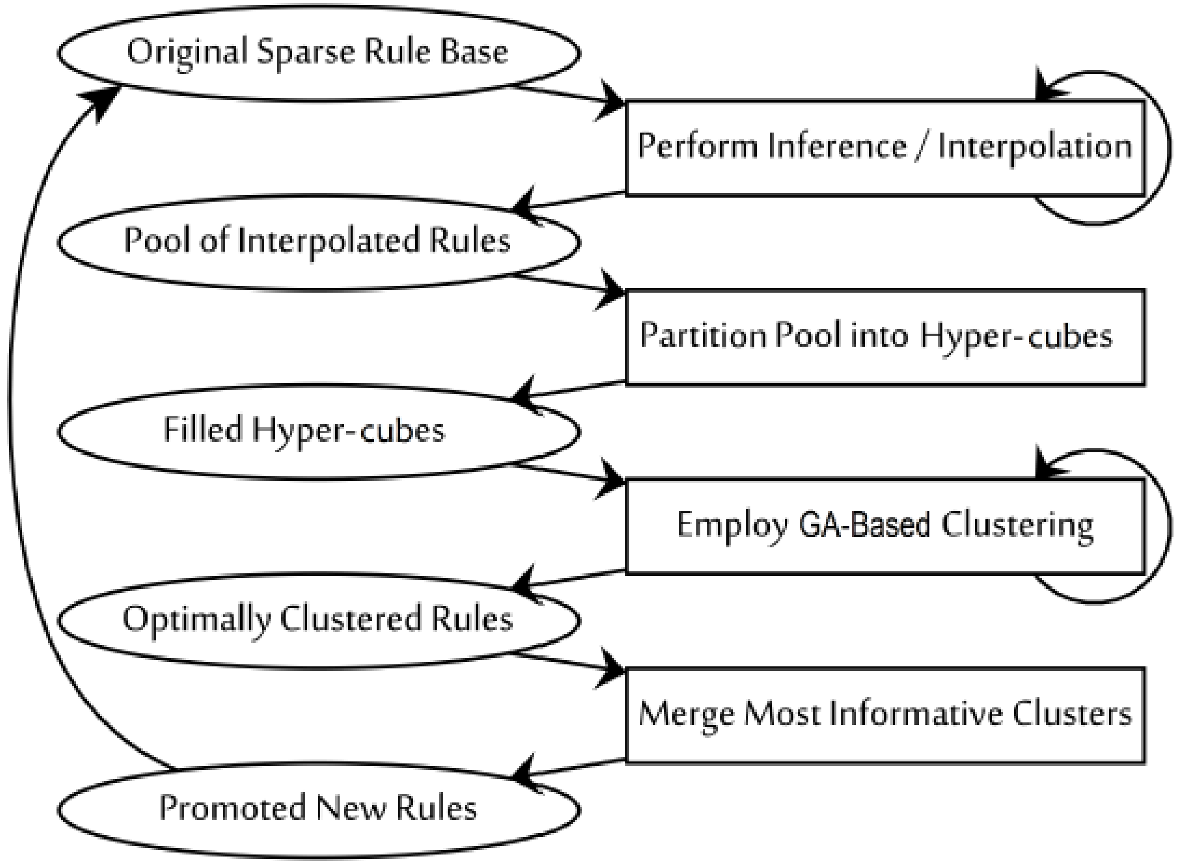

Figure 2.5: GA-based dynamic-FRI for Mandani models

and promoting all candidate rules in the following four procedures:

(1) Divide the input domain into same-size hyper-cubes (or multi-dimensional blocks). The size and number of hyper-cubes can be adjusted according to the requirements of inference. Suppose that the antecedent domain consists of N-dimensional variables, and each variable is divided into η intervals, then the total number of hyper-cubes is $\mathbb{H} = \eta^N$ (Naik *et al.*, 2014).

To deal with the situation where a rule may appear in more than one hyper-cube caused by rule antecedents are represented by fuzzy sets, the centre of gravity (COG) is applied to representing the original antecedents and determining which hyper-cuber the rule should belong to. Within all hyper-cubes, only non-empty ones are valuable and will be termed as input for subsequent GA-based clustering algorithm.

(2) Cluster these non-empty hyper-cubes with GA-based clustering algorithm. The Dunn Index is applied to designing the fitness function due to its advantages in evaluating clusters (Dunn, 1973), which is defined as the following:

$$f(X_i) = \min\{\frac{m_{pq}}{max(s_r)}\}(p \neq q) \tag{2.10}$$

where s_r is the intra-cluster distance measurement (the distance between an interpolated rule and the centroid of its cluster), and m_{pq} is inter-cluster distance measurement (the distance between the centroids of two clusters), which are respectively calculated by:

$$s_r = \sqrt{\Sigma_{R' \in C_r} \frac{d(R', \mu_r)^2}{|C_r|}} \tag{2.11}$$

$$m_{pq} = d(\mu_p, \mu_q) \tag{2.12}$$

where C_r represents the rth cluster in the result, and μ_r, μ_p, μ_q represent the centroid of C_r, C_p, C_q, respectively. Particularly, in Eqn. 2.11, the distance between R' and μ_r is given:

$$d(R', \mu_q) = \sqrt{\Sigma_{i=1}^{N} rep(A'_i) - \mu_{(q,i)})^2} \tag{2.13}$$

where $rep(A'_i)$ is implemented by the centre of gravity (COG) of the i-th antecedent of R', and $\mu_{(q,i)}$ is the corresponding variable of cluster centroid q. Distinct from conventional clustering algorithms in which the number of clusters has to be manually decided, such as K-means (MacQueen *et al.*, 1967), Gaussian mixture models (Bilmes *et al.*, 1998) and fuzzy c-means (Bezdek *et al.*, 1984), GA-based clustering algorithm makes automatically selecting the number of clusters without human intervention possible.

(3) Determine strong hyper-cubes and weak ones based on the number of rules in each rule cluster according to the predefined threshold. Merge weak hyper-cubes into their closest strong ones.

(4) Integrate and promote the candidate rules in the strong hyper-cubes to obtain representative rules. The obtained rules are added into original sparse rule bases, while the candidate rules are abandoned.

Note that the interpolation and rule promotion process of D-FRI is particularly designed for Mamdani type of fuzzy models, whose consequents are fuzzy sets. To establish a dynamic TSK model whose consequents are polynomial, an adapted approach for polynomial consequents is required.

2.3 Clustering Algorithms

A brief introduction to five popularly used and readily available clustering algorithms is given in this section, which will be adopted to implement the rule clustering task in the subsequent development. Any one of these may be employed to carry out the intended task, but these are collectively reviewed to facilitate comparison, in an effort to make an informed choice of the potentially most suitable. These five clustering algorithms represent different clustering mechanisms, K-means is the conventional distance-based clustering techniques, GMM is one of popular statistic-based clustering approaches, and fuzzy c-means and its two extensions represent fuzzy-based clustering algorithms.

2.3.1 K-means

As one of the most widely used fuzzy clustering algorithms, K-means (MacQueen *et al.*, 1967) clusters instances into K groups by iteratively updating cluster centres and assigning instances to their closest centres. The underlying objective function (known as the inertia or within-cluster sum-of-squares error) is defined as follows:

$$J_w(U, V) = \Sigma_{j=1}^{N} \Sigma_{i=1}^{K} \|x_{ji} - v_i\|^2 \tag{2.14}$$

where U denotes the set of instances, x_{ji} expresses that the j-th instance belongs to the i-th cluster, V stands for the set of cluster centres and $v_i \in V$, $\|x_j - v_i\|^2$ represents the Euclidean distance between the object x_{ji} and the centre v_i, K is the number of cluster centres, and N is the number of instances. The iterative process is detailed in Alg. 4.

Algorithm 4: K-means

Input: Instances x_j; Number of cluster K

Output: Instances with label; Cluster centres V

1. Initialize $V^{(0)}$ randomly.
2. Set max iteration number L, termination condition ε, and counter k, $k = 0, 1, ...L$
3. Compute the Euclidean distances between instances and all cluster centres: $D(x_j, v_i)$ $(v_i \in V^{(k)})$
4. Assign every instance to its closest centre.
5. Update $V^{(k+1)}$ with the mean values of instances in the corresponding clusters
6. Compare $V^{(k+1)}$ and $V^{(k)}$. If $\|V^{(k+1)}$ - $V^{(k)}\| < \varepsilon$ or k = L, then stop; otherwise, set $V^{(k)} = V^{(k+1)}$ and return to **step 3)** .
7. Return: Labeled instances; $V(K + 1)$.

This algorithm requires the number of clusters to be pre-specified. Many methodologies have been proposed in the literature to determine the number of clusters, K, such as those through the use of: the Elbow method (Thorndike, 1953), cross-validation, Bayesian Information Criterion (Bouveyron & Brunet-Saumard, 2014) and Silhouette-based technique (Kodinariya & Makwana, 2013). The Elbow method is fast and effective, and hence it is employed in this thesis.

The elbow method determines the value of K based on the criteria that adding another cluster does not lead to a much better modelling result (in terms of the given objective function Eqn. 2.14). Fig. 2.6 shows a brief example of the relationship between the performance of K-means (SSE) and the value of K. Obviously, 4 is the elbow point, thus the value of k, in this case will be set to 4.

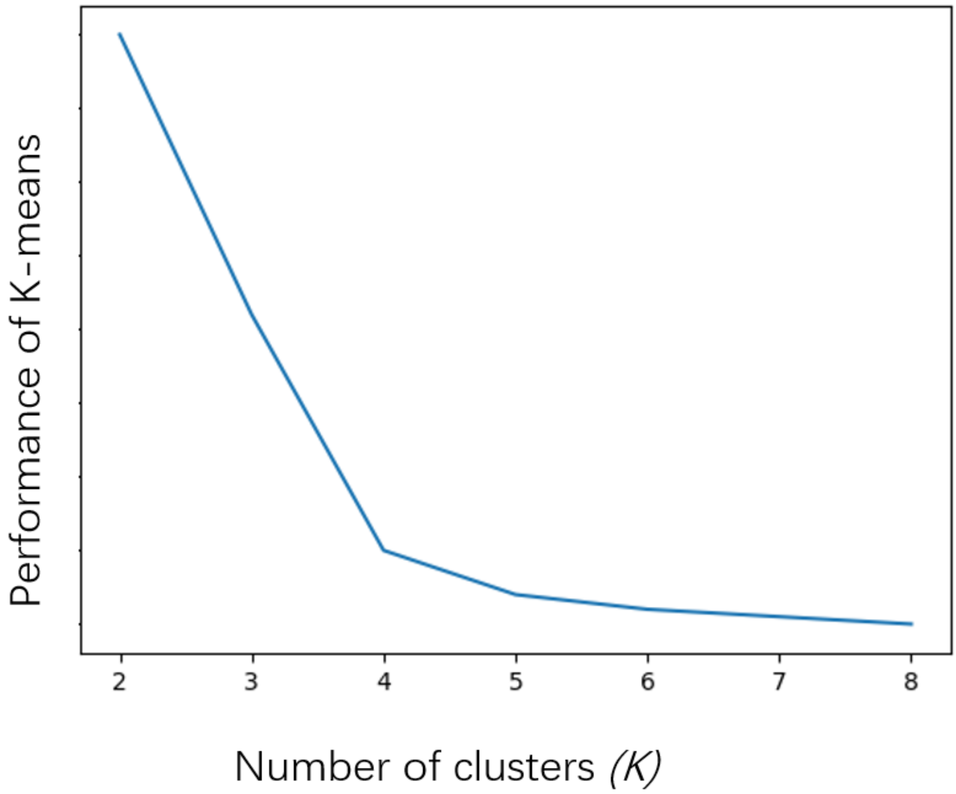

Figure 2.6: Determination of K using Elbow method.

2.3.2 Gaussian Mixture Models

Being another classical clustering algorithm, Gaussian Mixture Models(GMM) (Bilmes *et al.*, 1998) works by presuming that the distribution of instances conforms to the linear combination of multiple Gaussian distribution functions, which is defined by:

$$p(x_j) = \Sigma_{i=1}^{K} \pi_{ji} \mathcal{N}(x_j | \mu_i, \sigma_i) \tag{2.15}$$

where x_j denotes an instance to be clustered, π_{ji} represents the probability that the instance x_j belongs to the i-th cluster, and μ_i and σ_i respectively stands for the mean and standard deviation of the i-th Gaussian model.

Theoretically, this model can fit any type of data distribution. The expectation maximisation (EM) algorithm (Dempster *et al.*, 1977) is the most commonly used algorithm to construct GMM, which is outlined in Alg. 5,

Particularly, to identify the optimal partitions of GMM, the log-likelihood function to be maximized is given:

$$\log(\prod_{j=1}^{N} p(x_j)) = \eta_{j=1}^{N} log(\Sigma_{i=1}^{K} \pi_{ji} \mathcal{N}(x_j|\mu_i, \sigma_i)) \tag{2.16}$$

And the corresponding parameters are updated iteratively such that:

$$\begin{aligned} \mu_i &= \frac{1}{N_i} \Sigma_{i=1}^{N} \eta_{ji} x_j \\ \sigma i &= \frac{1}{N_i} \Sigma_{i=1}^{N} \eta_{ji} (x_j - \mu_i)(x_j - \mu_i)^T \\ \pi_{ji} &= \frac{N_i}{N} \end{aligned} \tag{2.17}$$

where $N_i = \Sigma_{j=1}^{N} \eta_{ji}$, and η_{ji} represent the conditional probability which is calculated by:

$$\eta_{ji} = \frac{\pi_{ji} \mathcal{N}(x_j|\mu_i, \sigma_i)}{\Sigma_{i=1}^{K} \pi_{ji} \mathcal{N}(x_j|\mu_i, \sigma_i)} \tag{2.18}$$

Algorithm 5: EM algorithm for GMM Construction

Input: Instances x_j; Number of cluster K

Output: Instances with label; Cluster centres V

1. Initialisation: Calculate initial means and variances of Gaussian models using global dataset statistics, with number of models fixed to predefined value K. Set termination condition ε.

2. E-step (Expectation): For each data point, compute its conditional probability under each model using current setup of parameters with Eqn. 2.18

3. M-step (Maximization): For each Gaussian model, update its parameters using data and their corresponding conditional probabilities with Eqn. 2.17

4. Iteration: Iterate Steps E and M until log likelihood (Eqn. 2.16 converges or becomes smaller than preset termination condition ε.

5. Return: Labeled instances; V.

2.3.3 Fuzzy C-means

Fuzzy c-means (FCM) (Bezdek *et al.*, 1984) is a conventional clustering algorithm, which has been popularly applied in dealing with various problems (e.g., for fuzzy rule base generation (Li *et al.*, 2018a) and social network modelling (Zhang & Shen, 2018)). Unlike crisp clustering algorithms such as K-means and GMM, FCM allows an instance to belong to different clusters at the same time with different membership degrees. It works by assigning membership degrees to each instance based on the relative distance measurement between individual cluster centres and that instance. The closer an instance is to a cluster centre, the higher the membership degree the instance has. To identify the optimal partitions of FCM, the objective function to be minimized is defined as the following:

$$J_w(U,V) = \Sigma_{j=1}^{N}\Sigma_{i=1}^{K}(u_{ij})^w \|x_j - v_i\|^2 \tag{2.19}$$

where x_j denotes an instance, w is a parameter that signifies the weight of each component, K stands for the number of cluster centres, N is the number of objects, V is the set of cluster centres with $v_i \in V$, U is the matrix of membership degree and $u_{ij} \in U$ represents the membership degree of the object x_j belonging to the centre v_i, and $\|x_j - v_i\|^2$ expresses the similarity between the object x_k and the centre v_i.

The membership u_{ik} and the centre v_i are updates iteratively as follows:

$$v_i = \frac{\Sigma_{j=1}^{N}(u_{ij})^w x_j}{\Sigma_{j=1}^{N}(u_{ij})^w} \tag{2.20}$$

$$u_{ij} = (\Sigma_{n=1}^{K}(\frac{\|x_n - v_i\|}{\|x_j - v_n\|})^{2/w-1})^{-1} \tag{2.21}$$

The procedure of the fuzzy c-means algorithm is summarised in Alg. 6.

Algorithm 6: Fuzzy C-means

Input: Instances x_j; Number of cluster K

Output: Instances with label; Cluster centres V; Matrix of membership degree U

1. Initialize $U^{(0)}$ randomly.
2. Set max iteration number L, termination condition ε, and counter k, $k = 0, 1, ...L$
3. Compute $V^{(k)}$ with $U^{(k)}$ with Eqn. 2.20.
4. Update $U^{(k+1)}$ by $V^{(k)}$ with Eqn. 2.21.
5. Compare $U^{(k+1)}$ and $U^{(k)}$. If $\|U^{(k+1)}$ - $U^{(k)}\| < \varepsilon$ or k = L, then stop; otherwise, set $U^{(k)} = U^{(k+1)}$ and return to **step 3)** .
6. Return: Labeled instances; $V(k+1)$; $U(k+1)$.

2.3.4 Kernel Fuzzy C-means

Being an extension to the standard FCM, Kernel fuzzy c-means (K-FCM) (Zhang & Chen, 2004) it works by replacing the original Euclidean distance with a kernel-induced distance. Kernel function is a nonlinear mapping that transforms a low dimensional input data space into a feature space with a much higher dimension, aiming at turning the original nonlinear problem into a potentially linear one so as to facilitate problem solving (Müller *et al.*, 2018).

There are three common used kernel function in the literature (Mller *et al.*, 2001), which are:

(1) Gaussian radial basis function (GRBF) kernel:

$$KN(x_j, v_i) = \exp \frac{-\|x_j - v_i\|^2}{\sigma^2} \tag{2.22}$$

(2) Polynomial kernel:

$$KN(x_j, v_i) = (1 + \langle x_j + v_i \rangle)^d \tag{2.23}$$

(3) Sigmoid kernel:

$$KN(x_j, v_i) = \tanh(\alpha \langle x_j + v_i \rangle + \beta) \tag{2.24}$$

where x_j states an instance, v_i means a cluster centre, σ^2 is the variance of the instances and d, α, β are the adjustable parameters of the above kernel functions. For the sigmoid function, only a set of parameters satisfying the Mercer theorem can be used to define a kernel function.

The Gaussian radial basis function (GRBF) kernel is the one employed in this thesis because no additional parameters are introduced. As it is a direct extension of FCM, the algorithm procedure is omitted here. However, note that adapted from its original, the underlying objective function is now:

$$J_w(U, V) = \Sigma_{j=1}^{N} \Sigma_{i=1}^{K} (u_{ij})^w (1 - KN(x_j, v_i)) \tag{2.25}$$

with Eqn. 2.20 and Eqn. 2.21 respectively transformed to:

$$v_i = \frac{\Sigma_{j=1}^{N}(u_{ij})^w KN(x_j, v_i)x_j}{\Sigma_{j=1}^{N}(u_{ij})^w KN(x_j, v_i)} \tag{2.26}$$

$$u_{ij} = (\Sigma_{n=1}^{K}(\frac{1 - KN(x_j, v_i)}{1 - KN(x_j, v_n)})^{1/w-1})^{-1} \tag{2.27}$$

2.3.5 Suppress Fuzzy C-means

Suppress fuzzy c-means (S-FCM) (Fan *et al.*, 2003) is another extension of FCM, following the motivation for improving on its convergence speed (Cannon *et al.*, 1986). It is developed on the basis of the rival-checked fuzzy c-means clustering algorithm (RCFCM) (Wei & Xie, 2000) that speeds up FCM with competitive learning. RCFCM proposed a membership modification approach to magnify the influence of the cluster with the largest membership degree to the instance after updating membership matrix U as the following:

$$\begin{aligned} u_{pj} &<= u_{pj} + (1 - \alpha)u_{sj} \\ u_{sj} &<= \alpha u_{sj} \end{aligned} \tag{2.28}$$

where u_{pj} represent the largest membership degree with the primary cluster, and u_{sj} means the second-largest membership degree.

Whilst providing the potential to improve the convergence speed, RCFCM has its own shortcomings. Particularly, the RCFCM algorithm only pays attention to the clusters with the largest and the second-largest membership degree while ignoring the impact of other clusters. Moreover, if α is not selected appropriately, it may lead the modified second biggest membership u_{sj} smaller than the other unrevised ones. Thus, the original order which reflects the natural relation between the instance and each cluster will be disturbed.

S-FCM is a development of RCFCM which makes the inference process more natural and more effective. The underlying mechanism of this approach is to magnify the largest membership degree, u_{pj} and suppress the others. In order to achieve such an objective, a membership modifying mechanism is added after iteratively updating the membership degrees U, such that

$$\begin{aligned} u_{pj} &<= 1 - \alpha \sum_{i \neq p} u_{ij} = 1 - \alpha + \alpha u_{pj} \\ u_{ij} &<= \alpha u_{ij} \end{aligned} \tag{2.29}$$

where $u_{pj} > u_{ij}, i \neq p$, and $0 \leq \alpha \leq 1$.

The inference procedure of S-FCM is demonstrated in Alg. 7.

Algorithm 7: Suppress Fuzzy C-means

Input: Instances x_j; Number of cluster K

Output: Instances with label; Cluster centres V; Matrix of membership degree$U$

1. Initialize $U^{(0)}$ randomly.
2. Set max iteration number L, termination condition ε, and counter k, $k = 0, 1, ...L$
3. Compute $V^{(k)}$ with $U^{(k)}$ with Eqn. 2.20.
4. Update $U^{(k+1)}$ by $V^{(k)}$ with Eqn. 2.21.
5. Modify obtained $U^{(k+1)}$ with Eqn. 2.29.
6. Compare $U^{(k+1)}$ and $U^{(k)}$. If $\|U^{(k+1)}$ - $U^{(k)}\| < \varepsilon$ or k = L, then stop; otherwise, set $U^{(k)} = U^{(k+1)}$ and return to **step 3)** .
7. Return: Labeled instances; $V(k+1)$; $U(k+1)$.

2.4 Missing Value Imputation

In real-world data mining applications, missing values are a common problem leading to a lack of data integrity and reduce the reliability of subsequent developments (Little & Rubin, 1986) (Pigott & Therese, 2001). Directly discarding instances containing missing values may work well in the situation where only a small number of missing values are included. Notwithstanding this, similar issues remain in datasets that contain much more missing values (Yuan *et al.*, 2017) (Farhangfar *et al.*, 2008). Missing value imputation has been introduced to deal with such an issue, and aims to fill missing values with reasonable ones learned from present data. Several approaches have been developed in the literature and successfully applied in many real-world applications, including microarray gene expression data (Cheng *et al.*, 2012) and breast cancer data (Jerez *et al.*, 2010). This section briefly reviews three popular and readily available methods.

2.4.1 Statistic Imputation

Examples of conventional missing value imputation approaches include statistical methods, of which mean and median imputation are the two most widely used. The underlying principle of both is to take the statistical value of the present data in each incomplete attribute as an estimate of the missing value, which can be respectively implemented from the mean and median values (Little & Rubin, 1986). Whilst having low computation overheads, the obtained results of statistical approaches also have low accuracy, due to the fact that only values of the same attribute are involved.

2.4.2 Linear Regression Imputation

In terms of classic machine learning algorithms, linear regression approaches have been adopted to perform missing value imputation over continuously-valued domains in the literature (Hyunsoo *et al.*, 2005) (Rana *et al.*, 2012). The basic assumption for this method is that the values of an individual attribute can be represented by a linear combination of the values of other attributes. Following the above presumption, the linear model is derived from complete instances (without missing values), taking the attribute with the missing value as output and other attributes as inputs. The missing value is then predicted by the linear approximation.

The preferred technique of creating a linear model for a dataset is through minimizing the residual sum of squares (RSS) between the observed instances (Freedman, 2009), which is defined by:

$$RSS = \Sigma_{i=1}^{n}(Y_i - f(X_i)^2 \tag{2.30}$$

where Y_i denotes the groundtruth of i-th value of the variable, and fX_i represents the corresponding predicted value.

2.4.3 Nonlinear Regression Imputation

In many real-world datasets, the relationship between each attribute may be nonlinear, which makes it difficult to precisely estimate them using linear regression approaches. TSK models have the potential to represent a nonlinear model (and a linear model of course) with several different linear ones described by 'if-then' rules. They have been successfully adopted to perform missing value imputation in (Lai *et al.*, 2019) and (Lai *et al.*, 2020).

This method works by clustering instances into different categories by FCM, and then deriving the polynomial consequent of an emerging rule through linear regression as described in (Freedman, 2009). As with linear regression techniques, the attribute with the missing value is taken to be the output and the other attributes as inputs. The calculation process for estimating missing values is as same as the inference with the conventional TSK model, which can be found in Alg. 1.

Whilst being generally powerful, this approach is not efficient enough for many practical applications given its nature of fundamentally requiring the input domain to be fully covered. As indicated previously, such a restriction is hard to satisfy for most real-world datasets, especially for those with missing values. In a situation where only sparse rule bases are available, TSK model-based approaches are unable to produce any conclusions by applying any of the classical rule-firing methods.

Chapter 3

FRI with K-Neighbours for Static TSK Models

Fuzzy rule-based inference systems are one successful representative of knowledge-based systems, the basic idea of which is to represent domain knowledge in the form of 'if-then' production rules while using fuzzy sets to describe imprecise variable values that are hard to be described precisely in traditional crisp rule-based systems. TSK models (Takagi & Sugeno, 1985) are one of the conventional and most exploited FRBSs in the literature. Within such a model, fuzzy sets are used as rule antecedents and polynomials as the consequents, whose application directly results in crisp conclusions and thereby, being particularly suitable for solving regression and prediction problems, over continuously-valued domains.

Whilst being generally powerful, conventional rule-based systems all suffer from an important limitation, be they fuzzy or not. That is, if the input domain is not completely covered, a novel observation may not always match any rule in the given rule base. In this case, they are unable to produce any conclusion by applying any of the classical rule-firing mechanisms. This is independent of what rule models are employed. To rectify, or at least to reduce the adverse impact of this limitation, fuzzy rule interpolation (FRI) has been introduced (Li *et al.*, 2021). If a newly presented input or observation does not match any of the rules available, FRI can help by generating an intermediate rule through the approximation of those rules close to the observation, from which a potentially relevant conclusion may then be obtained.

While conducting investigations on FRI, it is observed that most existing approaches are developed for reasoning with Mamdani models, and those designed for TSK models have inevitable limitations. This chapter presents a novel approach for performing FRI with TSK models. The work presents two advancements in developing FRI methodologies for TSK models, through two novel implementations: (i) interpolation with K Closest Rules (KCR) for sparse rule bases of a small size, and (ii) that with K Closest Rule Clusters (CRC) for rule bases of a large size. The underlying principle for both is to perform interpolation with K closest rules where K is normally a small number.

Another key issue concerned in this chapter is the determination of the rule weight. Generally, the weight of a certain rule is determined by an integrating operation (usually implemented by a minimum operator) on resulting individual similarity between individual antecedent variables of each rule and their counterparts in the observation. Distance measurements (e.g., (Chen, 1996); (Sridevi & Nadarajan, 2009)) are the most commonly used approaches to evaluate the similarity degree between two fuzzy sets in the literature, which usually produce acceptable results. A similarity measurement is recently proposed in (Li *et al.*, 2018b) to extend its sensitivity to distance. The most recent work of (Yang *et al.*, 2021) presents a different approach which establishes a mapping from the input domain to the output domain by Locally Linear Embedding (LLE) (Roweis & Saul, 2000). As such, the weight of each selected rule can be directly calculated without computing individual similarity. The efficacy of these three approaches will also be experimentally compared against each other in this chapter.

To have a fair comparison over different methods, a range of experimental studies are systematically carried out over ten benchmark datasets. Statistic analyses of the results demonstrate that KCR indeed improves the performance over TSK+ and CRC for cases involving small-sized sparse rule bases and that for cases involving large-sized sparse rule bases, CRC outperforms TSK+ and KCR. Of course, both KCR and CRC offer superior results over the existing approach TSK+.

The rest of this chapter is structured as follows. Section 3.1 and Section 3.2 detail the two aforementioned methods, respectively introduced for interpolation with K closest rules (KCR) and that with K closest rule clusters (CRC). Section 3.3 reviews three rule weight determination approaches involved in interpolative reasoning.

3.1 Interpolation with K Closest Rules (KCR)

From the specification as well as the application of TSK+, it is easy to reveal that those rules nearest to an unmatched observation generally have a much higher similarity degree than others. This indicates that the interpolated outcomes may be mainly determined by those closest rules, with the rest typically contributing substantially less. Note that in the FRI literature, it is often assumed that the interpolated conclusion is estimated by a certain aggregation of those neighbouring rules to the observation (Li *et al.*, 2021). That is, the nearest rules are (normally correctly) considered to contain the most relevant information whilst those rules far away from the observation are less relevant. Indeed, distant rules may introduce adverse biases into the results, with their use becoming counter-productive. As far-away rules generally have relatively smaller similarity measures against the observation, such biases do not necessarily impose much influence upon the interpolated results, but they do induce significant computational overheads if there are many rules in the rule base. Thus, such biases or the use of remote rules should be minimised, for both the efficiency and the effectiveness of the interpolative reasoning process.

To address the aforementioned issue, a significantly revised inference procedure, termed *interpolation with K closest rules (KCR)* is introduced in this section. The underpinning idea is that only K nearest neighbouring rules to a given unmatched input are exploited in performing the interpolation, rather than involving all the rules in the sparse rule base. In so doing, the final conclusion may avoid the adverse impact caused by rules with low similarities in a small size sparse rule base while reducing computation overheads.

3.1.1 KCR Procedure

Suppose that a TSK sparse fuzzy rule base is given, containing m rules with n antecedent variables, each rule is then defined by

$$R_i: \textbf{ if } x_1 \text{ is } A_{i1}, ..., x_n \text{ is } A_{in}, \textbf{ then } f_i(x_1, ..., x_n) = a_{i0} + a_{i1}x_1 + ... + a_{in}x_n \quad (3.1)$$

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 inference process of KCR can be described as Alg. 8.

Algorithm 8: Interpolation with K Closest Rules

Input: Rule base $\{R_i\}$; Unmatched observation O

Output: Interpolated conclusion $f(B_1, ..., B_n)$

1. Calculate overall Euclidean distance between representative values of individual variables within given observation and corresponding antecedent variables for each rule.

2. Select K closest rules by Quickselect (Hoare, 1961) (which is utilised purely for efficiency while any alternative selection mechanism may be employed if preferred).

3. Calculate similarity between observation O.

   and each of R_i that belongs to set of selected K closest rules: $S(A_{i1}, B_1), ..., S(A_{in}, B_n)$

4. Determine 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 working interpolated rule with following consequent parameters:

$$a_0 = \frac{\Sigma_{i=1}^{K} \alpha_i a_{i0}}{\Sigma_{i=1}^{K} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{K} \alpha_i a_{in}}{\Sigma_{i=1}^{K} \alpha_i} \quad (3.2)$$

6. Take observation O as input to fire interpolated rule such that consequent is obtained by computing:

   $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$

7. Return: $f(B_1, ..., B_n)$.

3.1.2 KCR Complexity

From the above process, it can be seen that the Euclidean distance 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. Note that the similarity measure is only applied K times for K selected rules rather than all the rules given in the rule base. As such, the proposed approach can significantly reduce the running time. Additionally, the Quickselect algorithm also helps decrease the computation of closest rule selection.

Indeed, given m rules each involving n antecedent attribute, the time complexity of the proposed KCR method is $O(mK+nK)$, where $O(mK)$ stands for the time complexity of K rules selection. In comparison, TSK+ has a complexity of $O(mn)$ since all rules are fired for final conclusion generation. AutoRS also has a complexity of $O(mn)$ due to the rule selection is required for every single feature. Note however that generally, K is much smaller than m or n. For example, there is 1000 rules each involving 100 antecedent attributes and 10 rules are selected for interpolation, the time complexities of TSK+ and AutoRS are $1000 \times 100 = 10^5$, while that of KCR is $1000 \times 10 + 100 \times 10 = 1.1 \times 10^4$.

3.2 Interpolation with K Closest Rule Clusters (CRC)

When applying KCR to solve problems that involve a large-sized sparse rule base, the K nearest rules with the greatest similarity degrees may appear to be rather more similar amongst themselves than the rest, in this case, rules are too close to each other. This may be expected intuitively, as illustrated in Fig. 3.1. Rule antecedents and the observation are represented by their representative values (in terms of the center of gravity).

Thus, if only these K rules are taken to implement interpolation, the results will be also similar to the linear combination of their consequents regardless of what similarity measures may be. Of course, this potential problem is not unique to KCR, but it can arise in TSK+ as well, despite that all rules are involved in rule interpolation there. This is simply because the similarities of the selected K rules are much larger than those measured over the rest.

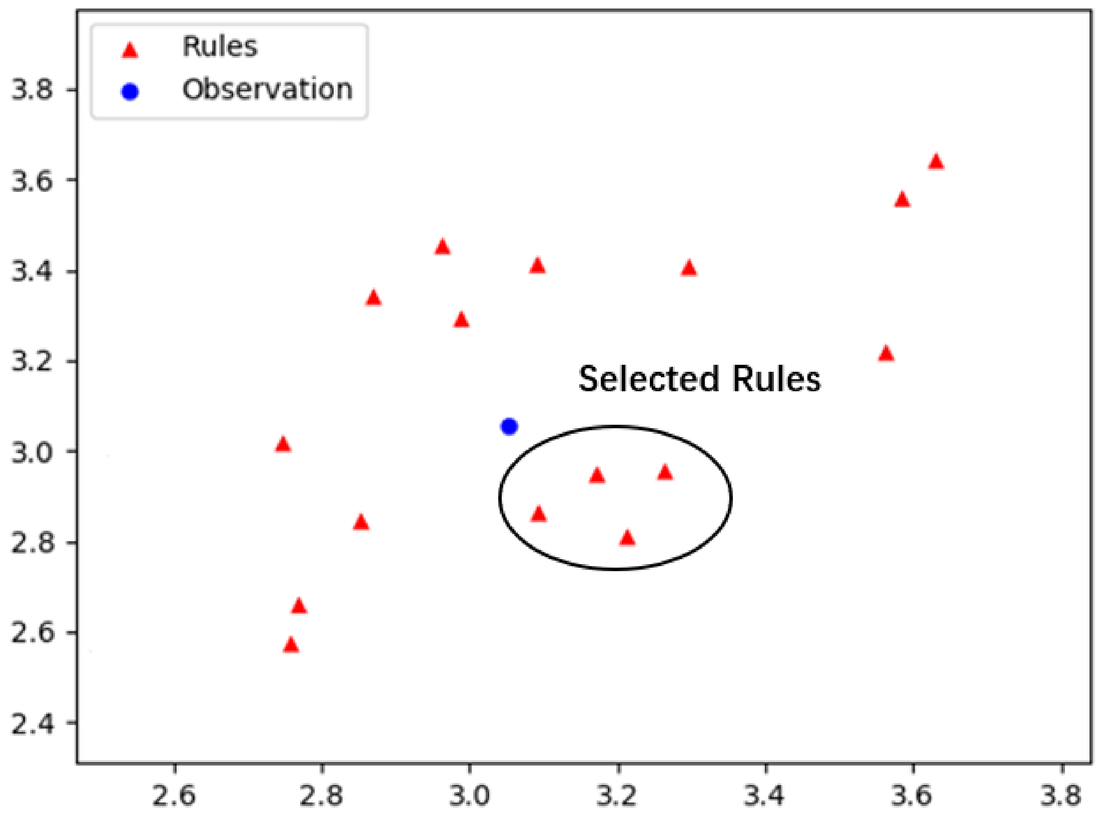

Figure 3.1: Example of rules selected with respect to unmatched observation by KCR in large-sized sparse rule base.

That is, the final interpolated result in TSK+ is also mainly determined by those nearest ones.

The above analysis prompts the need to extend the diversity of rules selected for use in the rule interpolation process, in an effort to avoid the involvement of far too many similar rules. Driven by this consideration, a clustering-aided interpolative reasoning process is presented here, termed *interpolation with K closest rule clusters (CRC)* hereafter.

In CRC, to maximise computational efficiency, all fuzzy values appearing in the rule antecedents are approximately represented using their representative values, which are firstly clustered into different groups, by the use of a clustering method. Those in the same cluster can be intuitively regarded as containing similar information about the mappings between the antecedents and the consequents; after all, they have been deemed to belong to the same cluster. Then, K clusters nearest to an unmatched observation are selected, with K being a small number, where the distance between a cluster and an observation is determined by the Euclidean measure between the cluster centre and the observation, which is defined as:

$$d(O, C) = \sqrt{\Sigma_{i=1}^{n}(B_i - c_i)^2} \tag{3.3}$$

where $O(B_1, ..., B_n)$ stands for the observation, and $C(c_1, ..., c_n)$ represents the centre of rule cluster.

From each of these selected clusters, the rule that is the closest to the observation is then selected as an element of a set of K nearest rules to be used for interpolation. Thus, rules measured without necessarily having the higher similarity measures are able to contribute to the creation of the final interpolated consequent. Note that through such a rule selection process, the rule that is of the overall closest distance to the observation is always included to participate in rule interpolation. This is obvious as it always is the representative of a certain cluster of rules since it has the greatest similarity to the observation amongst the entire rule base. In so doing, the proposed implementation resolves the problem of lacking diversity of rules being involved in subsequent interpolation where a large-sized sparse rule base is present.

Fig. 3.2 presents an illustrative example of rule clusters and rules selected from the closest clusters that are produced by the CRC algorithm. Similarly, rule antecedents and the observation are represented by their representative values.

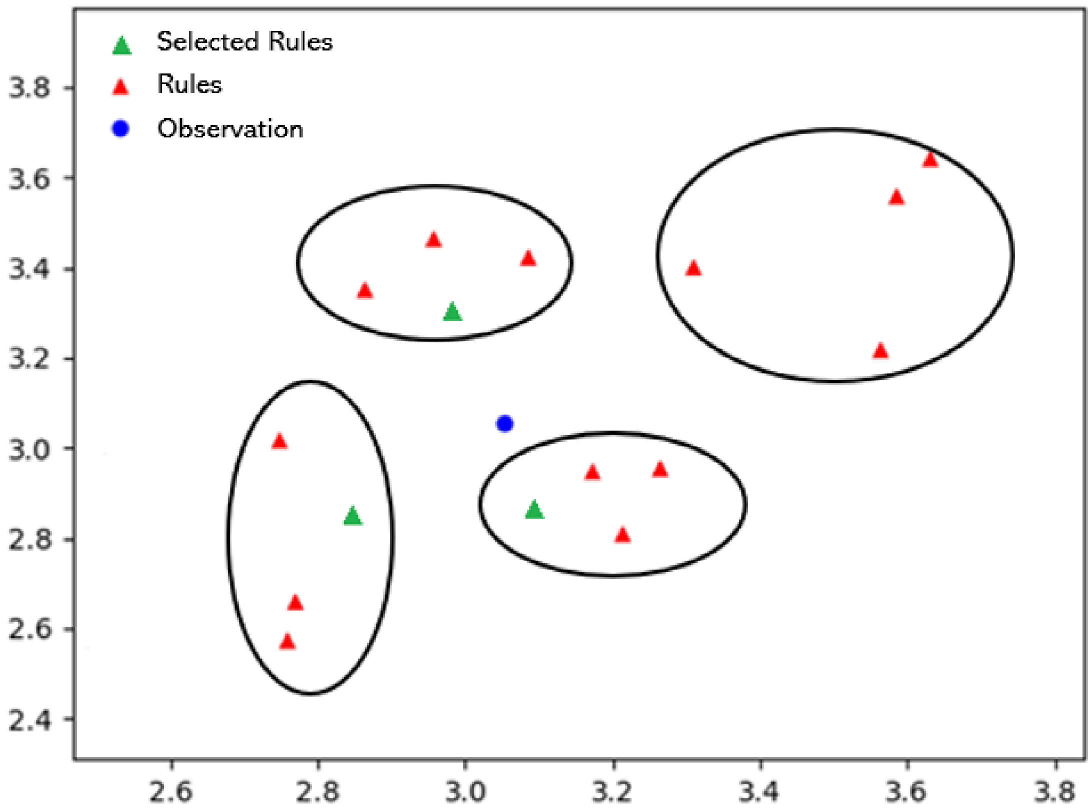

Figure 3.2: Example of rules selected with respect to unmatched observation by CRC in large-sized sparse rule base.

Note that the top right cluster is not selected because it is further away from the observation than the other three rule clusters. Besides, CRC only requires several nearest clusters instead of all clusters for rule interpolation.

3.2.1 CRC Procedure

The above idea is organised to form the proposed approach as follows. Suppose that a sparse rule base which contains m rules with n antecedents and an observation $O(B_1, ..., B_n)$ (which does not match any of the rules) are given, with each rule being specified as per Eqn. 3.1. The details of the proposed procedure for CRC is detailed in Alg. 9. Note that any of the five different clustering algorithms outlined in Section II-D (and indeed many other clustering methods if preferred) can be employed here to perform rule clustering, of course.

3.2.2 CRC Complexity

The time complexity of the proposed CRC procedure is $O(KC + KG + nK)$, 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, and n is the number of antecedent attributes in the rules. Compared with KCR, of which time complexity is $O(mK + nK)$, CRC can also decrease the computation effort. 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. For example, there is 1000 rules each involving 100 antecedent attributes, and 10 rules are selected for interpolation, suppose 100 rule clusters each including 10 rules are generated from the rule base, the time complexity CRC is $100 \times 10 + 10 \times 10 + 100 = 2.1 \times 10^3$. Therefore,

Algorithm 9: Interpolation with K Closest Rule Clusters

Input: Rule base $\{R_i\}$; Unmatched observation O

Output: Interpolated conclusion $f(O)$

1. Cluster all rules into C different groups using representative values of involved attributes throughout.
2. Calculate Euclidean distance between observation and all cores of the C clusters and select K $(K \leq C)$ closest clusters.
3. Take one C_i of K clusters, $i \in \{1, \ldots, K\}$, and compute distance between observation O and each and every rule in it.
4. Find closest rule R_i in C_i as its representative.
5. Determine weight of 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 interpolated rule, resulting in consequent parameters being:

$$a_0 = \frac{\Sigma_{i=1}^{K} \alpha_i a_{i0}}{\Sigma_{i=1}^{K} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{K} \alpha_i a_{in}}{\Sigma_{i=1}^{K} \alpha_i} \tag{3.4}$$

8. Fire interpolated rule with observation O and compute final consequent outcome with respect to O:

   $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$
9. Return: $f(B_1, ..., B_n)$.

$O(KC + KG)$ is generally smaller than $O(mK)$. In other words, the time complexity is lower than that of KCR.

3.2.3 Integration of KCR and CRC

In order to establish a static TSK type fuzzy inference system applicable for performing reasoning on rule bases under all different circumstances (whether sparse or dense, small-sized or large-sized). Both KCR and CRC procedures are integrated into a single algorithm, in conjunction with the conventional inference mechanism for TSK models. This integration is straightforward, deriving an overall novel approach, as shown in Fig. 3.3, which works by the use of just a small number of closest rules to infer the final conclusion while the conventional method is stuck when an observation matches no rules.

3.3 Determination of Rule Weight

The proposed algorithm computes the weights of selected rules via the employment of similarity measures. To offer a flexibility in practical applications, as well as to provide an opportunity for comparative evaluation of the algorithm, two distinct similarity measurement methods are introduced below. In addition, inspired by the most recent work of (Yang *et al.*, 2021), Locally linear embedding (LLE) (Roweis & Saul, 2000) can be regarded as an equivalent mechanism of weight calculation, which will also be discussed herein. Any one of these approaches may be employed to implement the intended task, but these are collectively reviewed to facilitate comparison, in an effort to make an informed choice of the potentially most suitable.

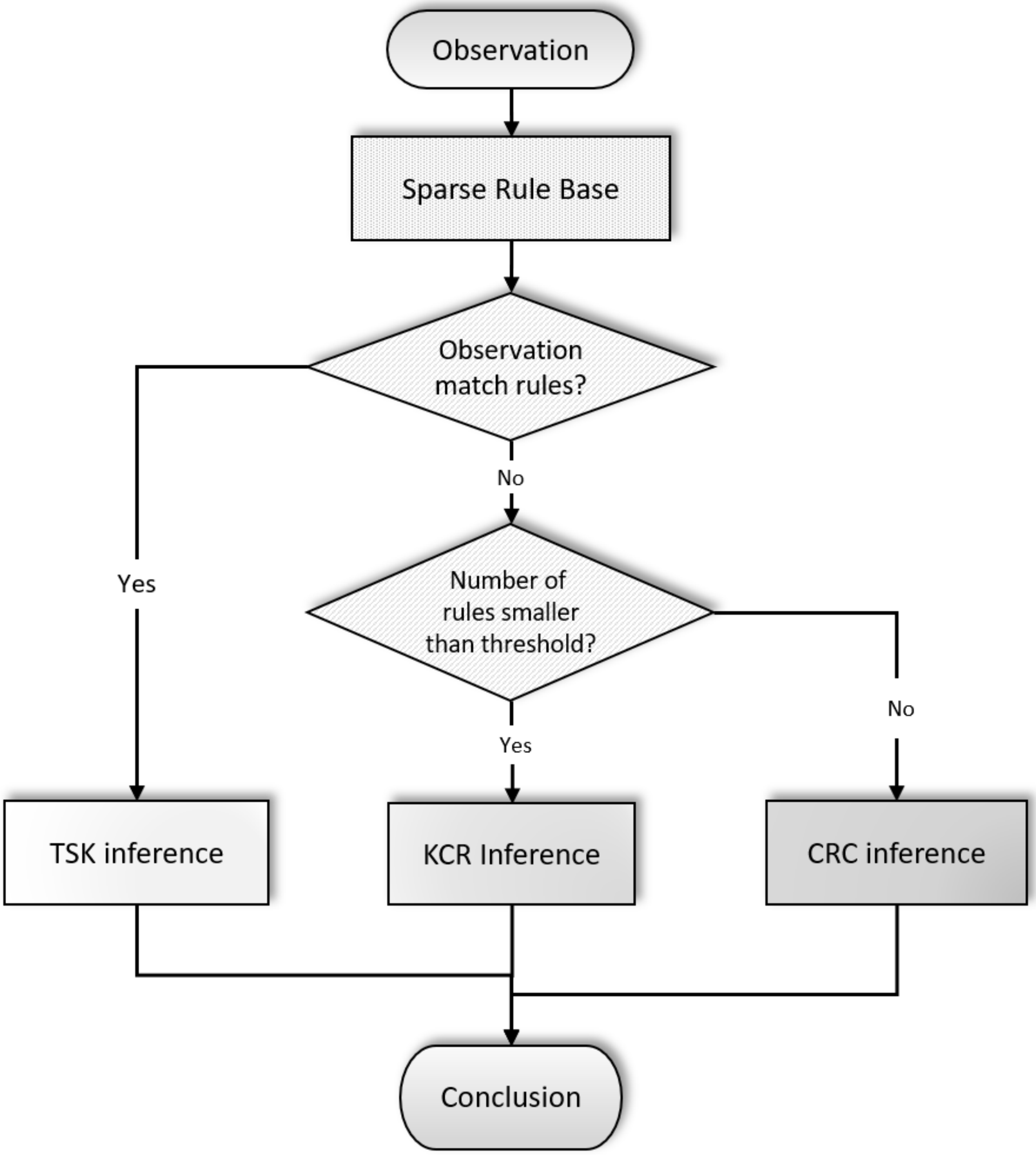

Figure 3.3: TSK fuzzy inference enhanced with two rule interpolation procedures.

3.3.1 Similarity Based on Distance Measures (Similarity-d)

Typically, the weight of each selected rule is produced by integrating the similarities between the observation features and the corresponding rule antecedents, both of which are represented by fuzzy sets. Similarity-d is one of the most widely applied similarity measurement methods (Chen, 1996), it calculates the similarity between two fuzzy sets directly based on the distance metric between them.

For illustration, suppose that two normalised fuzzy sets are given, represented by triangular membership function $A = (a_1, a_2, a_3)$ and $A' = (a'_1, a'_2, a'_3)$, respectively. Then, the implementation can be defined as the follows:

$$S(A, A') = \frac{1}{1 + d(A, A')} \tag{3.5}$$

where, $d(A, A') = \frac{\Sigma_{i=1}^{3}|a_i - a'_i|}{3}$.

Following the above specifications of these similarity measurement methods, it is clear that the larger the value of $S(A, A')$, the nearer and hence, the more similar the two fuzzy sets A and A'. In particular, $S(A, A')$ reaches the maximum value if and only if A and A' are identical.

3.3.2 Similarity with Distance Factor (Similarity-DF)

Whilst providing acceptable results, the above two approaches may still not be sufficiently flexible to support various fuzzy models, as the sensitivity of similarity degree to distance is fixed. To improve the distance sensitivity of above measurements, a centre of gravity method (COG) based similarity measure is proposed in (Chen & Chen, 2003). That is:

$$S(A, A') = (1 - \frac{\Sigma_{i=1}^{3}|a_i - a_i'|}{3}) \cdot (1 - |a_A^* - a_{A'}^*|) \tag{3.6}$$

where a_A^* and $a_{A'}^*$ is the COG of the fuzzy set A and A' which can be calculated by:

$$a_A^* = \frac{a_1 + a_2 + a_3}{3}, a_{A'}^* = \frac{a_1' + a_2' + a_3'}{3} \tag{3.7}$$

In the recent work of (Li *et al.*, 2017), this approach is further extended with a novel similarity measure introduced in (Shi-Jay *et al.*, 2003). Particularly, a distance factor (DF) is utilised to replace the COG function, which is defined as the following:

$$\begin{aligned} S(A, A') &= (1 - \frac{\Sigma_{i=1}^{3}|a_i - a_i'|}{3}) \cdot DF \\ DF &= 1 - \frac{1}{1 + e^{-sd+5}} \end{aligned} \tag{3.8}$$

where d represents the Euclidean distance between the gravity centres (or alternatively, representative values (Huang & Shen, 2008)) of the two fuzzy sets, s is a sensitivity factor (a larger value of s makes DF more sensitive to the distance measure), and β is a significantly large integer (which is empirically set to 5) to ensure that DF is approximately normalised as 1 when d is 0. The greater the value of $S(A, A')$ is, the closer and more similar the two fuzzy sets A and A' is.

Owing to their generality, both can be effective and applicable to capture and reveal the similarities between fuzzy sets (as to be experimentally verified later, while Similarity-DF generally performs better than Similarity-d).

3.3.3 LLE-Based Determination

Locally linear embedding (LLE) (Roweis & Saul, 2000) was originally developed to tackle dimensionality reduction problems. It provides an efficient nonlinear algorithm for mapping high-dimensional data to a low-dimensional observed space, which extends its capability to perform different types of machine learning tasks (e.g., (Chang *et al.*, 2004); (Rongrong *et al.*, 2017)).

As per the most recent work of (Yang *et al.*, 2021), LLE provides a mechanism similar to that of FRI. The space of rule antecedents can be regarded as the high-dimensional space and the consequents domain as the low-dimensional space. It has been successfully adopted to compute the similarities between the selected rules and the observation, thus constructing an adaptive neuro-fuzzy inference system (ANFIS) (Jang, 1993) through interpolation procedure. The LLE-based rule weight determination applied in this thesis works on the basis of the hypothesis which indicates that if an instance can be represented by a linear combination of the antecedences of its nearest rules, the required conclusion can be obtained by a linear combination of the corresponding rule consequents. The parameters of the linear function are deemed to be the weights of corresponding rules. The procedure of computing the weight of the selected rule is summarised in Alg. 10. This summary is given for the completeness, but further details regarding this algorithm can be found in (Roweis & Saul, 2000).

Algorithm 10: LLE-Based Rule Weight Determination

Input: Selected closest rules $\{R_i\}$; Unmatched observation O

Output: Interpolated conclusion $f(O)$

1. Represent observation by a linear combination of antecedences of $\{R_i\}$: $B_j = \Sigma_{i=1}^{K} w_i A_{ij}$
2. Compute W_i by minimizing error function:

$$E(W_{ji}) = \Sigma_{j=1}^{n}(\|B_j - \Sigma_{i=1}^{K} W_i A_{ij}\|)^2, s.t.\Sigma_{i=1}^{K} W_i = 1 \qquad (3.9)$$

3. Construct intermediate rule regarding W_j and corresponding rule consequences with following parameters for its consequent:

$$a_0 = \Sigma_{i=1}^{K} W_i a_{i0}, ..., a_n = \Sigma_{i=1}^{K} W_i a_{in} \qquad (3.10)$$

4. Take observation O as input to interpolated rule and calculate final consequent outcome: $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$
5. Return: $f(B_1, ..., B_n)$.

Chapter 4

Experimental Investigation of FRI for Static TSK Models

The performance of the above-introduced novel FRI approach for TSK models is experimentally evaluated in this chapter, in comparison to the state-of-the-art techniques, including the TSK+ and AutoRS (Chen *et al.*, 2019), over ten benchmark datasets. The robustness and effectiveness of the presented approach are also demonstrated by observing the consistency and efficiency of utilising different clustering methods in supporting CRC. In particular, for reasoning accuracy apart from fairness in comparison, three aforementioned rule weight determination approaches are systematically compared with each other at first, and the overall winner will be employed to determine the weight of selected rules in subsequent experiments.

The remainder of this chapter is structured as follows. Section 4.1 describes the setting of the experiments carried out. Section 4.2 presents and discusses the experimental results, in comparison with state-of-the-art alternatives. Finally, Section 4.3 concludes the chapter and points out interesting future research.

4.1 Experimental Setup

This section presents the necessary setting for conduction experimental investigations, including the details of the ten applied benchmark datasets, the criteria for evaluating performance, the generation of the sparse rule bases, and the parameters for implementing KCR and CRC.

Moreover, to ensure the proposed approach does not rely on any specific clustering algorithm, the proposed implementation for CRC is systematically evaluated using five different clustering techniques (Kmeans, GMM, FCM, K-FCM and S-FCM) to conduct rule clustering.

4.1.1 Datasets Used

The datasets run include one nonlinear artificial model and nine real-world datasets (for regression problems) that have been taken from the UCI machine learning (Dua & Graff, 2017), function approximation (Guvenir *et al.*, 2000) and evolutionary learning repositories (Alcalá-Fdez *et al.*, 2011). The details of these employed benchmark datasets are summarised in Table 4.1. Note that the Polynomial dataset in the table is produced by randomly sampling from the following 3-dimensional nonlinear function:

$$F(x, y) = sin\frac{x}{\pi} \cdot sin\frac{y}{\pi} \tag{4.1}$$

This nonlinear function has been used to produce a benchmark dataset in (Li *et al.*, 2018b) and (Rezaee & Zarandi, 2010), and the random sampling method has been frequently employed in the literature (e.g., (Li *et al.*, 2017); (Bellaaj *et al.*, 2013)).

Table 4.1: Details of benchmark datasets run

Datasets	Features	Instances	Output Domain	Generated Rules
Dee	6	365	[0.765853,5.11875]	20
AutoMPG6	5	392	[9.0,46.6]	40
Stock	9	950	[34,62]	90
Laser	4	993	[0.0, 255.0]	100
Friedman	5	1200	[0.6640, 28.5903]	120
Polynomial	2	2000	[-1,1]	200
Quake	3	2178	[5.8,6.9]	200
Delta_ail	5	7129	[-0.0021, 0.0022]	500
Delta_elv	6	9517	[-0.014, 0.013]	1000
Pole	26	14998	[0.0,100.0]	1500

4.1.2 Performance Evaluation Criteria

To enable thorough evaluation and fair comparison, experimental results are obtained by the average and variance of 10×10-fold cross-validation on all the ten above introduced datasets. Training sets are used to generate sparse fuzzy rule bases (see next) while testing sets assess the performance described by RMSE (root-mean-square error, in relation to the ground truth). The smaller the value of RMSE is, the higher accuracy the approach has. Note that the training sets (the learned rule bases) and testing sets are exactly the same for the experiments on all approaches.

4.1.3 Sparse Rule Base Generation

In this experimental study, sparse fuzzy rule bases are artificially created from the dense fuzzy rule bases that are induced from the original datasets. This allows a fair and firm ground upon which to investigate the experimental results. In particular, a sparse rule base is generated by randomly removing a number of rules from the original dense rule base that has been learned by employing a data-driven learning approach. To emphasise on the sparsity of the knowledge available, in order to compare against conventional approximate reasoning and state-of-the-art FRI mechanisms (both running on TSK models), only 80% rules are retained to form the sparse fuzzy rule base for each problem case.

The following simple data-driven fuzzy rule learning procedure is employed to generate the original dense rule base: instances in a given training dataset are clustered into different categorises using fuzzy c-means (Bezdek *et al.*, 1984). Since fuzzy c-means allows an instance to belong to more than one cluster with different membership values, the worst rule-learning assumption is made here, with the least biased threshold of 0.5 membership value used to determine whether an attribute is taking on a certain fuzzy set as its value. The polynomial consequent of an emerging rule is learned through the popular linear regression approach as described in (Freedman, 2009).

For computational simplicity as well as for fair comparison, only triangular membership functions are used throughout to represent the fuzzy values of the antecedent attributes in the rules (and also for the observations) unless otherwise stated. The three parameters of a triangular membership function are implemented by the infimum, centre and supremum of the corresponding cluster. Of course, if fine-tuned membership functions are available and used, the performance of all interpolation approaches can be expected to have further improvement. The number of rules in the created sparse fuzzy rule base for each dataset is also listed in Table 4.1. Note that for easing illustration the threshold for determining large-sized sparse rule bases is set to be 95, in order to help evaluate the performance of KCR and CRC in relation to the sizes of the sparse rule bases concerned.

For easy illustration, consider the sparse rule base generated from the polynomial dataset. The following are two random examples of the learned rules:

- **If** x_1 is $A_1 = (0.092, 0.159, 0.227)$ and x_2 is $A_2 = (0.727, 0.767, 0.808)$, **then** $f_i(x_1, ..., x_n) = -1.759x_1 + 0.468x_2 + 0.029$
- **If** x_1 is $A_1 = (0.437, 0.455, 0.472)$ and x_2 is $A_2 = (0.388, 0.413, 0.438)$, **then** $f_i(x_1, ..., x_n) = -1.214x_1 - 0.764x_2 + 1.441$

Running the TSK model consisting of the 200 such rules (of the sparse rule base) without FRI is illustrated in Fig. 4.1 and Fig. 4.2. As can be seen, there are substantial amounts of space that are not covered by the learned rules, indicating that plenty of observations have matched no rule, resulting in missing values in the output domain.

Running the TSK model consisting of 200 such rules (of the sparse rule base) without FRI is illustrated in Fig. 4.1 and Fig. 4.2, reflecting the inference results drawn from a different inspection angle. In particular, Fig. 4.1 gives a sideview of running on the entire sparse rule base, and Fig. 4.2 shows a bird's-eye view. As can be seen, there are substantial amounts of space that are not covered by the learned rules, indicating that plenty of observations have matched no rule, resulting in missing values in the output domain. These two sub-figures collectively demonstrate the poor outcome of just exploiting the incomplete knowledge in the given problem domain, without the support of FRI.

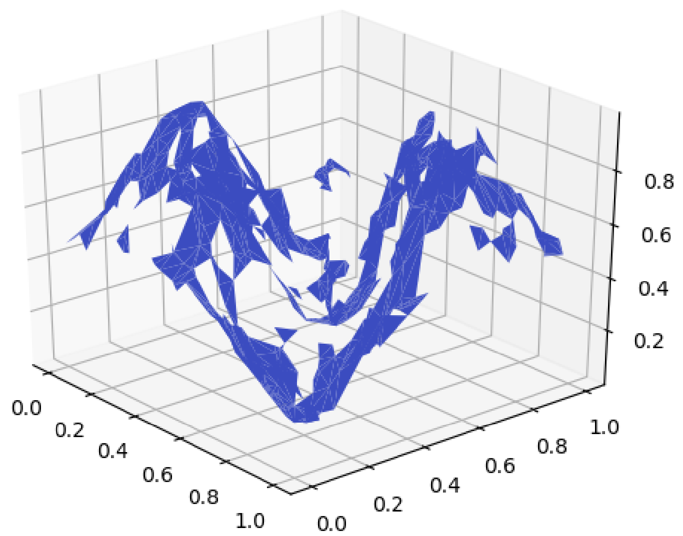

Figure 4.1: Inference results running TSK model on sparse rule base without FRI on Polynomial dataset: sideview.

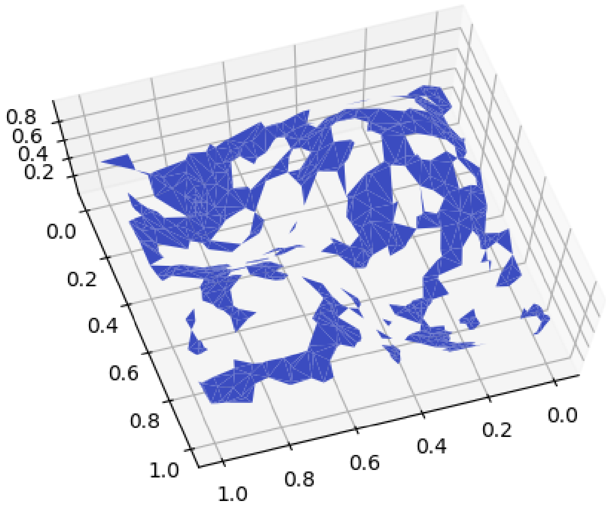

Figure 4.2: Inference results running TSK model on sparse rule base without FRI on Polynomial dataset: bird's-eye view.

4.1.4 Algorithmic Parameters

For completeness, the parameters used to implement KCR and CRC for the experiments on different datasets are listed in Table 4.2. Note that the number K of selected rules for both KCR and CRC is determined by a trial and error process. Particularly, the processes of determining the K for KCR and CRC are exemplified on the polynomial dataset, as shown in Fig. 4.3 and Fig. 4.4, respectively. For CRC, the Elbow method is employed to determine the number of clusters required.

Table 4.2: Setting of KCR and CRC for different datasets

Datasets	KCR	CRC	
	Selected rules(K)	Rule clusters	Selected rules(K)
Dee	3	5	2
AutoMPG6	3	5	2
Stock	3	5	2
Laser	3	10	3
Friedman	4	10	3
Polynomial	3	20	3
Quake	3	20	4
Delta_ail	5	50	5
Delta_elv	9	80	5
Pole	4	100	5

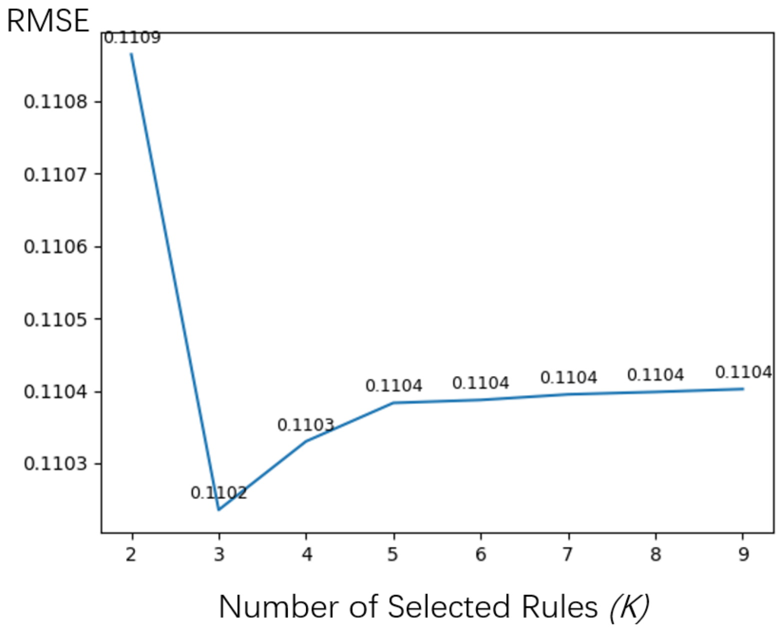

Figure 4.3: Model RMSE vs. number of rules K on KCR.

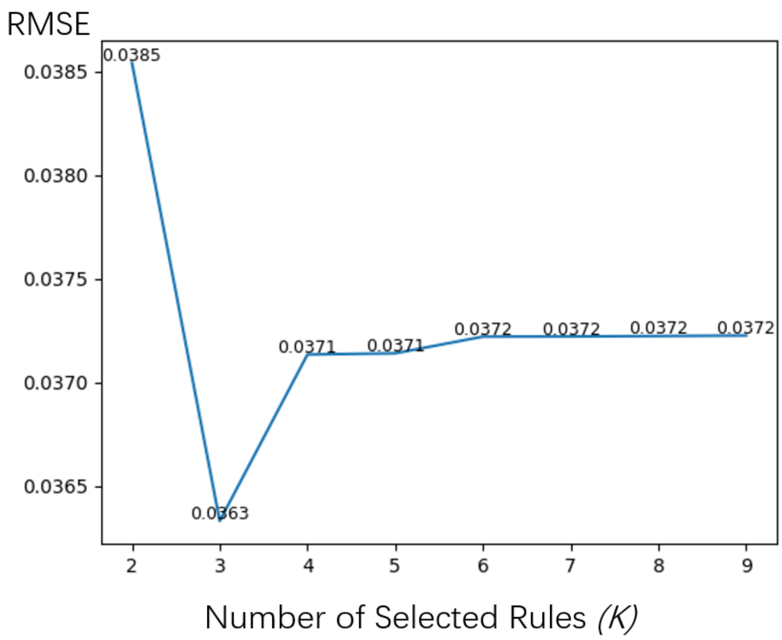

Figure 4.4: Model RMSE vs. number of rules K on CRC.

4.2 Results and Discussion

The results of systematic comparative experimental investigations over ten benchmark datasets are illustrated and analysed in this section. The performance of the proposed FRI approach is discussed with respect to the following four aspects, respectively: the outcomes of the three rule weight determination methods; the results on the small-sized datasets; the results on the large-sized datasets; and the analysis of several interesting observations.

4.2.1 On Use of Different Rule Weight Determination Methods

Table 4.3 presents the means and standard deviations of the interpolation results produced by KCR and CRC (with FCM as representative), with each supported by the use of either of the two similarity measurement methods and the LLE-based determination method introduced previously.

As can be seen, the FRI methods employing Similarity-DF can produce better results on all bar one dataset than those using Similarity-d, and on all datasets than those with LLE-based method. The effectiveness of utilising Similarity-DF is more obvious on more complex datasets. Note that KCR with Similarity-d leads to the best inference outcome on the dataset Dee. One possible reason for this is that selected rule have similar weights concerning the conclusion whilst their distances to the observation are different. Having taken notice of these results, in the following presentation of the experimental investigations, only Similarity-DF is used.

Table 4.3: Performance of employing different rule weight determination methods in KCR and CRC (with FCM as representative)

Datasets		KCR			CRC with FCM		
		Similarity-d	LLE	Similarity-DF	Similarity-d	LLE	Similarity-DF
Small	Dee	**0.367 ± 0.130**	1.196 ± 1.25	0.539 ± 0.204	0.498 ± 0.207	1.019 ± 0.806	0.623± 0.304
	AutoMPG6	4.010 ± 2.716	5.429 ± 2.826	**3.419 ± 1.384**	3.627 ± 3.161	5.687 ± 2.501	3.667 ± 1.503
	Stock	2.703 ± 2.458	2.271 ± 0.985	**1.099 ± 0.362**	8.095 ± 6.604	1.955 ± 0.796	1.144 ± 0.402
Large	Laser	14.98 ± 10.20	29.82 ± 23.84	12.45 ± 6.51	22.76 ± 20.30	28.76 ± 22.34	**11.29 ± 5.95**
	Friedman	2.909 ± 1.712	4.152 ± 2.809	3.400 ± 2.110	1.654 ± 1.033	2.291 ± 1.810	**1.615 ± 1.204**
	Polynomial	0.083 ± 0.011	0.092 ± 0.037	0.088 ± 0.034	0.245 ± 0.091	0.073 ± 0.030	**0.037 ± 0.011**
	Quake	1.414 ± 0.764	2.784 ± 1.148	1.152 ± 0.897	1.638 ± 3.129	1.923 ± 1.334	**0.575 ± 0.479**
	Delta_ail	1.516 ± 0.582	2.961 ± 0.553	0.860 ± 0.630	1.477 ± 0.357	2.706 ± 0.894	**0.358 ± 0.243**
	Delta_elv	2.880 ± 2.012	4.254 ± 2.075	2.175 ± 1.872	2.420 ± 1.247	3.157 ± 2.068	**0.327 ± 0.280**
	Pole	101.67 ± 42.99	110.50 ± 48.66	90.26 ± 23.48	50.34 ± 21.75	60.82 ± 30.80	**13.70 ± 12.75**

4.2.2 On Small-sized Rule Bases

Table 4.4 shows the means and standard deviations of interpolation results, which are averaged outcomes over 10×10-fold cross-validation, for each of the eight compared approaches on the datasets that involve the number of rules being less than or equal to 90.

Obviously, in this table, the notion of 'CRC with C', with C being K-means, GMM, FCM, K-FCM and S-FCM standing for the CRC algorithm implemented with the corresponding clustering method to create the rule clusters, respectively. The comparison with the conventional TSK models is not included herein, due to TSK models alone can not derive any conclusion when an observation does not match any of the rules in the rule base. Naturally, all rule interpolation methods significantly outperform the direct utilisation of the TSK models involving a sparse rule base across all problems.

As can be seen from Table 4.4, regarding those models comprising a sparse rule base of a small size, on average, the overall best results are obtained by KCR. AutoRS also produces similar results, one possible reason is that the rule selected by AutoRS and KCR for interpolation are similar in small-sized rule bases. Note that as introduced previously, the time complexity of AutoRS is much higher than KCR. TSK+ has slightly lower accuracies, which is due to the fact that all rules are involved, bringing forward adverse biases. It can also be observed that TSK+ takes a longer computation time to obtain the reasoning outcomes. This empirically verifies the fact that the proposed approach enjoys a significantly lower time complexity than TSK+, successfully reducing computation overheads.

Table 4.4: Performance of different rule selection methods on small-sized sparse rule bases

Datasets	RMSE (Mean ± Standard deviation)							
	TSK+	AutoRS	KCR	CRC with Kmeans	CRC with GMM	CRC with FCM	CRC with K-FCM	CRC with S-FCM
Dee	0.572±0.222	0.554±0.201	**0.539±0.204**	0.620±0.302	0.623±0.302	0.623±0.304	0.597±0.207	0.597±0.258
AutoMPG6	3.467±1.419	**3.267±1.313**	3.419±1.384	3.681±1.551	3.645±1.470	3.667±1.503	3.579±1.477	3.642±1.514
Stock	1.141±0.390	1.149± 0.385	**1.099±0.362**	1.131±0.402	1.116±0.411	1.144±0.402	1.137±0.435	1.129±0.409

Independent of which of the five clustering methods is used, the CRC algorithm does not work well on small-sized sparse fuzzy rule bases. This can be expected as those potentially highly relevant rules are most likely to have been clustered into one single cluster whilst the other clusters of rules offer little useful information to the conclusion. Thus, all rules bar one contribute misleading information to the calculation of the final results, leading to inaccurate interpolated outcomes.

4.2.3 On Large-sized Rule Bases

Table 4.5 presents the means and standard deviations of interpolation results returned by each of the eight compared approaches, regarding the seven TSK models that involve sparse rule bases of a large size. All algorithms are ranked according to their means and standard deviations of the results on each dataset, while those of the same results are deemed to have the same rank. The lower the ranking value the higher the model accuracy. In summary, the bottom row of this table provides the total rank values calculated by the sum of individual ranks across all seven benchmark datasets.

Clearly, CRC outperforms the rest consistently for such more complex datasets, with CRC implemented with FCM and CRC with K-FCM, for rule clustering, performing the best. Examining the experimental results more closely, CRC supported by K-FCM has highly impressive results on the Quake, Delta_ail and Delta_elv datasets, significantly improving the accuracy of fuzzy interpolative reasoning. Interestingly, this indicates that the kernel function (Eqn. 2.22) applied in K-FCM has the ability to make the clustering results more suitable for the distribution of the corresponding sparse rule bases.

Table 4.5: Performance of different rule selection methods on large-sized sparse rule bases

Datasets	RMSE (Mean ± Standard deviation)							
	TSK+	AutoRS	KCR	CRC with Kmeans	CRC with GMM	CRC with FCM	CRC with K-FCM	CRC with S-FCM
Laser	13.85±6.21	12.97±6.68	12.45±6.51	**11.21±6.05**	11.51±6.04	11.29±5.94	12.03±6.26	11.79±6.84
Friedman	3.325±2.02	3.632±1.948	3.400±2.110	1.753±1.151	1.567±1.0002	1.615±1.204	1.528±1.168	**1.522±1.112**
Polynomial	0.119±0.040	0.103±0.040	0.088±0.034	0.038±0.009	**0.037±0.011**	**0.037±0.011**	0.088±0.055	0.105±0.062
Quake	1.155±0.750	1.147±0.895	1.152±0.897	0.663±0.474	0.683±0.523	0.575±0.479	**0.408±0.319**	0.497±0.396
Delta_ail(e-4)	1.788±1.449	0.951±0.604	0.860±0.630	0.357±0.230	0.359±0.241	0.358±0.243	**0.243±0.161**	0.359±0.246
Delta_elv(e-4)	4.778±1.582	7.216±3.256	2.175±1.872	0.525±0.472	0.602±0.208	0.327±0.280	**0.299±0.230**	0.360±0.422
Pole	100.97±73.54	75.50±26.15	90.26±23.48	39.24±13.06	45.59±15.90	**13.70±12.75**	15.50±14.75	19.31±18.98
Rank	52	47	48	24	27	16	16	24

As reflected by these experimental outcomes, the utilisation of any of the five clustering algorithms enables CRC to outperform the other algorithms that do not involve clustering on the sparse rule bases, when these rule bases are of significant size. This demonstrates the significance of clustering-aided fuzzy rule interpolation.

Indeed, these results positively reflect the intuition that similar-rule crowds do have a negative impact on the accuracy of the final interpolated conclusions. With the support of clustering, CRC successfully avoids involving otherwise far too many similar rules and extends the diversity of rules used for subsequent interpolation. In addition, the results also confirm that although all (the sparse) rules are involved in deriving the final outcome, TSK+ does not cope with the problem well. Importantly, the narrow-banded standard deviation values given in Table 4.5 further demonstrate that the performance of the proposed approach is robust.

4.2.4 Further Examination on Performance

In general, CRC performs very well for a large-sized sparse rule base, as demonstrated above. However, to systematically exploit the proposed approach as per Fig. 3.3, a mechanism is required to determine the threshold at which to decide on whether KCR or CRC is to be used (although in the event when it is unrealistic to identify such a threshold, both methods may be applied to provide suggestions that are still useful for interpolative decision-making). According to Table 4.1, on the stock and laser datasets, the sparse rule bases include 90 or 100 rules. Consider Tables 4.4 and 4.5, by comparing against the outcomes achieved on the other datasets, algorithms with and without clustering have more similar results on these two datasets. Therefore, borrowing the underlying idea of the Elbow method, the threshold of large-sized sparse rule bases can be determined as at least including 95 rules (the average number between 90 and 100).

Apart from the issue of determining the threshold, there are occasional situations where the generality of CRC outperforming KCR for large datasets does not necessarily hold. This is evident by considering the cases where CRC supported by K-FCM or S-FCM is utilised to carry out interpolative reasoning with the sparse rule bases which are induced from the polynomial dataset. Fig. 4.5 depicts the distribution of the results in a box-plot, across all approaches investigated. As revealed by this figure, CRC with K-FCM or S-FCM is significantly underperformed than its peers, in comparison to CRC with K-means, GMM or the original version of FCM.

To analyse the causes of such performance deviation, the rules selected for a given observation on the polynomial dataset are outlined in Table 4.6, represented by their index. To analyse the causes of such occasional performance deviation, the rules selected for a given observation on the polynomial dataset are outlined in Table 4.6, represented by their index.

Table 4.6: Index of selected rules (for one observation) on Polynomial dataset

Approaches	Index of selected rules
TSK+	All rules in sparse rule base
AutoRS	[36, 39, 41, 45, 21]
KCR	[36, 39, 41]
CRC with Kmeans	[36, 22, 59]
CRC with GMM	[36, 21, 57]
CRC with FCM	[36, 21, 57]
CRC with K-FCM	[36, 41, 21]
CRC with S-FCM	[36, 45, 56]

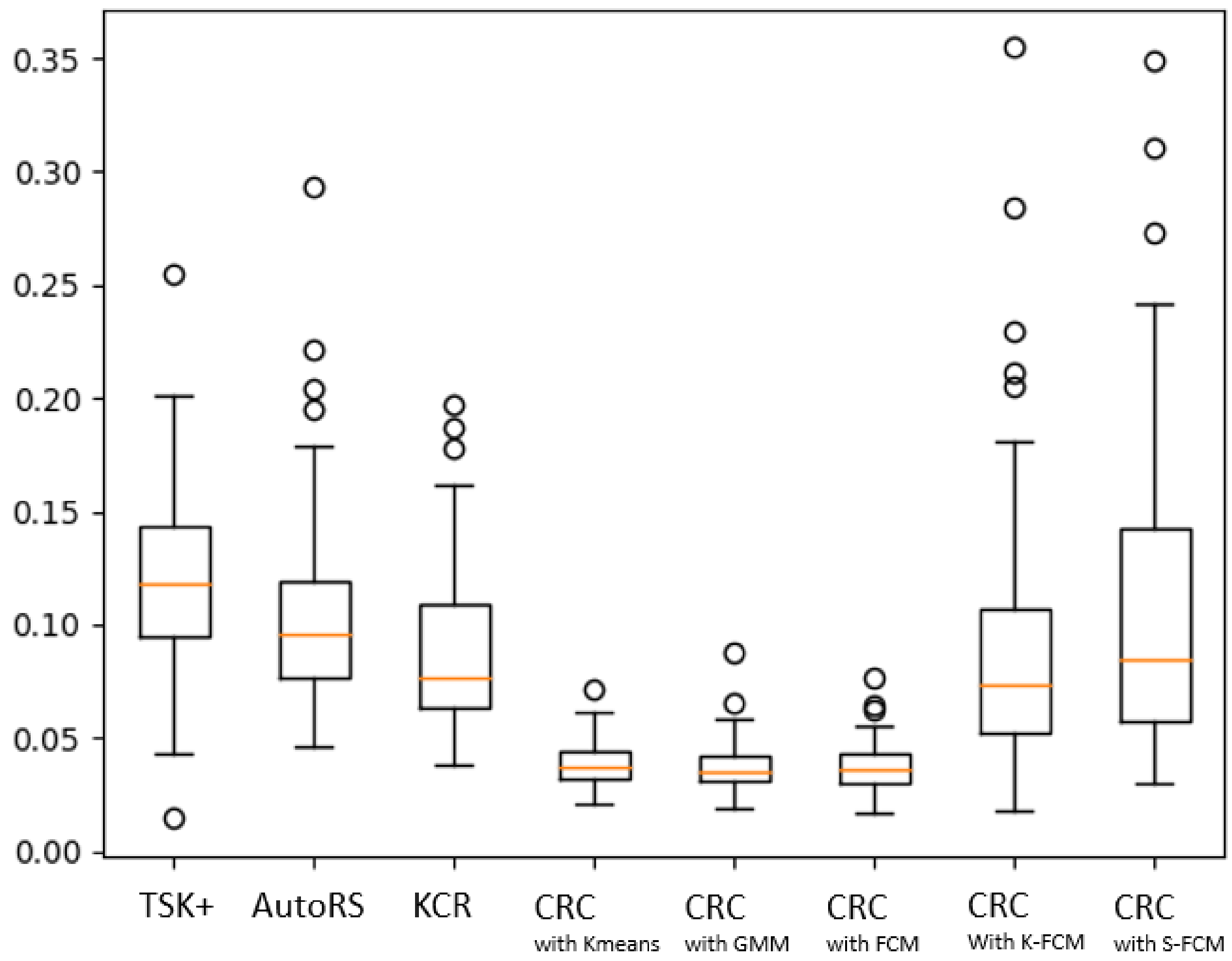

Figure 4.5: Box-plot of results over polynomial dataset.

It can be seen that rule 36 is the nearest rule to the particular observation and is always selected by all rule selection methods. Rules selected by K-FCM and S-FCM are clearly distinct from the ones that lead to satisfactory results. The likely reason for this is that these two clustering algorithms fail to derive appropriate rule clusters, adversely affecting the subsequent rule selection.

Another interesting observation is that what KCR selected is the top three rules also selected by AutoRS. However, with another two rules being selected further to those three AutoRS performs relatively worse than KCR.

Note that rule 21 is selected by AutoRS, which is also taken by the top performers like CRC with GMM or FCM (both of which happen to utilise the same selected rules). However, AutoRS underperforms in comparison to the two CRC implementations. This is probably due to the fact that those three (rules 36, 21 and 57) jointly offer the best information for producing accurate outcomes. Although rule 21 is also taken by AutoRS, it is being treated as one of the lowest weights (as it is taken the last amongst the five rules selected). Thus, its potential contribution is delimited, whilst adding more computation costs to reach the (less well) interpolated result. This reinforces the significance of the proposed approach.

To qualitatively visualise the strengths of the proposed approach, Fig. 4.6 and Fig. 4.7 illustrates the best inference results produced by the method without clustering (KCR) and that with clustering (CRC with FCM) over the polynomial dataset, respectively. Together with Fig. 5, it can be seen that the proposed approach enables appropriate conclusions to be generated for data instances unmatched by the sparse rules. Additionally, for this particular dataset, CRC with FCM can produce smoother results than KCR.

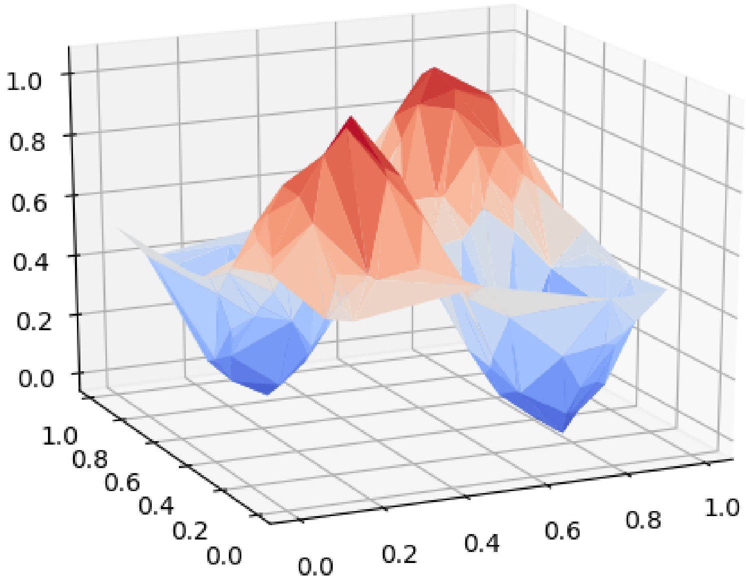

Figure 4.6: Inference results of TSK supported by KCR on polynomial dataset.

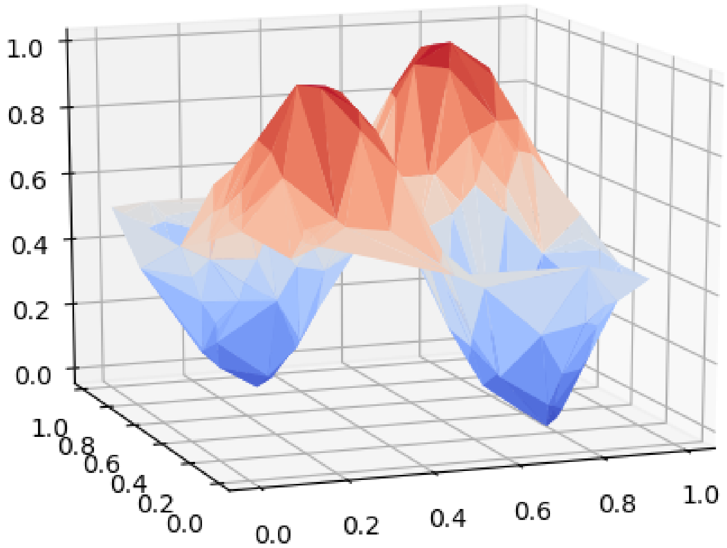

Figure 4.7: Inference results of TSK supported by CRC (with FCM) on polynomial dataset.

4.3 Summary

These two chapters have presented a novel approach with two implementations suitable for performing fuzzy rule interpolation (FRI) with TSK models. The work has been motivated by the observation that existing FRI approaches are almost completely devised for reasoning with Mamdani models, whilst those developed for TSK models are inefficient. It makes two major contributions to the literature: (i) For models involving small-sized sparse rule bases, the innovative KCR algorithm only requires the use of a small number of closest rules to derive accurate interpolation. (ii) For models with a large-sized sparse fuzzy rule base, the novel CRC algorithm first employs a clustering method to categorise the rules into groups and then, utilise one closest rule from each of a small number of the resulting rule clusters to perform the interpolation. Systematic comparative experimental studies over a range of benchmark datasets have demonstrated the efficacy of both implementations.

The proposed work offers many opportunities for further development. For example, the required algorithmic parameters, such as the number of closest rules and that of the nearest clusters, are currently empirically set manually; creating an automated method to determine these parameters from the training data requires significant further research. Also, AutoRS offers a means for automated selection of the number of closest rules for interpolation, how it may be integrated with the proposed approach to minimise human intervention is worth investigating. Furthermore, in real-world systems, the inputs are usually time-dependent, the requirements of fuzzy systems may change over time. Therefore, designing a novel system that can dynamically maintains the rule base is required.

Chapter 5

Monitoring System for Water Treatment Plant Supported by FRI

The water treatment plant dataset (Dua & Graff, 2017) comes from a real life monitoring system of an urban water treatment plant. It is collected by daily sensor measurements and utilised to establish a classification reasoning system to monitor the status of the plant without the benefit of an expert in the field. The dataset has been chosen for its prominence in the literature as a sample dataset to be used when tackling this type of problem, including that reported in (Shen & Jensen, 2004) and (Shen & Chouchoulas, 2002).

However, nearly all of the attributes in this dataset contain missing values, as they are usually resulting from human faults during storage, or other events hindering measurement (F *et al.*, 2013). This may lead to reduced reliability of the resulting classification systems. This problem remained unsolved in previously developed approaches.

As introduced in Chapter 2, several missing value imputation mechanisms have been presented in recent decades, with TSK model-based imputation being the most recently. However, this method may collapse if the resulting rule base is sparse (or incomplete), which is a common issue in real-world applications, especially for those involving missing values, such as the water treatment plant. Following the motivation for extending the applicability of the TSK model-based imputation and improving the accuracy of the subsequent classification, this chapter presents a novel fuzzy rule-based interpolative reasoning system for missing value imputation on the water treatment plant dataset.

The rest of this chapter is structured as follows. Section 5.1 introduces the water treatment plant and the feature selection approach to eliminate irrelevant attributes. Section 5.2 details the proposed missing value imputation approach, including the generation of the TSK system and the inference process of the missing value. Section 5.3 presents and discusses the experimental results, in comparison with a range of alternatives methods. Finally, Section 5.4 concludes the chapter and points out interesting future research.

5.1 Water Treatment Plant Dataset

The water treatment plant dataset collected historical data over 521 days, with 38 different input attributes measured per day. All measurements are continuous values. The group of measurements in the same day forms one instance. Fig. 5.1 demonstrates the following five aspects of the plant's operation reflected by 38 conditional features:

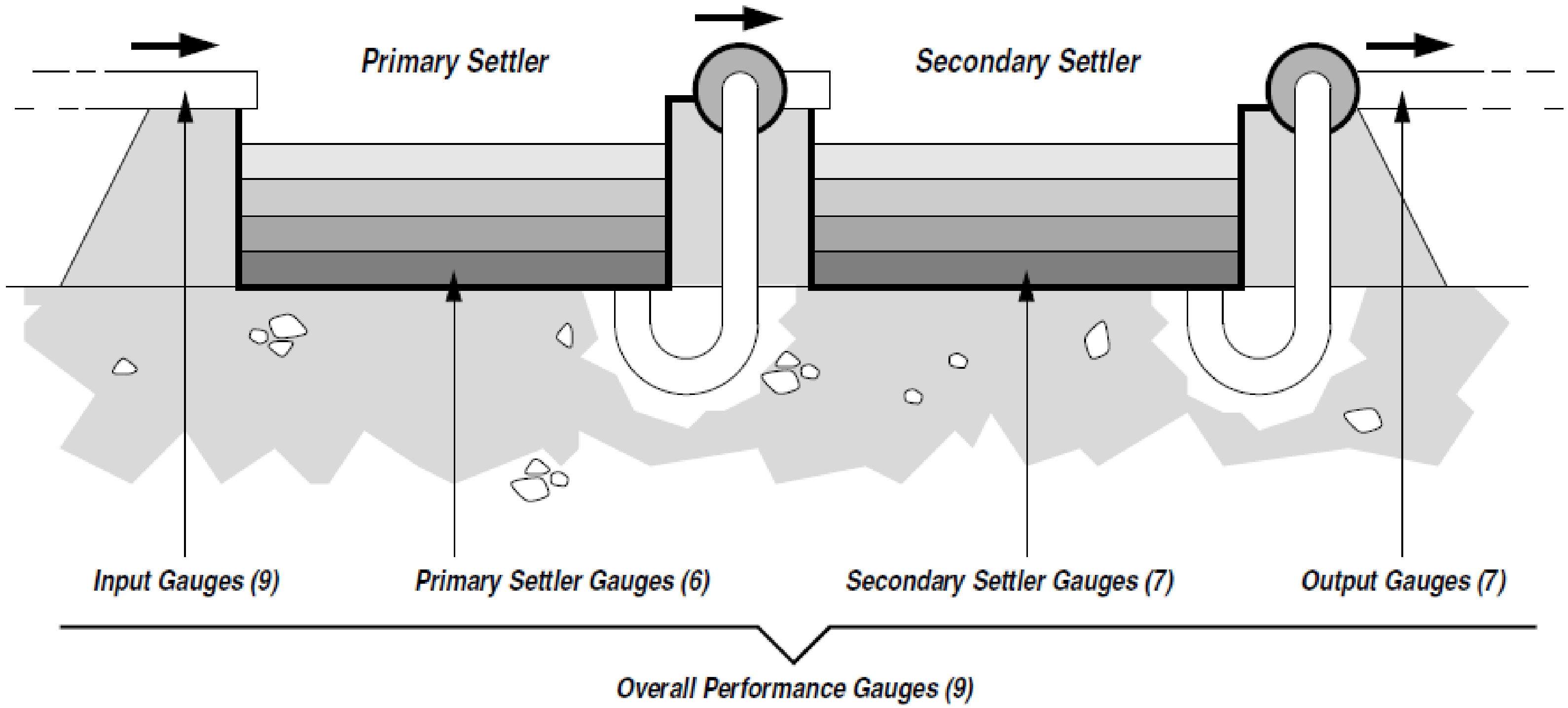

Figure 5.1: Schematic diagram of water treatment plant, showing sensors at different sampling points.

1. Input to plant (9 features)
2. Input to primary settler (6 features)
3. Input to secondary settler (7 features)
4. Output from plant (7 features)
5. Overall plant performance (9 features)

Each instance is classified into one of thirteen categories according to the plant operation status of the day, including four types of normal states and nine distinct faulty states. As all data are taken from the actual plant, all faults are immediately resolved. Therefore, most faulty states contain only a few instances (usually only one). This is a clear obstacle to the construction of an effective monitoring system (Shen & Jensen, 2004). To rectify, or at least to reduce the adverse impact of this limitation, the thirteen statuses are collapsed into just two or three categories (Normal and Faulty, or Normal, Good and Faulty, respectively) to increase the number of faulty examples. The thirteen states of the water treatment plant and their corresponding categories are demonstrated in Table 5.1.

In summary, the resulting binary monitoring system consists of 507 samples for acceptable performance and 14 samples for malfunctions. The resulting ternary monitoring system is composed of 114 instances of 'Good' status, 393 instances of 'Normal' status and 14 instances of 'Faulty' status.

Table 5.1: States of water treatment plant and corresponding categories in binary and ternary models.

Statu	Instances	Status Description	Ternary	Binary
1	275	Normal situation	Normal	Normal
9	65	Normal situation, low influent		
11	53	Normal situation		
5	114	Good situation	Good	
2	1	Secondary settler, type-1	Faulty	Faulty
3	1	Secondary settler, type-2		
4	4	Secondary settler, type-3		
6	3	Solids overload, type-1		
7	1	Secondary settler type-4		
8	1	Storm, type-1		
10	1	Storm, type-2		
12	1	Storm, type-3		
13	1	Solids overload, type-2		

5.1.1 Feature Selection

A large plant like the water treatment plant tends to collect a vast array of conditional features, many of which are of high similarity. However, in general, not all of these features are significant in determining the operational status of the plant. Therefore, a dimensionality reduction approach is usually required to remove redundant, or even possibly irrelevant attributes and simplify the design and implementation of the actual monitoring (classification) system. In a realistic system, these input features are obtained through a series of sensors, which also have a monetary cost, thus it is typically desirable to lower the input quantity.

The fuzzy–rough set based feature selection (FRFS) approach (Jensen &

Shen, 2004) (Jensen & Shen, 2009) is applied in this chapter to discover the dependence between subsets of features and to choose the most informative features. This approach defines a feature subset dependency function with the fuzzy positive region on the basis of fuzzy–rough sets, which is defined as follows:

$$\begin{aligned}
\gamma_P(Q) &= \frac{|\mu_{POS_{P(Q)}}(x)|}{|U|} = \frac{\Sigma_{x\in U}\mu_{POS_{P(Q)}}(x)}{|U|} \\
\mu_{POS_{P(Q)}}(x) &= \sup_{x\in U/Q} \mu_{\underline{P}X}(x) \\
\mu_{\underline{P}X}(x) &= \sup_{F\in U/P} min(\mu_F(x), \inf_{y\in U}\{\mu_F(y) \rightarrow \mu_X(y)\})
\end{aligned} \tag{5.1}$$

where U stands for all objects, $\gamma_P(Q)$ represents the dependence, $\mu_{POS_{P(Q)}}(x)$ denotes the membership of the object x belonging to the fuzzy positive region, and $\mu_{\underline{P}(X}(x)$ means the fuzzy lower approximation. This summary is given for the completeness, but further details regarding this algorithm can be found in (Shen & Jensen, 2004).

This function is employed to evaluate the relative importance degree of a certain subset of features to categories. This approach can also be utilised to calculate the degree of each individual feature by setting a single attribute in each feature subset.

Table 5.2 demonstrates the resulting selected features for the binary and ternary monitoring systems.

5.1.2 Generation of Fuzzy Classification Rule

An effective fuzzy rule induction algorithm (RIA), as proposed in (Au & Chan, 2002), has been successfully applied in the construction of the classification monitoring system for the water treatment plant dataset and used to obtain

Table 5.2: Selected rules for binary and ternary monitoring problems.

Class	Selected Features	Amount
2-Class	{0, 2, 6, 10, 12, 15, 22, 24, 26, 37}	10
3-Class	{2, 3, 6, 10, 12, 15, 17, 22, 27, 29, 37}	11

impressive results in the literature (Shen & Chouchoulas, 2002) and (Shen & Jensen, 2004).

The rule induction procedure divides each input feature into a range of corresponding linguistic terms and the output feature into several linguistic class labels. The instances are then categorised into subsets based on their labels with the highest value. The fuzzy subsethood (Kosko, 1986) (Yuan & Shaw, 1995) is utilised to represent the relationships between the decisions of the subset and the term of each feature in each subset, which is defined by:

$$S(L, A_i) = \frac{M(L \cap A_i)}{M(L)} = \frac{\Sigma_{u \in U} min(\mu_L(u), \mu_{A_i}(u))}{\Sigma_{u \in U} \mu_L(u)} \tag{5.2}$$

where $\mu_L(u)$ denotes the membership degree of the object u belonging to the class L, and $\mu_A(u)$ represents the overlap degree of the object u to the i-th term of feature A.

An empirically predefined threshold is applied herein to determine the terms for the final classification system. Following the above definition, the term with the highest subsethood value is selected. Note that the threshold must ensure that at most one term is selected from each feature.

The efficiency of RIA and FRFS has been empirically verified as superior to many existing possible alternatives for the water treatment plant dataset in (Shen & Chouchoulas, 2002) and (Shen & Jensen, 2004).

5.2 Missing Value Imputation with TSK Models Supported by FRI

It is observed that the water treatment plant dataset contains many missing values in a number of instances, which also appear in the reduced 2-class and 3-class datasets (118 and 112 instances containing missing values respectively). This is a common problem with real-world datasets that can arise from equipment failure, the limitation of data collection and other reasons, all of which decrease the reliability and efficiency of the classification system.

To rectify the adverse impact of missing values, and to address the limitations of existing methodologies, this chapter presents a novel missing value imputation approach with TSK models supported by FRI. The underlying principle is that features for classification problems are not independent. Thus, each feature can be calculated by a linear or nonlinear combination of other features. Fortunately, TSK models have the potential to represent such a model with a range of linear functions.

5.2.1 Fuzzy Rule Base for Missing Value Imputation

For each feature with missing values, a distinct fuzzy rule base of TSK models is established by employing a data-driven learning approach to the instances without missing values. Feature selection also helps to significantly reduce the complexity of this procedure.

The data-driven fuzzy rule learning procedure regards attributes with the missing values as outputs and others as inputs. It works by clustering instances into different categories through fuzzy c-means (Bezdek *et al.*, 1984) according to their input features. The rule antecedents are implemented by fuzzy sets representing input features within the same cluster. The corresponding rule consequent, which is a polynomial, is derived from the output feature in the category through linear regression (Freedman, 2009).

Borrowing the underlying idea of the Elbow method, the number of rules in the fuzzy rule base can be determined as ten rules. For easy illustration, taking feature 37 as the output, the following are two random examples of the learned normalised rules in 2-class and 3-class models, respectively:

- **If** x_0 is (0.19, 0.59, 1), x_2 is (0.31, 0.65, 0.99), x_6 is (0.39, 0.69, 1), x_{10} is (0.29, 0.59, 0.88), x_{12} is (0.10, 0.55, 1), x_{15} is (0.28, 0.59, 0.90), x_{22} is (0.11, 0.23, 0.34), x_{24} is (0, 0.21 0.42), and x_{26} is (0.37, 0.64, 0.92),
  then $f_X = 0.0081x_0 + 0.0270x_2 - 0.0013x_6 - 0.0312x_{10} - 0.0025x_{12} - 0.0431x_{15} - 0.0905x_{22} - 1.4768x_{24} - 0.0155x_{26} + 1.0579$

- **If** x_2 is (0.25, 0.53, 0.81), x_3 is (0, 0.33, 0.67), x_6 is (0, 0.40, 0.81), x_{10} is (0, 0.26, 0.52), x_{12} is (0.076, 0.44, 0.81), x_{15} is (0, 0.45, 0.90), x_{17} is (0, 0.30, 0.61), x_{22} is (0, 0.19, 0.38), and x_{27} is(0, 0.28, 0.57) x_{29} is(0.06, 0.53, 1),
  then $f_X = 0.0037x_2 + 0.0114x_3 + 0.0008x_6 + 0.0003x_{10} - 0.0131x_{12} + 0x_{15} + 0.0121x_{17} - 0.0039x_{22} - 1.5886x_{27} + 0.0004x_{29} + 0.9982$

5.2.2 FRI Approach for Missing Value Imputation

In a situation where only sparse rule bases are presented, which is a common issue in real-world applications, it is possible that a novel instance 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 (be they fuzzy or not). This issue is more serious for datasets containing missing values as only limited complete instances are available to train the model. In the water treatment plant status classification problem, 58 out of 118 instances match no rule in the binary models, and 60 out of 112 instances match no rule in the ternary models.

Inspired by the approach introduced in Chapter 4, this section presents an FRI approach to carry out missing value imputation for unmatched observations. According to the previously stated results, a rule base containing ten rules is considered small-sized. Therefore, the KCR procedure is applied to approximate missing values in the water treatment plant dataset.

The presented method works by first calculating the Euclidean distance between individual antecedent variables of each rule R_i and their counterparts in the observation, implemented via the COG of the corresponding fuzzy sets. K closest rules are then selected based on their distance to the observation. The weight of each selected rule in the observation is calculated by employing the similarity measure specified as per Eqn. 5.3. The individual conclusions of the K selected rules are produced by taking the observation as the input to the consequent polynomial of selected rules. Finally, the conclusion with respect to the unmatched observation can be derived through the integration of the obtained sub-conclusions, which is implemented by the weighted average in this chapter. A more detailed description can be found in Alg. 8.

Fig. 5.2 shows the proposed water treatment plant monitoring system, including training and application procedure. This system is derived from the integration of all the above-introduced approaches. The proposed missing value imputation approach makes three major contributions to the system: (i) filling the incomplete dataset with estimated values; (ii) improving the accuracy of the resulting classification system; (iii) generating a reasonable classification conclusion for a novel observation containing missing values in practice.

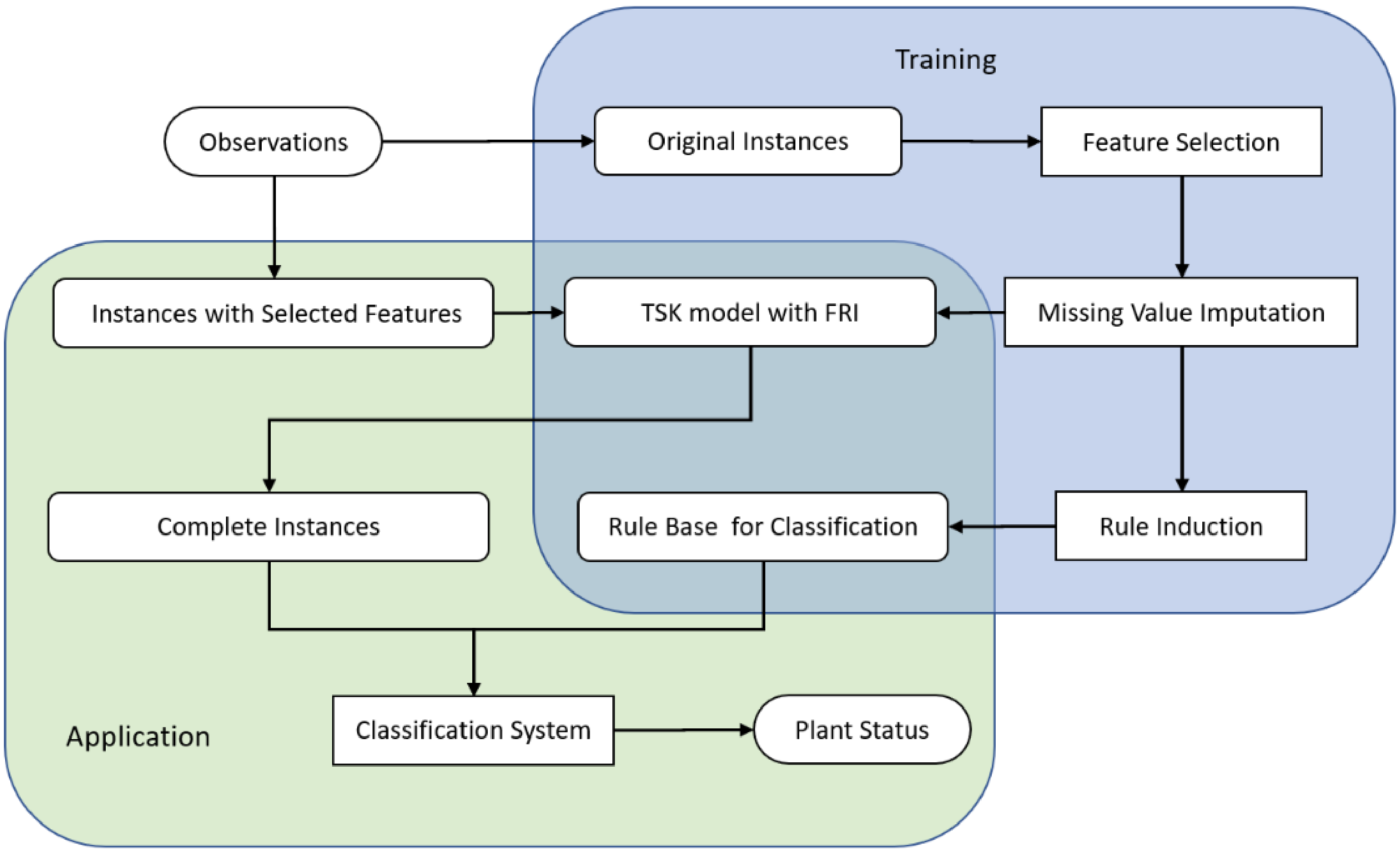

Figure 5.2: Workflow of proposed water treatment plant monitoring system.

5.3 Experimental Evaluation

The performance of the proposed novel missing value imputation approach is experimentally evaluated in this section, in comparison with several conventional techniques, such as mean, median imputation and linear regression imputation. In particular, to provide flexibility in implementing the proposed approach, ensuring that it does not rely on any specific FRI algorithm, the AutoRS based imputation is also systematically evaluated. To demonstrate the potential of the proposed method, the following experimental investigations focus on two major aspects: the accuracy of the predicted results and the improvement on the classification system.

5.3.1 On Missing Value Imputation

This section describes the setting of the experiments carried out in terms of the accuracy of the predicted conclusions, and discusses the corresponding results of comparative experimental evaluations.

For each feature with missing values, a specific TSK model that takes this feature as the output is derived from complete instances (without missing value) by employing a data-driven learning approach. Missing values in the corresponding feature are then approximated using the resulting model.

5.3.1.1 Performance Evaluation Criteria

To investigate the predicted result and enable fair comparison, the experimental results are represented by the average and variance of 10 $\times$ 5-fold cross-validation on complete datasets. There are 389 instances for binary systems and 395 instances for ternary systems. Training sets are used to generate the rule base of TSK models, and testing sets are employed to evaluate the performance based on the accuracy of the imputation in terms of the root-mean-square error (RMSE, in relation to the ground truth). The smaller the RMSE value, the more accurate the approach.

5.3.1.2 Results and Discussion

Table 5.3 and 5.4 show the means and standard deviations of the predicted results obtained by 2-class and 3-class models, respectively. The notions in the column of 'Output Feature' represent the specific missing value imputation methodologies constructed for the corresponding feature, whilst 'AutoRS' and 'KCR' stand for the approach with TSK models supported by corresponding FRI approaches. Note that the comparison with the TSK model-based method is not included herein. This is because TSK models alone can not draw any conclusions for an observation which does not match any of the rules in the rule base, as indicated previously.

As reflected by these experimental outcomes, the TSK systems supported by FRI approaches obtain the overall best results on average, which reveals the efficiency of the proposed approach. It is noticed that the results produced by the approaches with AuoRS and KCR are highly similar. One possible reason is that in such small-sized rule bases with only ten rules, the rules these two FRI methods select for interpolation are similar, leading to similar results. It can be expected that there will be a greater difference between them in large-sized rule bases. In addition, the regression imputation performs better when compares with the rest methods. This empirically verifies the fact that features for classification problems are not independent. One feature may be represented by a linear or nonlinear combination of other features.

Table 5.3: Performance of different missing value imputation approaches for case with 10 features and 2 classes.

Output Feature	RMSE (Mean ± Standard deviation)				
	Mean	Median	Regression	AutoRS	KCR
Feature 0	6727.56 ± 976.74	6849.20 ± 1075.91	6133.68 ± 738.26	**3765.88 ± 765.40**	3901.90 ± 804.21
Feature 2	0.233 ± 0.032	0.233 ± 0.033	0.133 ± 0.007	0.084 ± 0.019	**0.082 ± 0.017**
Feature 6	12.583 ± 1.679	12.832 ± 1.766	5.765 ± 1.411	3.712 ± 1.204	**3.635 ± 1.243**
Feature 10	72.786 ± 8.151	73.230 ± 8.482	66.155 ± 9.264	41.753 ± 6.901	**40.141 ± 7.935**
Feature 12	12.634 ± 1.863	12.820 ± 2.147	5.847 ± 1.894	**5.076 ± 1.851**	5.248 ± 1.898
Feature 15	0.192 ± 0.018	0.194 ± 0.018	0.096 ± 0.009	0.062 ± 0.012	**0.060 ± 0.010**
Feature 22	0.153 ± 0.019	0.154 ± 0.018	0.122 ± 0.007	0.087 ± 0.015	**0.082 ± 0.013**
Feature 24	35.183 ± 8.563	35.289 ± 8.688	29.623 ± 5.514	**15.658 ± 4.531**	17.179 ± 4.410
Feature 26	9.172 ± 1.128	9.221 ± 1.193	8.654 ± 1.034	**5.258 ± 1.438**	5.703 ± 1.479
Feature 37	3.094 ± 3.796	2.999 ± 3.916	3.857 ± 2.573	2.432 ± 2.944	**2.302 ± 2.252**

Table 5.4: Performance of different missing value imputation approaches for case with 11 features for 3 classes.

Output Feature	RMSE (Mean ± Standard deviation)				
	Mean	Median	Regression	AutoRS	KCR
Feature 2	0.239 ± 0.020	0.239 ± 0.019	0.135 ± 0.019	0.098 ± 0.031	**0.091 ± 0.025**
Feature 3	59.056 ± 8.262	59.191 ± 8.456	43.570 ± 9.775	**28.747 ± 10.306**	29.342 ± 10.638
Feature 6	12.595 ± 1.560	12.759 ± 1.772	5.886 ± 1.398	**3.904 ± 1.491**	3.982 ± 1.211
Feature 10	70.527 ± 5.425	70.896 ± 6.307	34.527 ± 7.211	23.904 ± 5.596	**21.382 ± 5.263**
Feature 12	12.645 ± 1.624	12.804 ± 1.779	**6.234 ±1.591**	6.844 ± 1.931	6.235 ± 1.591
Feature 15	0.194 ± 0.021	0.195 ± 0.020	0.105 ± 0.019	0.088 ± 0.022	**0.069 ± 0.025**
Feature 17	68.927 ± 6.618	68.871 ± 6.729	42.002 ± 4.855	29.893 ± 9.004	**27.663 ± 8.018**
Feature 22	0.170 ± 0.063	0.171 ± 0.063	0.141 ± 0.067	0.109 ± 0.066	**0.103 ± 0.062**
Feature 27	0.127 ± 0.173	0.124 ± 0.175	0.053 ± 0.061	0.033 ± 0.051	**0.030 ±0.052**
Feature 29	14.789 ± 1.506	14.813 ± 1.488	9.374 ± 1.359	**5.647 ± 1.817**	6.168 ± 1.577
Feature 37	3.172 ± 3.622	3.124 ± 3.714	1.382 ± 1.417	0.713 ± 1.379	**0.706 ± 1.300**

Examining the experimental results more closely, in the experiments on feature 12, the estimated value produced by the regression method and the proposed approach have results that are more similar in both the binary and ternary systems, than the results of the experiments on the other features. This may indicate that the relationship between feature 12 and other features will more likely fit a linear function. It also demonstrates that TSK models have the potential to simulate both linear and nonlinear models with a range of linear functions.

5.3.2 On Improvement of Classification System

To emphasise on the potential for strengthening the classification system, the proposed missing value imputation approach and other alternative methods are used to estimate all missing values in the dataset. Afterwards, a unique classification system can be derived for each corresponding resulting complete dataset. This section describes the setting of the experiments carried out and discusses the results of comparative experimental evaluations. For completeness, comparison with the delete method (which directly discards instances containing missing values) is also included.

5.3.2.1 Performance Evaluation Criteria

To demonstrate the improvement of the classification system, the missing values contained in instances are first estimated using a range of imputation approaches. The corresponding classification systems are then learned from the resulting complete datasets. To verify its generality over different classification methods, fuzzy rule classification and support vector machines (SVM) (Saunders *et al.*, 2002) will be applied to derive classification systems from obtained datasets. The performance of the systems with respect to each imputation approach is evaluated according to their F1-score on the training and testing sets (in relation to the actual class label), which are represented by the average obtained from 10 × 5-fold cross-validation. A larger F1-score denotes that the corresponding method can obtain higher reasonable results.

5.3.2.2 Results and Discussion

The means and standard deviations of the macro-F1 score of the classification results returned by each of the five compared approaches, regarding binary and ternary models, are illustrated in Table 5.5 and 5.6, respectively. The macro-F1 score applied herein is defined by:

$$\begin{aligned} F1 &= \frac{\Sigma_{i=0}^{n} F1_i}{n} \\ F1_i &= 2 \times \frac{Percision_i \times Recall_i}{Percision_i + Recall_i} \end{aligned} \tag{5.3}$$

where n represents the number of classes.

In these two tables, the notions in the column of 'Approaches' represents different missing value approximation methodologies used to obtain the complete dataset for deriving the subsequent classification system, whilst 'Delete' stands for the method that directly discards the instances containing missing values.

The results make it clear that for both SVM and fuzzy classification systems, the systems generated from the datasets whose missing values are estimated by the FRI0supported TSK model outperform those from other methods. This shows the potential of TSK-based missing value imputation approaches to improve the accuracy of the classification system, and demonstrates its general applicability to different classification systems.

Interestingly, it is also observed that, in certain situations, classification systems learned from datasets with missing values filled with mean and median

Table 5.5: Classification results for 2-class database whose missing value are processed with different approaches.

Approaches	Instances	Fuzzy Classification		SVM	
		Training	Testing	Training	Testing
Delete	389	0.853 ± 0.062	0.788 ± 0.198	0.841 ± 0.044	0.695 ± 0.189
Mean	507	0.859 ± 0.037	0.798 ± 0.115	0.866 ± 0.038	0.698 ± 0.141
Median	507	0.848 ± 0.042	0.738 ± 0.177	0.868 ± 0.051	0.725 ± 0.133
Regression	507	0.876 ± 0.056	0.815 ± 0.164	0.868 ± 0.071	0.741 ± 0.137
AutoRS	507	0.880 ± 0.027	0.854 ± 0.159	**0.885 ± 0.014**	**0.777 ± 0.153**
KCR	507	**0.884 ± 0.026**	**0.858 ± 0.151**	0.883 ± 0.022	0.770 ± 0.124

Table 5.6: Classification results for 3-class database whose missing value are processed with different approaches.

Approaches	Instances	Fuzzy Classification		SVM	
		Training	Testing	Training	Testing
Delete	395	0.671 ± 0.045	0.599 ± 0.141	0.657 ± 0.145	0.550 ± 0.204
Mean	507	0.651 ± 0.018	0.594 ± 0.122	0.649 ± 0.092	0.557 ± 0.090
Median	507	0.667 ± 0.021	0.611 ± 0.113	0.647 ± 0.046	0.521 ± 0.124
Regression	507	0.676 ± 0.020	0.641 ± 0.124	0.655 ± 0.090	0.561 ± 0.081
AutoRS	507	**0.681 ± 0.014**	0.659 ± 0.111	0.684 ± 0.091	0.566 ± 0.061
KCR	507	0.678 ± 0.018	**0.662 ± 0.107**	**0.686 ± 0.096**	**0.571 ± 0.058**

imputation may produce worse results than those using the delete method. This is probably due to the fact that in both methods, only values in the same attribute are involved in the missing value approximation, which may import undesirable bias into the results. Meanwhile, this also demonstrates the significance of considering the relationship between features in missing value imputation.

5.4 Summary

This chapter introduces a status monitoring system for the water treatment plant. The missing values contained in the dataset are estimated by TSK models with the support of a novel FRI approach. The systematic experimental studies have demonstrated that the proposed approach not only outperforms other methods in producing more accurate results, but also has the potential to help to improve the accuracy of the resulting classification system.

There are many opportunities for the future development of the method contained herein. This chapter considers only one missing value within each instance. However, in most real-world datasets, an instance may contain two or more missing values. In the case where two missing values exist in an instance is present, one possible approach is to estimate one value with the mean or median method, predict the other using the proposed approach, and then re-evaluate the first value using the predicted result. Thus, these two values can be iteratively updated until convex. Furthermore, the current work only involves the rule base of a small size (containing ten rules). How the other implementation of aforementioned novel FRI approach (CRC) may be adopted to perform missing value imputation for large-sized datasets can be further investigated.

Chapter 6

Dynamic TSK Systems Supported by Fuzzy Rule Interpolation

While performing fuzzy systems in real-world applications, the inputs are usually time-dependent and the requirements of fuzzy systems may change over time. If the frequently appearing unmatched observations are of high similarity, the use of a static rule base will repeat the similar FRI process and further affect the efficacy of fuzzy inference systems.

Therefore, dynamically maintaining the rule base is necessary to improve the efficacy of the system concerned through enhanced coverage of the rule base. Fortunately, FRI has the potential to provide additional information which is required to design a dynamic TSK system. This is due to most existing FRI approaches tend to produce a large amount of interpolated rules over time, which are generally discarded once the results are derived. Note that the goal of FRI is to compute an interpolated consequent that corresponds to the input observation rather than to produce an intermediate rule through interpolative reasoning. Such interpolated rules may contain potentially useful information, such as that reflecting the input-output relationships which were not covered by the original sparse rule base. Exploiting these interpolated rules may help update the original sparse rule base, thereby constructing a dynamic system.

This chapter presents a dynamic TSK system in extending the underlying ideas of the previous proposed FRI method for static TSK models and D-FRI. Particularly, for unmatched observations in a sparse rule base of a small size, the corresponding conclusions are produced by the interpolation with a small number of closest rules while the interpolated rules are directly integrated into the original sparse rule base. When the number of rules (be they original or interpolated) reaches a certain threshold, the rule base is regarded as a large-sized one. In this case, rules are clustered first and the corresponding conclusions are computed using the closest rules selected from a small number of closest rule clusters. The interpolated rules are then, also integrated into the rule base. Systematic experimental comparisons against the static approach demonstrate that the proposed dynamic TSK system can both improve the overall reasoning accuracy and reduce the interpolation overheads by extending the rule base coverage.

The rest of this chapter is structured as follows. Section 6.1 briefly outlines the static TSK systems with a sparse rule base. Section 6.2 details the framework of the dynamic TSK inference systems. Section 6.3 discusses the results of comparative experimental evaluations. Finally, Section 6.4 provides a summary of the work reported in this chapter.

6.1 Static TSK Inference Systems

The overall procedure of static TSK inference system is presented in Fig. 6.1. Due to the fact that the rule base is sparse, it is possible that a new observation does not match any rule. Thus, the first step is to determine whether an observation matches any rule in the sparse rule base. For the matched input observations, the conventional TSK inference mechanism is fired to obtain the conclusion; as for the unmatched ones, TSK interpolation approach will be applied, using either KCR or CRC (which is purely for efficiency while any alternative FRI mechanism, e.g., TSK+ (Li *et al.*, 2018b), AutoRS (Chen *et al.*, 2019), may be employed if preferred). The details of the above approaches have been presented in Alg. 1 (TSK models), Alg. 8 (KCR) and Alg. 9 (CRC), respectively.

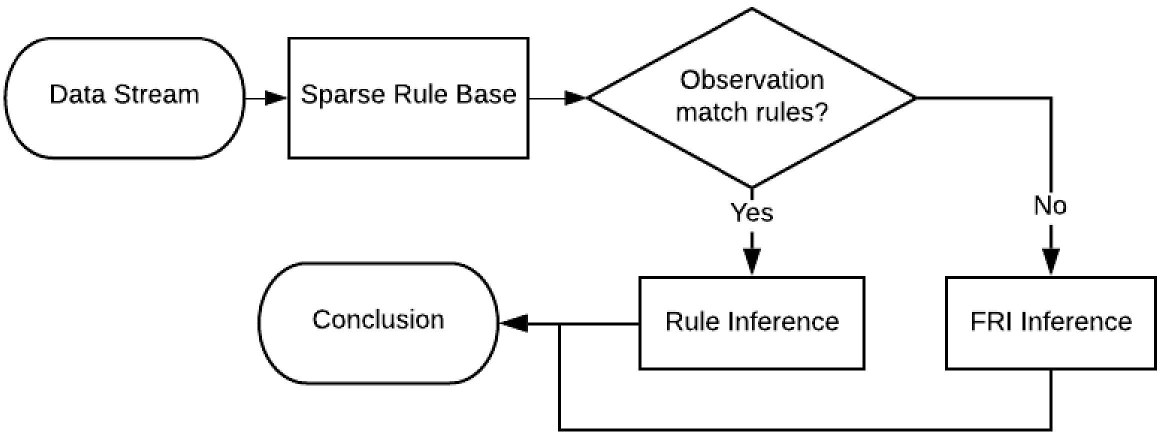

Figure 6.1: Inference process of static TSK fuzzy inference system

6.2 Dynamic TSK Inference Systems

The workflow of the proposed dynamic TSK fuzzy inference system is shown in Fig. 6.2. The system checks whether an input instance overlaps with any rules in the sparse rule base at the beginning. If so, it applies the conventional TSK inference mechanism to obtain the outcome. Otherwise, the FRI process becomes active, with the size of the rule base being checked first. For a sparse rule base of a small size, the KCR method is employed to generate the consequent of the interpolated rule and compute the result by the obtained polynomial. The antecedent of the interpolated rule is constructed by fuzzifying the observation with fuzzy sets represented by the triangular membership function. The complete interpolated rule is then directly inserted into the rule base. When the number of rules, be they original or interpolated, reaches a certain threshold, it becomes a large-sized rule base. For a rule base of a large size, the CRC approach is applied to obtain the consequents of the interpolated rule for conclusion calculation. The interpolated rule is also added into the rule base after integrating with antecedents. KCR and CRC methods will be described in detail in subsequent sections, respectively.

In doing so, with the increase of the number of rules in the sparse rule base, the overall system's coverage will rise and the interpolation overheads will reduce. The rule interpolation for dynamic systems and promotion processes are detailed in the following, in relation to whether KCR or CRC is to be used. Note that any of the five different clustering algorithms outlined in Section 2.3 (and indeed many other clustering methods if preferred) can be employed here to perform rule clustering.

In this algorithm, it is assumed, without losing generality, that a sparse rule base contains m rules (with each specified as per Eqn. 3.1) and an observation $O(B_1, ..., B_n)$ are given.

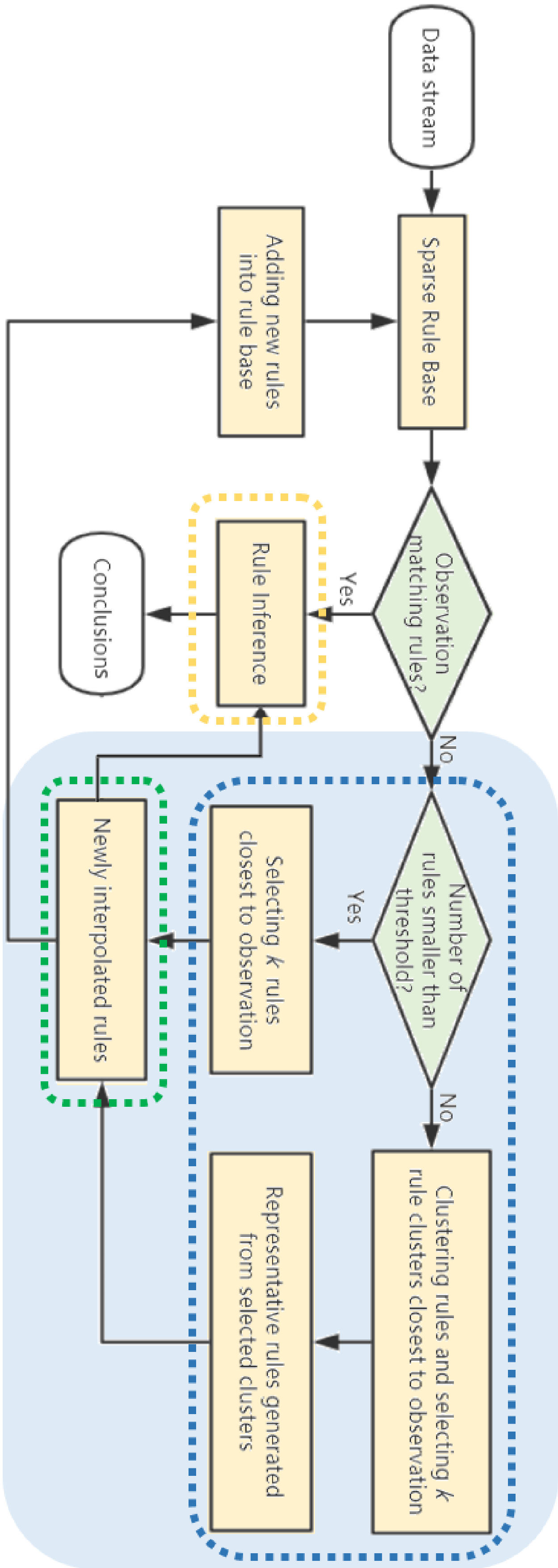

Figure 6.2: Inference process of dynamic TSK fuzzy inference system

The conclusion computation and interpolated rule construction in support of KCR procedure for sparse rule bases of a small size and CRC procedure for large-sized sparse rule bases is detailed in Alg. 11 and Alg. 12, respectively.

6.3 Experimental Evaluation

In this section, the performance of the proposed dynamic TSK inference system is experimentally compared against the static system over three benchmark datasets. The datasets run include a nonlinear mathematical model and two real-world datasets (Stock and Plastic (Guvenir *et al.*, 2000)).

6.3.1 Experimental Setup

The necessary setting for conduction experimental studies is outlined in this section, including the details of the three benchmark datasets, the criteria for evaluating performance, the generation of the sparse rule bases, and the parameters for implementing static and dynamic systems.

6.3.1.1 Three datasets used

(1) Polynomial Function: This dataset is produced by randomly sampling from the following 3-dimensional nonlinear function:

$$F(x, y) = sin(\frac{x}{\pi}) \cdot sin(\frac{y}{\pi}) \tag{6.7}$$

This nonlinear function has been used to produce a benchmark dataset in (Li *et al.*, 2018b) and (Rezaee & Zarandi, 2010), and the random sampling method has been frequently employed in the literature (e.g., (Li *et al.*, 2017); (Bellaaj *et al.*, 2013)). Two thousand points are randomly sampled as the original dataset, with the output domain being [-1, 1].

Algorithm 11: FRI with KCR on small-sized sparse rule bases for dynamic TSK systems.

Input: Rule base $\{R_i\}$; time-dependent data stream O

Output: Interpolated conclusion $f(O)$; interpolated rule $R*$

1. Calculate overall Euclidean distance between representative values of individual variables within observation and those of antecedent variables for each given rule.

2. Select K closest rules by Quickselect algorithm (Hoare, 1961).

3. Calculate similarity between observation O and each of R_i that belongs to set of selected K closest rules:

$$S(A_{i1}, B_1), ..., S(A_{in}, B_n)$$

.

4. Determine 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 following parameters for its consequent:

$$a_0 = \frac{\Sigma_{i=1}^{K} \alpha_i a_{i0}}{\Sigma_{i=1}^{K} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{K} \alpha_i a_{in}}{\Sigma_{i=1}^{K} \alpha_i} \tag{6.1}$$

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

7. Construct n triangular fuzzy sets as antecedents of interpolated rule:

$$(B_1 - \epsilon, B_1, B_1 + \epsilon), ..., (B_n - \epsilon, B_n, B_n + \epsilon) \tag{6.2}$$

with ϵ being a small number (set to 0.01 in this chapter).

8. Add the interpolated rule into rule base.

$$\begin{aligned} R* : \textbf{if } & x_1 \text{ is } (B_1 - \epsilon, B_1, B_1 + \epsilon), ..., x_n \text{ is } (B_n - \epsilon, B_n, B_n + \epsilon), \\ \textbf{then } & f_i(x_1, ..., x_n) = a_0 + a_1 x_1 + ... + a_n x_n \end{aligned} \tag{6.3}$$

9. Return: $f(B_1, ..., B_n)$; $R*$.

Algorithm 12: FRI with CRC on large-sized sparse rule bases for dynamic TSK systems.

Input: Rule base $\{R_i\}$; time-dependent data stream O

Output: Interpolated conclusion $f(O)$; interpolated rule $R*$

1. Cluster all rules into C different groups by their representative values.

2. Calculate Euclidean distance between observation and all cores of C clusters and select K ($K \leq C$) closest clusters.

3. Compute distance between observation O and each rule within each of chosen K clusters.

4. Find closest rule R_i in each selected cluster as representative of that cluster.

5. Determine weight of the rule R_i: $\alpha_i = S(A_{i1}, B_1) \wedge ... \wedge S(A_{in}, B_n)$.

6. Integrate all K similarities to obtain interpolated rule, with parameters of consequent being:

$$a_0 = \frac{\Sigma_{i=1}^{K} \alpha_i a_{i0}}{\Sigma_{i=1}^{K} \alpha_i}, ..., a_n = \frac{\Sigma_{i=1}^{K} \alpha_i a_{in}}{\Sigma_{i=1}^{K} \alpha_i} \tag{6.4}$$

7. Take observation O as input to consequent of interpolated rule such that outcome is computed by: $f(B_1, ..., B_n) = a_0 + a_1 B_1 + ... + a_n B_n$.

8. Construct n triangular fuzzy sets as antecedents of interpolated rule:

$$(B_1 - \epsilon, B_1, B_1 + \epsilon), ..., (B_n - \epsilon, B_n, B_n + \epsilon) \tag{6.5}$$

   with ϵ being a small number (set to 0.01 in this chapter).

9. Add interpolated rule into rule base.

$$\begin{aligned} R* : \textbf{if } & x_1 \text{ is } (B_1 - \epsilon, B_1, B_1 + \epsilon), ..., x_n \text{ is } (B_n - \epsilon, B_n, B_n + \epsilon), \\ \textbf{then } & f_i(x_1, ..., x_n) = a_0 + a_1 x_1 + ... + a_n x_n \end{aligned} \tag{6.6}$$

10. Return: $f(B_1, ..., B_n)$; $R*$.

(2) Stock Dataset: This dataset concerns stock prices for ten aerospace companies. The task is to predict the price for the 10th company given the prices for the rest (Guvenir *et al.*, 2000). The dataset consists of 950 instances and 9 features (i.e., antecedent variables), with the output in the range of [34, 62].

(3) Plastic Dataset: This dataset contains 1650 instances and 2 antecedent variables. The task is through regression to predict how much pressure the plastic materials can hold according to their strength in different temperature settings (Guvenir *et al.*, 2000). The output values are in the domain of [10.0, 20.0].

6.3.1.2 Performance Evaluation Criteria

To enable thorough evaluation and fair comparison, 10 $\times$ 5-fold cross-validation is employed. Note that different from the traditional cross-validation method, in the following experiments, to demonstrate the significance of the proposed dynamic TSK system, one fold is employed as a training dataset while the other four as testing data (or time-dependent data stream). The results are evaluated with respect to the following two criteria: (i) the accuracy of the computed consequents, in terms of RMSE (root-mean-square error, in relation to the ground truth); (ii) the coverage of the rule bases, in terms of the percentage of the number of instances matching the rules out of the number of instances in the whole testing set.

T-test is applied to detect whether there is any statistically significant difference between the results of the dynamic and static systems. The corresponding null hypothesis is that the results obtained by the two types of systems have no statistical difference. P-values represents the probability to accept the null hypothesis. Thus, if the p-value is smaller than a predefined significance level, the null hypothesis will be rejected, which indicates that there certainly is a significant difference between the results of the two systems. The significance level is herein set to 0.005.

6.3.1.3 Sparse Rule Base Generation

In the present experimental study, a simple data-driven fuzzy rule base generation method is employed to create the rules. The instances in a given dataset are firstly clustered into different categories through FCM. In general, FCM allows a data point to belong to more than one cluster with different membership values, in this work, for simplicity an instance is permitted to belong to two clusters that involve the two largest membership values. 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 (Freedman, 2009).

6.3.1.4 Algorithmic Parameters

Throughout this initial experimental investigation, observations are regarded as unmatched if the matching degree is less than 0.3 for all rules. Also, in the following experiments, 20 rules created from the training sets constitute the original fuzzy rule base. According to the experimental conclusion demonstrated in Chapter 4, The threshold of the large size sparse rule base is at least 95 rules included in the rule base. Regarding KCR, the number of closest rules K is empirically set to 3, and regarding CRC, the number of clusters C is set to 10 with the number of closest rule clusters K set to 3.

Particularly, according to the results demonstrated in Chapter 4, Similarity-DF leads to much improved outcomes than the other approach of rule weight determination, and CRC supported by FCM and K-FCM outperform other approaches, while FCM is less affected by data distribution. Thus, for generality and effectiveness, the similarity measure is utilised to calculate the weight of each selected rule and FCM is applied to perform rule clustering in CRC procedure for the sparse rule base of a large size in subsequent experiments.

6.3.2 Results and Discussion

This section illustrates and analyses the results of systematic comparative experimental investigations over three benchmark datasets. As indicated previously, the results are discussed from the following three aspects: the accuracy of the predicted conclusions (described by RMSE), the coverage of the rule bases, and the difference between the systems (in terms of t-test).

6.3.2.1 RMSE

The average values and variances are utilised to describe the distribution of all cross-validation results, as listed in Table 6.1. These results show that for all these three datasets, the conclusions obtained by the proposed dynamic TSK inference system has smaller errors and variances than those generated by the static one. It indicates that the proposed dynamic system can produce both higher accuracy and more stable results.

Examining more closely, together with the results of corresponding datasets shown in Table 4.4 and 4.5, although the same approaches are adopted on the same datasets, the accuracy of the static TSK system applied herein is much lower. This is probably due to the utilisation of the modified 5-fold cross-validation in the present research, which makes only one of five folds instances employed to generate the initial rule base, resulting in the decline of predicted accuracy.

To qualitatively visualise the strengths of the proposed approach, Fig.

Table 6.1: Performance in terms of RMSE

Datasets	Mean ± Variance	
	Dynamic	Static
Nonlinear model	0.201±0.0155	0.235±0.0175
Stock Dataset	1.116±0.161	1.818±0.283
Plastic Dataset	1.536±0.107	1.617±0.164

6.3 illustrates how the inference results (represented by RMSE) change while conducting the proposed dynamic TSK system over the stock dataset.

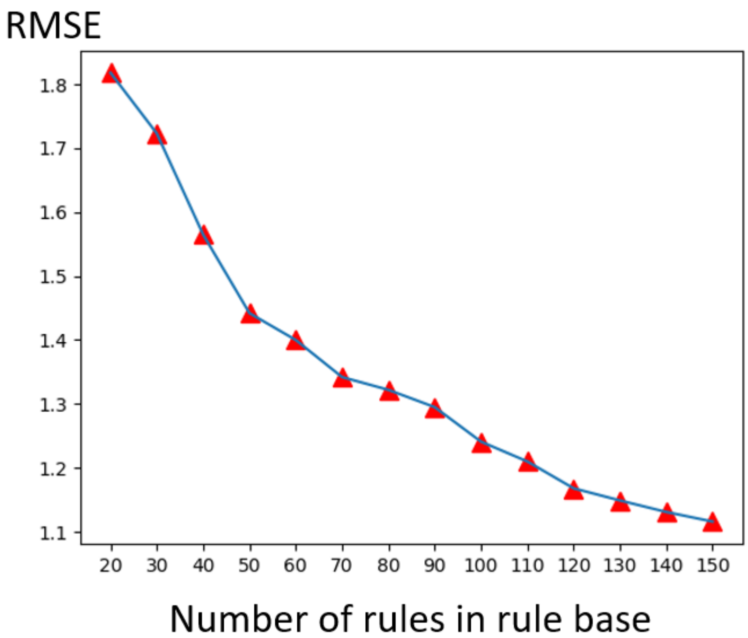

Figure 6.3: Model RMSE vs. number of rules K on proposed dynamic TSK system.

6.3.2.2 Coverage of rule bases

Table 6.2 lists the coverage of the rule bases by running the dynamic and static TSK inference systems. As reflected by these experimental outcomes, more instances overlap with the rules in the dynamic rule base than those in the static one. The dynamic system can gradually improve the coverage over time as new rules are promoted and added into the rule base. Benefiting from this, the dynamic TSK inference system avoids the need for interpolations when new observations which are similar to those previously experienced are presented, thereby reducing computational overheads.

Table 6.2: Performance in terms of coverage

Datasets	Coverage	
	Dynamic	Static
Nonlinear model	32.0%	26.0%
Stock Dataset	21.1%	15.7%
Plastic Dataset	36.4%	21.2%

Fig. 6.4 demonstrates how the coverage of the rule base promote while conducting the proposed dynamic TSK system over the stock dataset.

6.3.2.3 T-tests

By examining the experimental results of t-tests as given in Table 6.3, it can be seen that for all three datasets, the p-values are all smaller than the predefined significance level (0.005). Thus, the null hypothesis that there is no statistical difference between the results of the dynamic system and those of the static one is rejected. From this, together with previous conclusions on accuracy and coverage, it can be said that the proposed dynamic TSK inference system significantly improves the performance of the static one.

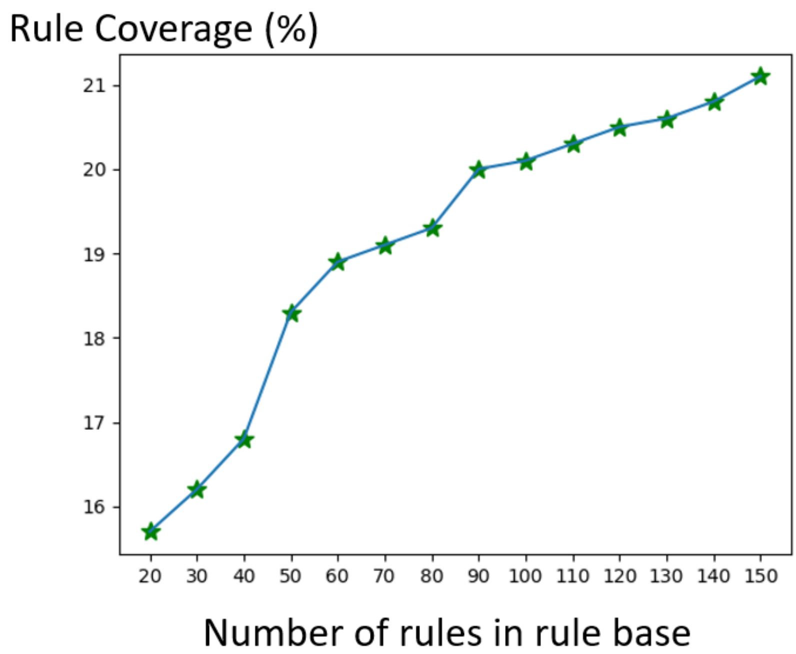

Figure 6.4: Model RMSE vs. number of rules K on proposed dynamic TSK system.

Table 6.3: Experimental results on t-test

Datasets	p-value	Hypothesis (0.005)
Nonlinear model	1.613e-11	Reject
Stock Dataset	9.825e-8	Reject
Plastic Dataset	4.27e-3	Reject

6.4 Summary

This chapter has presented an initial investigation into a novel dynamic TSK inference system suitable for working with sparse rule bases. Systematic comparative experimental results have demonstrated that compared to the existing static TSK inference approach, the proposed system increases the overall reasoning accuracy while being able to decrease the interpolation overheads by avoiding the need for interpolations for observations similar to those experienced.

The presented work iteratively add the newly interpolated rule into the existing rule base. However, it may make the rule base overly complicated, incurring significant computational overheads in any subsequent rule-firing process. Introducing an approach to refine the dynamically enriched rule base remains a challenge. The idea from D-FRI which is clustering and merging such interpolated rules on Mamdani models may help to resolve this problem.

Chapter 7

Conclusion

This chapter concludes the thesis. First, a summary of the research presented in this thesis is given, which also re-states the contributions made from the study. Secondly, possible future work is outlined, including several further developments for FRI techniques for static and dynamic systems.

7.1 Summary

The core works presented in this thesis are a novel approach with two implementations suitable for performing FRI with static and dynamic TSK fuzzy inference models. The proposed techniques not only avoid adverse impact caused by the rules of low similarities and undesirable bias but also help significantly reduce the computational overheads of the inference process. Particularly, the utilisation of clustering algorithms resolves the problem of lacking diversity of rules for interpolation caused by the situation where a large-sized sparse rule base is present. Meanwhile, the interpolated rules discarded by FRI approaches are further exploited to support promoting rule bases and constructing dynamic TSK systems. The following summarises the main works of the previous chapters from two viewpoints.

7.1.1 Static TSK Systems Supported by FRI

The proposed FRI approach for TSK models extends its capability to perform TSK models on sparse fuzzy rule bases while entailing the generation of more accurate interpolated results. It resolves limitations of existing FRI approaches in the literature. The proposed approach provides two implementations: (i) the innovative KCR algorithm for models involving small-sized sparse rule bases, which only involve a small number of closest rules to derive precise interpolation; (ii) the novel CRC algorithm for models with a large-sized sparse fuzzy rule base, which first categorises the rules into groups with a clustering procedure and then, utilises the closest rule from each of a small number of the resulting rule clusters to perform the interpolation. The proposed implementation for the second method is systematically evaluated using several different clustering techniques as alternatives for rule clustering tasks. Interestingly, it is observed that the distribution of rule bases has an significant influence on the final result calculation, which forms another piece of active research. Systematic experimental studies over a range of benchmark datasets have demonstrated the efficacy of both implementations.

Furthermore, to investigate the practicability of the proposed FRI approach in real-world applications, this thesis presents a status monitoring system for the water treatment plant. The missing values in the datasets are estimated using KCR-supported TSK models. The experimental results indicate that this novel approach have such capability to perform missing value imputation and derive a more accurate classification system with more instances.

7.1.2 Dynamic TSK Systems Supported by FRI

Given the proposed FRI approach and inspired by the aforementioned D-FRI (Naik *et al.*, 2017b), a novel dynamic TSK inference system suitable for working with sparse rule bases is established. With the support of the information contained in the interpolated rule discarded by FRI approaches, the proposed dynamic system conduct reasoning by constructing interpolated rules and iteratively integrating the rules into a rule base. Compared to the existing static TSK inference approach, the proposed system increases the overall reasoning accuracy and decreases the interpolation overheads by avoiding the need for interpolations for observations similar to those experienced.

7.2 Future Work

The works proposed in this thesis offer many opportunities for further development. This section points some of these opportunities for future research. Particularly, several attempted methods to implement FRI for TSK models in this project are briefly introduced. Although they may not draw satisfactory conclusions under the present circumstances, these methods have the potential to resolve practical problems in future development.

7.2.1 Adaptive Rule Selection for FRI

The adaptive FRI is one of the methods attempted in this thesis to support FRI for TSK models to deal with high-similarity issues where a large-sized sparse rule base is present. The underlying principle is to cluster rules according to their distance to the unmatched observation. It works by clustering rules into three categories based on their distance to the observation: strong, medium and weak. The strong rule set is the closest one. The nearest to the observation rule within each cluster is then selected for use as an element of the set of K closest rules. In so doing, the closest rules with high similarity are divided into the same categories and the diversity of rules used for interpolation is extended. In addition, distinct rule clusters are generated for individual observation, which provides the potential to conduct reasoning in real-time systems. An example of a relevant rule-clustering procedure is illustrated in Fig. 7.1.

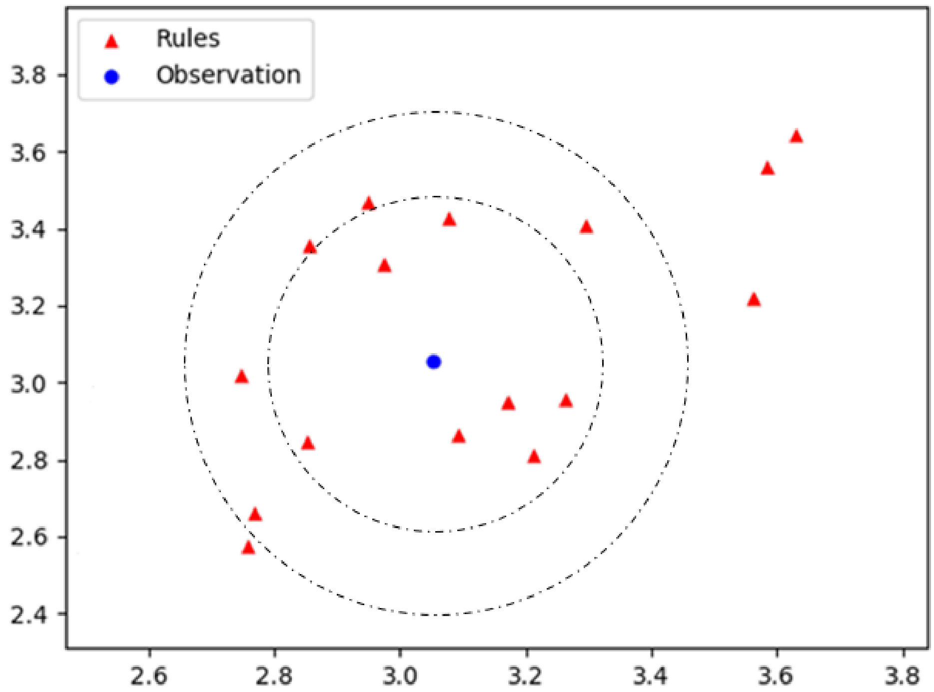

Figure 7.1: Example of rule clustering process in adaptive FRI approach.

Whilst being an applicable method, this attempt may wrongly cluster the closest rules which are not similar in the strong rule set. Therefore, these rules can not contribute to the final conclusion. The utilisation of the rules in medium and weak rule sets also violates the principle that the interpolated conclusion is mainly determined by its neighbour. Moreover, the distance based rule clustering method breaks the underlying similarity relation between the rules.

Although this attempted approach does not perform well in practice, it inspires the implementation of the rule clustering approach as indicated previously and provides the potential to conduct FRI adaptively to solve real-time problems. Furthermore, when big data is involved, this method may also be used for preliminary screening to select the rules relevant to the interpolation, rather than using the whole rule base directly.

7.2.2 Weighted FRI for TSK models

In the present implementations and most conventional FRI mechanisms, all antecedent variables are treated equally. This is a strong assumption in practical applications, often leading to less accurate interpolated results. How weighted representations as per the most recent work of (Li *et al.*, 2018a)(Li *et al.*, 2020) may be extended to accommodating interpolation with TSK models forms another interesting piece of active research. The work exploits attribute ranking approaches to help determine the relative importance of attributes (represented by rule antecedents) in a sparse rule base and then enable system integration of the individual attribute weights with the corresponding FRI methodologies. The attribute ranking approaches have been successfully applied in feature selection (FS) tasks, including the supervised attribute evaluation methods such as information gain (Mitchell, n.d.)(Shannon & C., 2001), Relief-F (Kononenko, 1994), and fuzzy-rough set based FS (Jensen & Shen, 2007)(Jensen & Shen, 2004), and unsupervised ones like local learning-based clustering for FS (Hong & Cheung, 2011) and Laplacian score (He *et al.*, 2005). The effectiveness and applicability of the weighted representations have been proven in practical applications (Li *et al.*, 2020) (Li *et al.*, 2020). Note that the above-mentioned supervised approaches are for dealing with classification problems using Mamdani models.

In this attempted approach, one of each supervised and unsupervised method is applied to implement attribute weight determination for TSK systems, which are information gain for regression (used in decision tree C4.5) and Laplacian score (He *et al.*, 2005). The corresponding weight integrated distance, which is also defined as the following:

$$d(R_i, O) = \sqrt{\Sigma_{j=1}^{N}(w_i(Rep(A_j) - Rep(B_j)))^2}, s.t.\Sigma_{j=1}^{N}w_j = 1 \tag{7.1}$$

Although systematic comparative experimental results indicate that neither implementations of the attempted approach can not obtain a better conclusion, discovering integration methodologies distinct from the one employed in Mamdani models to extend the attribute weight to support the FRI for TSK models forms an interesting piece of active research.

7.2.3 Representative of Rule Clusters

Moreover, in each selected rule cluster, the rule closest to the observation is regarded as the representative for subsequent interpolation because it contains most of the underlying information relevant to the desired conclusions. In general, the rule closest to the cluster centre or the rule intergraded by all rules in the same cluster is considered to include the main consensus information of the group. However, systematic experiments demonstrate that the proposed method outperforms these two implements. Thus, finding a representative of each cluster applicable to every observation forms another interesting task in future research that may help further reduce computation overheads.

7.2.4 Advanced Clustering Algorithms

In the present work, only five clustering methods are considered to support the CRC approach. They may not generate the most appropriate categories for all situations. It is worth to investigate whether performance may be reinforced if more extensions of fuzzy clustering mechanisms are used as an alternative (e.g., the global fuzzy c-means (Wang *et al.*, 2006); the possibilistic fuzzy c-means (Pal *et al.*, 2005)). More advanced clustering methods (e.g., (Boongoen *et al.*, 2011); (Pan *et al.*, 2017)) may help even more. For example, fuzzy clustering by local approximation of membership (FLAME) (Fu & Medico, 2007) utilises neighbourhood density for each object to support fuzzy clustering, which makes the final results more stable and accurate than the conventional FCM. Moreover,because the object with the highest density among its neighbors is identified as the cluster centre, FLAME also provides a novel mechanism to determine the representative of rule clusters.

7.2.5 Automated Parameter Selection

Currently, the specification of the required algorithmic parameters, such as the number of closest rules and that of the nearest clusters, is another interesting point. Currently, these parameters are empirically set manually. Introducing an automated way to decide on these parameters from the training data remains a challenge. Also, AutoRS offers a means for automated selection of the number of closest rules for interpolation. How it may be integrated with the proposed approach to minimise human intervention is worth investigating.

7.2.6 Candidate Rule Integration

Last but not least, the proposed FRI approach for dynamic TSK systems repeats itself whenever a new unmatched observation is interpolated and all newly interpolated rules are added to the existing rule base directly. However, it may not be completely necessary to add all such interpolated rules as many of them may be clustered and merged as done in the D-FRI work. One possible solution is that if the interpolation process repeats a certain number of times in a certain region, this region will be identified as a frequent region. All candidate interpolated rules in the frequent region will be merged into a new rule according to their distance to the region centre. The obtained rule is deemed to be the representative of the region and added into the rule base. This may help simplify any subsequent rule-firing process without making a rule base overly complicated.

Appendix A

Publications from Thesis

A number of publications have been generated from the research carried out within this PhD project. The resultant publications that are in close relevance to the thesis are listed following, including three published papers and three journal publications currently under review.

1. Pu Zhang, Changjing Shang and Qiang Shen. “Fuzzy Rule Interpolation with K-Neighbours for Takagi-Sugeno-Kang Models.” Accepted for publication by IEEE Transaction on Fuzzy Systems.

2. Pu Zhang and Qiang Shen.“Missing Value Imputation by TSK Models for Water Treatment.” Under review for journal publication.

3. Pu Zhang and Qiang Shen.“Dynamic Takagi-Sugeno-Kang Systems with Sparse Rule Bases” Under review for journal publication.

4. Pu Zhang and Qiang Shen. “Dynamic TSK Systems Supported by Fuzzy Rule Interpolation: An Initial Investigation.” in 2020 IEEE World Congress on Computational Intelligence (WCCI). Being selected as one of finalists in WCCI 2020s Best Student Paper Awards.

5. Pu Zhang and Qiang Shen. “A Novel Framework of Fuzzy Rule Interpolation for Takagi-Sugeno-Kang Inference Systems.” in 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

6. Mou Zhou, Changjing Shang, Pu Zhang, Guobin Li, Shangzhu Jin, Jun Peng and Qiang Shen. "Towards Rule-ranking Based Fuzzy Rule Interpolation." in 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

Appendix B

List of Acronyms

ANFIS Adaptive Neuro-Fuzzy Inference System

AutoRS Automated Rule Selection

COG Centre of Gravity

CRC Interpolation with K Closest Rule Clusters

CRI Compositional Rule of Inference

DF Distance Factor

D-FRI Dynamic Fuzzy Rule Interpolation

EM Expectation Maximization

FCM Fuzzy C-means

FLAME Fuzzy clustering by Local Approximation of Membership

FRBSs Fuzzy Rule-Based Systems

FRFS Fuzzy-Rough Feature Selection

FRI Fuzzy Rule Interpolation

GA Genetic Algorithm

GMM Gaussian Mixture Models

GRBF	Gaussian Radial Basis Function
LLE	Local Linear Embedding
KCF	Interpolation with K Closest Rules
K-FCM	Kernel Fuzzy C-means
KH	Kczy and Hirota (Linear Fuzzy Rule Interpolation)
RCFCM	Rival Checked Fuzzy C-means
RIA	Rule Induction Algorithm
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
S-FCM	Suppress Fuzzy C-means
SSE	Sum-of-squares Error
SVM	Support Vector Machine
T-FRI	Scale and Move Transformation-based Fuzzy Rule Interpolation
TSK	Takagi-Sugeno-Kang
TSK+	TSK inference extension

References

Alcala-Fdez, J., & Alonso, Jose M. 2016. A Survey of Fuzzy Systems Software: Taxonomy, Current Research Trends, and Prospects. *IEEE Transactions on Fuzzy Systems*, **24**(1), 40–56.

Alcalá-Fdez, Jesús, Fernández, Alberto, Luengo, Julián, Derrac, Joaquín, García, Salvador, Sánchez, Luciano, & Herrera, Francisco. 2011. Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic & Soft Computing*, **17**.

Angelov, Plamen P. 2003. An evolutionary approach to fuzzy rule-based model synthesis using indices for rules. *Fuzzy sets and systems*, **137**(3), 325–338.

Angelov, Plamen P, & Buswell, Richard A. 2003. Automatic generation of fuzzy rule-based models from data by genetic algorithms. *Information Sciences*, **150**(1-2), 17–31.

Asai, Kiyoji, Michio, Terano, & Toshiro. 1994. *Applied fuzzy systems*. Applied Fuzzy Systems.

Åström, Karl J, & Wittenmark, Björn. 2013. *Adaptive control*. Courier Corporation.

Astrom, KJ, & Witternmark, B. 1995. *Adaptive Control, Lund Institute of Technology, Sweeden*.

Au, W. H., & Chan, Kcc. 2002. An effective algorithm for discovering fuzzy rules in relational databases. *In: 1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36228).*

Baranyi, P., Tikk, D., Yam, Y., Koczy, L. T., & Nadai, L. 1999. A new method for avoiding abnormal conclusion for /spl alpha/-cut based rule interpolation. *In: IEEE International Fuzzy Systems Conference.*

Bellaaj, Hatem, Ketata, Rouf, & Chtourou, Mohamed. 2013. A new method for fuzzy rule base reduction. *Journal of Intelligent & Fuzzy Systems*, **25**(3), 605–613.

Bezdek, James C, Ehrlich, Robert, & Full, William. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, **10**(2-3), 191–203.

Bilmes, Jeff A, *et al.* 1998. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, **4**(510), 126.

Bo, K. W., & Lai, V. S. 2011. A survey of the application of fuzzy set theory in production and operations management: 19982009. *International Journal of Production Economics*, **129**(1), 157–168.

Boongoen, Tossapon, Shang, Changjing, Iam-On, Natthakan, Shen, & Qiang. 2011. Extending Data Reliability Measure to a Filter Approach for Soft Subspace Clustering. *IEEE Transactions on Systems, Man & Cybernetics: Part B.*

Bouveyron, Charles, & Brunet-Saumard, Camille. 2014. Model-based clustering of high-dimensional data: A review. *Computational Statistics and Data Analysis*, **71**, 52–78.

Cannon, Robert L, Dave, Jitendra V, & Bezdek, James C. 1986. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE transactions on pattern analysis and machine intelligence*, 248–255.

Castillo, O., & Melin, P. 2012. A review on the design and optimization of interval type-2 fuzzy controllers. *Applied Soft Computing Journal*, **12**(4), 1267–1278.

Chang, H., Yeung, D. Y., & Xiong, Y. 2004. Super-resolution through neighbor embedding. *In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*

Chang, Pei-Chann, & Fan, Chin-Yuan. 2008. 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)*, **38**(6), 802–815.

Chang, Y. C., Chen, S. M., & Liau, C. J. 2008. Fuzzy Interpolative Reasoning for Sparse Fuzzy-Rule-Based Systems Based on the Areas of Fuzzy Sets. *IEEE Transactions on Fuzzy Systems*, **16**(15), 1285–1301.

Chen, C., MacParthalain, N., Li, Y., Price, C., Quek, C., & Shen, Q. 2016. Rough-fuzzy rule interpolation. *Information Sciences*, **351**, 1–17.

Chen, S. M., & Adam, S. I. 2017. Adaptive fuzzy interpolation based on general representative values of polygonal fuzzy sets and the shift and modification techniques. *Information Sciences*, **414**.

Chen, S. M., & Chen, Z. J. 2016. Weighted fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on piecewise fuzzy entropies of fuzzy sets. *Information Sciences*, **329**, 503–523.

Chen, S. M., & Lee, L. W. 2011. Fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on interval type-2 fuzzy sets. *Expert Systems with Applications*, **38**(8), 9947–9957.

Chen, S. M., Chang, Y. C., Chen, Z. J., & Chen, C. L. 2013. MULTIPLE FUZZY RULES INTERPOLATION WITH WEIGHTED ANTECEDENT VARIABLES IN SPARSE FUZZY RULE-BASED SYSTEMS. *International Journal of Pattern Recognition and Artificial Intelligence*, **27**(5), 1359002.1–1359002.15.

Chen, Shi-Jay, & Chen, Shyi-Ming. 2003. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on fuzzy systems*, **11**(1), 45–56.

Chen, Shyi-Ming. 1996. New methods for subjective mental workload assessment and fuzzy risk analysis. *Cybernetics & Systems*, **27**(5), 449–472.

Chen, Tianhua, Shang, Changjing, Yang, Jing, Li, Fangyi, & Shen, Qiang. 2019. A new approach for transformation-based fuzzy rule interpolation. *IEEE Transactions on Fuzzy Systems*, **28**(12), 3330–3344.

Cheng, K., Law, N., & Siu, W. 2012. Iterative bicluster-based least square framework for estimation of missing values in microarray gene expression data. *Pattern Recognition*.

Cheng, S. H., Chen, S. M., & Chen, C. L. 2016. Adaptive fuzzy interpolation based on ranking values of polygonal fuzzy sets and similarity measures between polygonal fuzzy sets. *Information Sciences*, 176–190.

Dempster, A., Laird, N., & Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, **39**, 1–38.

Dua, Dheeru, & Graff, Casey. 2017. *UCI Machine Learning Repository*.

Dunn, Joseph C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

F, Cismondi, Fialho, Andre S., Vieira, S. M., Reti, S. R., Sousa, Joao M. C., & Finkelstein, S. N. 2013. Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine*, **58**(1), 63–72.

Fan, Jiu-Lun, Zhen, Wen-Zhi, & Xie, Wei-Xin. 2003. Suppressed fuzzy c-means clustering algorithm. *Pattern Recognition Letters*, **24**(9-10), 1607–1612.

Farhangfar, A., Kurgan, L., & Dy, J. 2008. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, **41**(12), 3692–3705.

Freedman, David A. 2009. *Statistical models: theory and practice*. cambridge university press.

Fu, L., & Medico, E. 2007. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *Bmc Bioinformatics*, **8**.

Graves, D., & Pedrycz, W. 2010. Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. *Fuzzy Sets & Systems*, **161**(4), 522–543.

Grosan, Crina, & Abraham, Ajith. 2011. Rule-based expert systems. *Pages 149–185 of: Intelligent Systems*. Springer.

Guvenir, H Altay, Uysal, Ilhan, & Repositor, Function Approximation. 2000. Function approximation repository. *Bilkent University. URL http://funapp. cs. bilkent. edu. tr/DataSets*.

He, X., Cai, D., & Niyogi, P. 2005. Laplacian Score for Feature Selection. *In: Advances in Neural Information Processing Systems 18 [Neural Infor-*

mation Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada].

Hoare, Charles AR. 1961. Algorithm 65: find. *Communications of the ACM*, **4**(7), 321–322.

Hong, Z., & Cheung, Y. M. 2011. Feature Selection and Kernel Learning for Local Learning-Based Clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **33**(8), 1532–47.

Huang, Zhiheng, & Shen, Qiang. 2008. Fuzzy interpolation and extrapolation: A practical approach. *IEEE Transactions on Fuzzy Systems*, **16**(1), 13–28.

Hyunsoo, K., Golub, G. H., & Haesun, P. 2005. Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics*, 187–98.

Jang, J.-S.R. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans on Smc*, **23**(3), 665–685.

Jensen, R., & Shen, Q. 2004. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches. *IEEE Transactions on Knowledge & Data Engineering*, **16**(12), 1457–1471.

Jensen, R., & Shen, Q. 2009. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, **17**(4), 824–838.

Jensen, Richard, & Shen, Qiang. 2007. Fuzzy-Rough Sets Assisted Attribute Selection. *IEEE Transactions on Fuzzy Systems*, **15**(1), 73–89.

Jerez, J M, Molina, I., Garca-Laencina, P J, Alba, E., & Franco, L. 2010. Missing data imputation using statistical and machine learning approaches in a real breast cancer problem. *Artificial Intelligence in Medicine.*

Jiang, Y., Chung, F. L., Wang, S., Deng, Z., Wang, J., & Qian, P. 2015. Collaborative Fuzzy Clustering From Multiple Weighted Views. *IEEE Trans Cybern*, **45**(4), 688–701.

Joo, I., Koczy, L. T., Tikk, D., & Varlaki, P. 2002. Stability of interpolative fuzzy KH-controllers. *In: Proceedings of 6th International Fuzzy Systems Conference.*

Kasabov, N. K., & Song, Q. 2002. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Transactions on Fuzzy Systems*, **10**(2), 144–154.

Kczy, L, & Hirota, K. 1993. Approximate reasoning by linear rule interpolation and general approximation. *International Journal of Approximate Reasoning*, **9**(3), 197–225.

Kczy, L, & Hirota, K. 2015. Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases. *Information Sciences*, **71**(1-2), 169–201.

Koczy, Laszlo, T., Hirota, & Kaoru. 1997. Size reduction by interpolation in fuzzy rule bases. *IEEE Transactions on Systems, Man and Cybernetics: Part B.*

Koczy, L. T., Hirota, K., & Muresan, L. 2000. Interpolation in hierarchical fuzzy rule bases. *In: IEEE IEEE International Conference on Fuzzy Systems.*

Kóczy, LászlóT, & Hirota, Kaoru. 1993. Approximate reasoning by linear rule interpolation and general approximation. *International Journal of Approximate Reasoning*, **9**(3), 197–225.

Kodinariya, Trupti M, & Makwana, Prashant R. 2013. Review on determining number of Cluster in K-Means Clustering. *International Journal*, **1**(6), 90–95.

Kononenko, Igor. 1994. Estimating attributes: analysis and extension of relief.

Kosko, B. 1986. Fuzzy entropy and conditioning. Inf Sci. *Information Sciences*, **40**(2), 165–174.

Kovarik, Vincent J. 2009. Chapter 12 - Cognitive Research: Knowledge Representation and Learning. *Pages 367–399 of:* Fette, Bruce A. (ed), *Cognitive Radio Technology (Second Edition)*, second edition edn. Oxford: Academic Press.

Kovcs, S. 2006. *Extending the Fuzzy Rule Interpolation "FIVE" by Fuzzy Observation*. Springer Berlin Heidelberg.

Kovcs, S. 2014. *Practical Aspects of Fuzzy Rule Interpolation*. Advances in Soft Computing, Intelligent Robotics and Control.

Kovcs, S, & Kczy, LT. 2004. Application of interpolation-based fuzzy logic reasoning in behaviour-based control structures. *In: IEEE International Conference on Fuzzy Systems*.

Kovcs-Lszl, S, & Kczy, T. 1999. Application of an approximate fuzzy logic controller in an AGV steering system, path tracking and collision avoidance strategy. *tatra mt.math.publ*, **16**(Part II), 325–338.

Kurfess, Franz J. 2003. Artificial Intelligence. *Pages 609–629 of:* Meyers, Robert A. (ed), *Encyclopedia of Physical Science and Technology (Third Edition)*, third edition edn. New York: Academic Press.

Lai, X., Liu, X., Zhang, L., Lin, C., & Hsiao, K. F. 2019. Missing Value Imputations by Rule-Based Incomplete Data Fuzzy Modeling. *In: ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*.

Lai, X., Zhang, L., & Liu, X. 2020. Takagi-Sugeno Modeling of Incomplete

Data for Missing Value Imputation With the Use of Alternate Learning. *IEEE Access*.

Li, F., Li, Y., Shang, C., & Shen, Q. 2019. Fuzzy Knowledge-Based Prediction Through Weighted Rule Interpolation. *IEEE Transactions on Cybernetics*, 1–10.

Li, Fangyi, Shang, Changjing, Li, Ying, Yang, Jing, & Shen, Qiang. 2018a. Fuzzy Rule-Based Interpolative Reasoning Supported by Attribute Ranking. *IEEE Transactions on Fuzzy Systems*, **26**(5), 2758–2773.

Li, Fangyi, Shang, Changjing, Li, Ying, & Shen, Qiang. 2020. Interpretable mammographic mass classification with fuzzy interpolative reasoning. *Knowledge-Based Systems*, **191**, 105279.

Li, Fangyi, Shang, Changjing, Li, Ying, Yang, Jing, & Shen, Qiang. 2021. Approximate reasoning with fuzzy rule interpolation: background and recent advances. *Artificial Intelligence Review*, **54**, 4543–4590.

Li, Jie, Qu, Yanpeng, Shum, Hubert PH, & Yang, Longzhi. 2017. TSK inference with sparse rule bases. *Pages 107–123 of: Advances in Computational Intelligence Systems*. Springer.

Li, Jie, Yang, Longzhi, Qu, Yanpeng, & Sexton, Graham. 2018b. An extended Takagi–Sugeno–Kang inference system (TSK+) with fuzzy interpolation and its rule base generation. *Soft Computing*, **22**(10), 3155–3170.

Liao, Shu-Hsien. 2005. Expert system methodologies and applicationsa decade review from 1995 to 2004. *Expert systems with applications*, **28**(1), 93–103.

Little, Rja, & Rubin, D. B. 1986. Statistical analysis with missing data. *Wiley*.

Liu, Bin-Da, Chen, Chuen-Yau, & Tsao, Ju-Ying. 2001. Design of adaptive fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **31**(1), 32–53.

MacQueen, James, *et al.* 1967. Some methods for classification and analysis of multivariate observations. *Pages 281–297 of: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. Oakland, CA, USA.

Mamdani, Ebrahim H, & Assilian, Sedrak. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*, **7**(1), 1–13.

Mitchell, T. M. Machine learning (mcgraw-hill. *machine learning*.

Mizumoto, Masaharu, & Zimmermann, Hans-Jürgen. 1982. Comparison of fuzzy reasoning methods. *Fuzzy sets and systems*, **8**(3), 253–283.

Mller, KR, Mika, S., Rtsch, G, Tsuda, K., & Schlkopf, B. 2001. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, **12**(2), 181.

Mohan, Sudeept, & Bhanot, Surekha. 2006. Comparative study of some adaptive fuzzy algorithms for manipulator control. *International Journal of Computational Intelligence*, **3**(4), 303–311.

Müller, Klaus-Robert, Mika, Sebastian, Tsuda, Koji, & Schölkopf, Koji. 2018. An introduction to kernel-based learning algorithms. *Handbook of Neural Network Signal Processing*, 4–1.

Nagy, S., Sziov, B, & Kczy, LT. 2018. The effect of image feature qualifiers on fuzzy colorectal polyp detection schemes using KH interpolation - towards hierarchical fuzzy classification of coloscopic still images. *In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.

Naik, N., Ren, D., Shang, C., Shen, Q., & Jenkins, P. 2017a. D-FRI-WinFirewall: Dynamic fuzzy rule interpolation for Windows Firewall. *In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).*

Naik, Nitin, Diao, Ren, & Shen, Qiang. 2014. Genetic algorithm-aided dynamic fuzzy rule interpolation. *Pages 2198–2205 of: Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on.* IEEE.

Naik, Nitin, Diao, Ren, & Shen, Qiang. 2017b. Dynamic fuzzy rule interpolation and its application to intrusion detection. *IEEE Transactions on Fuzzy Systems*, **26**(4), 1878–1892.

Pal, Nikhil R, Pal, Kuhu, Keller, James M, & Bezdek, James C. 2005. A possibilistic fuzzy c-means clustering algorithm. *IEEE transactions on fuzzy systems*, **13**(4), 517–530.

Pan, S., Shang, C., Chen, T., & Qiang, S. 2017. Exploiting data reliability and fuzzy clustering for journal ranking. *IEEE Transactions on Fuzzy Systems*, **25**(5), 1306–1319.

Pigott, & Therese, D. 2001. A Review of Methods for Missing Data. *Educational Research & Evaluation*, **7**(4), 353–383.

Rana, S., John, A. H., & Midi, H. 2012. Robust regression imputation for analyzing missing data. *In: International Conference on Statistics in Science.*

Rezaee, Babak, & Zarandi, MH Fazel. 2010. Data-driven fuzzy modeling for Takagi–Sugeno–Kang fuzzy system. *Information Sciences*, **180**(2), 241–255.

Rongrong, Ji, Hong, Liu, Liujuan, Cao, Di, Liu, Yongjian, & Wu. 2017. Toward Optimal Manifold Hashing via Discrete Locally Linear Embedding. *IEEE Transactions on Image Processing*, **26**(11), 5411–5420.

Ross, T. J. 2004. *Fuzzy Logic with Engineering Applications*. McGraw-Hill,Inc.

Roweis, St, & Saul, Lk. 2000. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, **290**(5500), 2323–2326.

Saunders, C., Stitson, M. O., Weston, J., Holloway, Royal, Bottou, L., Scholkopf, B., & Smola, A. 2002. Support Vector Machine. *Computer Science*, **1**(4), 1–28.

Scherer, Rafa. 2012. *Introduction to Fuzzy Systems*. Multiple Fuzzy Classification Systems.

Shannon, & C., E. 2001. A mathematical theory of communication. AT and T Tech J. *Acm Sigmobile Mobile Computing and Communications Review*, **5**(1), 3–55.

Shen, Q., & Chouchoulas, A. 2002. A rough-fuzzy approach for generating classification rules. *Pattern Recognition*, **35**(11), 2425–2438.

Shen, Q., & Jensen, R. 2004. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition*, **37**(7), 1351–1363.

Shi-Jay, Chen, Shyi-Ming, & Chen. 2003. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Fuzzy Systems, IEEE Transactions on*.

Sridevi, B., & Nadarajan, R. 2009. Fuzzy Similarity Measure for Generalized Fuzzy Numbers. *International Journal of Open Problems in Computer Science and Mathematics*, **2**(2).

Takagi, Tomohiro, & Sugeno, Michio. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, 116–132.

Thorndike, R. 1953. Who belongs in the family? *Psychometrika*, **18**(4), 267–276.

Tikk, D, & Baranyi, P. 2000a. Comprehensive analysis of a new fuzzy rule interpolation method. *IEEE Transactions on Fuzzy Systems.*

Tikk, D., & Baranyi, P. 2000b. Comprehensive analysis of a new fuzzy rule interpolation method. *IEEE Transactions on Fuzzy Systems*, **8**(3), 281–296.

Vincze, D., & Kovacs, S. 2008. Using fuzzy rule interpolation based automata for controlling navigation and collision avoidance behaviour of a robot. *In: IEEE International Conference on Computational Cybernetics.*

Wang, Weina, Zhang, Yunjie, Li, Yi, & Zhang, Xiaona. 2006. The global fuzzy c-means clustering algorithm. *Pages 3604–3607 of: 2006 6th World Congress on Intelligent Control and Automation*, vol. 1. IEEE.

Wei, L-m, & Xie, W-x. 2000. Rival checked fuzzy c-means algorithm. *Acta Electronica Sinica*, **28**(7), 63–66.

Wu, Shiqian, Er, Meng Joo, & Gao, Yang. 2001. A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks. *IEEE Transactions on Fuzzy Systems*, **9**(4), 578–594.

Yam, Y., Baranyi, P., Tikk, D., & Kczy, LT. 1999. Eliminating the Abnormality Problem of -cut Based Fuzzy Interpolation. *In: Eighth International Fuzzy Systems Association World Congress Vol.*

Yang, Jing, Shang, Changjing, Li, Ying, Li, Fangyi, & Shen, Qiang. 2021. Anfis construction with sparse data via group rule interpolation. *IEEE Transactions on Cybernetics*, **51**(5), 2773–2786.

Yang, L., & Sheng, Q. 2013. Closed form fuzzy interpolation. *Fuzzy Sets and Systems*, **225**(aug.16), 1–22.

Yang, Longzhi, & Shen, Qiang. 2011. Adaptive fuzzy interpolation. *IEEE Transactions on Fuzzy Systems*, **19**(6), 1107–1126.

Yuan, J., Chen, M., Jiang, T., & Li, T. 2017. Complete Tolerance Relation based Parallel Filling for Incomplete Energy Big Data. *Knowledge-Based Systems*, **132**(sep.15), 215–225.

Yuan, Y., & Shaw, Mjp. 1995. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*.

Zadeh, L. A. 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man, and Cybernetics, SMC (vol)3:1:28-45. *IEEE Transactions on Systems Man and Cybernetics*, **SMC-3**(1), 28–44.

Zadeh, Lotfi A. 1965. Fuzzy sets. *Information and control*, **8**(3), 338–353.

Zhang, Dao-Qiang, & Chen, Song-Can. 2004. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial intelligence in medicine*, **32**(1), 37–50.

Zhang, Huaguang, & Bien, Zeungnam. 2000. Adaptive fuzzy control of MIMO nonlinear systems. *Fuzzy sets and systems*, **115**(2), 191–204.

Zhang, P., & Shen, Q. 2020. Dynamic TSK Systems Supported by Fuzzy Rule Interpolation: An Initial Investigation. *In: 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).*

Zhang, Pu, & Shen, Qiang. 2018. Fuzzy c-means based coincidental link filtering in support of inferring social networks from spatiotemporal data streams. *Soft Computing*, **22**(21), 7015–7025.

Zhang, Pu, Shang, Changjing, & Shen, Qiang. 2021. Fuzzy Rule Interpolation with K-Neighbours for TSK Models. *IEEE Transactions on Fuzzy Systems*. DOI:10.1109/TFUZZ.2021.3136359.