Ordered weighted aggregation of fuzzy similarity relations and its application to detecting water treatment plant malfunction

Su, Pan; Shen, Qiang; Chen, Tianhua; Shang, Changjing

Published in:
Engineering Applications of Artificial Intelligence

DOI:
10.1016/j.engappai.2017.08.009

Publication date:
2017

Citation for published version (APA):
Su, P., Shen, Q., Chen, T., & Shang, C. (2017). Ordered weighted aggregation of fuzzy similarity relations and its application to detecting water treatment plant malfunction. *Engineering Applications of Artificial Intelligence*, *66*, 17-29. https://doi.org/10.1016/j.engappai.2017.08.009

Ordered Weighted Aggregation of Fuzzy Similarity Relations and its Application to Detecting Water Treatment Plant Malfunction

Pan Su

School of Control and Computer Engineering, North China Electric Power University, Baoding, China

Qiang Shen, Tianhua Chen and Changjing Shang*

Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, Aberystwyth, UK

Abstract

Ordered weighted aggregation procedures have been introduced in many applications with promising results. In this paper, an innovative approach for ordered weighted aggregation of fuzzy relations is proposed. It allows the integration of component relations generated from different perspectives of a certain observation to form an overall fuzzy relation, deriving a useful similarity measure for observed data points. Two types of aggregation are investigated: a) min/max operators are employed for the aggregation of component relations defined by the minimum T-norm; and b) sum/product operators are employed for the aggregation of component relations defined by the Łukasiewicz T-norm. The resultant ordered weighted aggregations prove to preserve the desirable reflexivity and symmetry properties, with T-transitivity also conditionally preserved if appropriate weighting vectors are adopted. The conditions upon which the proposed aggregated relations preserve T-transitivity are studied. The characteristics of applying an aggregated relation in combination with clustering procedures is also experimentally examined, where fuzzy similarity relations regarding individual features are aggregated to support hierarchical clustering. An application to the detection of water treatment plant malfunction demonstrates that better results can be obtained with the transitive fuzzy relations acting as the required similarity measures, as compared to the use of non-transitive ones. By introducing transitivity to the aggregation the interpretability of the detection system is also enriched.

Keywords: Fuzzy relations, similarity measures, water treatment, OWA, aggregator transitivity, hierarchical clustering.

1. Introduction

Methods for aggregation of different pieces of information into an integrated form are an indispensable tool, not only for theoretical development in e.g., mathematics and physics, but for many real-world applications in engineering, economical, social, and other fields. Having recognised this, a significant number of aggregation operators have been developed, ranging from simple arithmetic mean to more complicated fuzzy methods, including minimum/maximum, uninorm, and other alternative T-norms/T-conorms [1, 2, 3]. In particular, a class of parameterised mean-like aggregation operators, commonly named as ordered weighted averaging (OWA), have been introduced in the literature [4] and successfully applied in different areas [5, 6, 7, 8]. Intuitively, with an appropriate specification of a weighting vector, an OWA operator helps

*Corresponding author

Email addresses: supan1986@yahoo.com (Pan Su), {qqs, tic4, cns}@aber.ac.uk (Qiang Shen, Tianhua Chen and Changjing Shang)

to capture and reflect the uncertain nature of human judgments in problem-solving, generating an aggregated result that lies between the (conventional) two extremes of minimum or maximum combination of multi-featured data objects [9].

In general, relations holding amongst data points form the basis for many developments and applications of fuzzy systems. However, in their applications to supporting multicriteria decision making [10, 11], which forms a major challenge for practical fuzzy systems, a key question is what underlying properties of the data can be preserved in the process of constructing or aggregating similarity relations. For certain applications like prototype-based reasoning where clusters of objects that are similar to certain prototypical samples are sought [12], properties such as reflexivity and transitivity [13] may not be necessary. Yet, there are many other situations in which it is desirable to maintain the symmetry and a degree of transitivity over the homogeneous similarity classes or granules whose members possess these properties as symmetric and transitive classes or granules support intuitive interpretation of the reasoning process involved [14, 15, 16].

To enhance the mechanism for aggregation of fuzzy relations with such desired properties entailed, this paper presents two novel types of OWA-based aggregation methods, where the component relations are sorted first and subsequently aggregated with assigned weights. These techniques allow the aggregated results to retain the $T_{\min}$-transitive and T_{L}-transitive similarities, respectively. It is theoretically proven that the aggregated relations can hold the respective T-transitivity if the weights are arranged in ascending order. To illustrate the effectiveness of such ordered weighted aggregation of fuzzy relations, it is systematically evaluated over the task of clustering both synthetical and UCI datasets, by following the strategy of hierarchical clustering. In this experimental evaluation, similarities between data patterns are measured through ordered weighted aggregation of component fuzzy relations which hold amongst individual features. The work is applied to the detection of water treatment plant malfunction, demonstrating that the aggregated T_{L}-transitive similarities lead to better hierarchical clusters than those of non-transitive similarities.

The paper is organised as follows. Section 2 introduces the basic concepts of the aggregation of fuzzy relations. Section 3 presents two types of ordered weighted aggregation of fuzzy relations, with a detailed discussion of their properties, including the use of stress functions to decide on the weighting vectors for them. Section 4 describes the experimental investigation into the proposed aggregation of fuzzy relations in performing clustering tasks, evaluated over a number of classic datasets. Section 5 presents an application of the proposed aggregator to detecting malfunctions of a water treatment plant. The paper is concluded in Section 6, with a discussion of further research.

2. Preliminaries

2.1. Fuzzy relations

The concept of similarity is a preliminary notion in human cognition, playing an essential role in many tasks such as taxonomy, recognition, and inference (e.g., case-based reasoning). Particularly, fuzzy sets and relations [17] are of great significance in both theoretical development and industrial applications of constructing similarity metrics when dealing with imprecise situations [18, 19, 20].

Definition 1. *Let X be a nonempty universe. A fuzzy relation $R = [r(a,b)] : X \times X \to [0,1]$ is*
- *reflexive iff $\forall a \in X, r(a,a) = 1$;*
- *symmetric iff $\forall a,b \in X, r(a,b) = r(b,a)$;*
- *T-transitive iff $\forall a,b,c \in X, r(a,b) \geq T(r(a,c), r(c,b))$,*

where T is a T-norm [21], e.g., a mapping $T(x,y) : [0,1] \times [0,1] \to [0,1]$ which satisfies
1) commutativity: $T(x,y) = T(y,x)$;
2) monotonicity: $T(x,y) \leq T(x',y')$, if $x \leq x'$ and $y \leq y'$;
3) associativity: $T(x, T(y,z)) = T(T(x,y), z)$; and
4) boundary condition: $T(x,1) = x$.

A number of T-norms have been proposed in the literature, including (but not limited to):
- the minimum T-norm: $T_{\min}(x,y) = \min(x,y)$,
- the product T-norm: $T_{\text{p}}(x,y) = x \cdot y$, and

• the Łukasiewicz's T-norm: $T_{\text{L}}(x, y) = \max(x + y - 1, 0)$.

There exist many different definitions of similarity metrics which have been employed with success for different purpose such as clustering, classification, recognition and diagnostics. However, it is very challenging to validate the effectiveness of a similarity metric in real application scenarios. In this paper, the proposed aggregation methods focus on the use of transitive similarity metrics in support of water treatment plant monitoring.

2.2. Aggregation of fuzzy relations

In describing many engineering problems, an entity is commonly represented by a set of features or evaluated by a set of characteristic indicators [22]. As such, the evaluation of similarity between two entities is usually based on their feature/indicator-values. When multiple indicators are considered, an aggregator is typically employed to combine multiple similarity values into a single one. For example, the similarity degrees derived from individual water quality indicators can be aggregated using a weighted sum in an effort to construct an overall water quality index for rivers [23]. In the following, relevant concepts and properties regarding aggregation of fuzzy relations are introduced.

Formally, let X denote a finite set, $R_j = [r_j(a,b)] : X \times X \to [0,1], a, b \in X, j = 1, \cdots, m$ denote m fuzzy relations (named as component relations) on X, and $w_1, \cdots, w_m \in [0,1]$ denote weights, respectively associated with these relations. The aggregation process aims at providing a relation $R = [r(a,b)], a, b \in X$, summarising the component relations $R_1, \cdots, R_m$ in conjunction with the information implied by the weights $w_1, \cdots, w_m$. Here, the aggregated degree $r(a,b) \in [0,1]$ at position $(a,b), a, b \in X$ depends on the local compositions $r_1(a,b), \cdots, r_m(a,b)$. The component relations usually represent the similarities of patterns from different perspectives such as opinions from different experts, multiple criteria of evaluation and different features of describing data.

Definition 2. *[24] The aggregation of component relations $R_1 = [r_1(a,b)], \cdots, R_m = [r_m(a,b)], a, b \in X$, with weights $w_1, \cdots, w_m$, is a relation R over X such that*

$$r(a,b) = Agg(r_1(a,b), \cdots, r_m(a,b), w_1, \cdots, w_m) \tag{1}$$

where $a, b \in X$ and Agg is a mapping $[0,1]^{2m} \to [0,1]$, non-decreasing in the first m places and satisfying:

$$Agg(0, \cdots, 0, w_1, \cdots, w_m) = 0, \textit{ and } Agg(1, \cdots, 1, w_1, \cdots, w_m) = 1$$

Both the weighted and non-weighted aggregation procedures have been studied in the literature. For the purpose of aggregating fuzzy relations, typical methods investigated include the norm-conorm and sum-product operators. Usually, the T-norm/conorm operators are employed to aggregate a more general type of fuzzy relations while the sum-product operators are employed to aggregate fuzzy relations which preserve T_{L} transitivity [24, 25]. An aggregator may be described as optimistic or pessimistic: An optimistic aggregator produces outputs that are closer to the maximum of its inputs, and the outputs of a pessimistic one are closer to the minimum of its inputs.

Definition 3. *[24] Given component fuzzy relations $R_j = [r_j(a,b)], j = 1, \ldots, m$, the optimistic aggregated fuzzy relation over these relations is*

$$R_{opt} = [r_{opt}(a,b)] : \; r_{opt}(a,b) = S_{j=1,\cdots,m} T(w_j, r_j(a,b)); \tag{2}$$

and the pessimistic aggregated fuzzy relation over these relations is

$$R_{pess} = [r_{pess}(a,b)] : \; r_{pess}(a,b) = T_{j=1,\cdots,m} S(N(w_j), r_j(a,b)); \tag{3}$$

where, T is a T-norm, S is a T-conorm and N is a strong negation.

Intuitively, the weight w_j here reflects the relative importance of R_j. These two aggregators may be explained with the specific case where all the m weights are assumed to be either 0 or 1 (representing negligible or significant, respectively). In this case, Eqn. (2) and Eqn. (3) can be rewritten as $R_{\text{opt}} = S_{\{j|w_j=1\}} R_j$ and $R_{\text{pess}} = T_{\{j|w_j=1\}} R_j$, respectively. Thus, $r_{\text{opt}}(a, b)$ can be viewed as the degree of truth of the statement that "there exists at least one significant criterion for which a hold the relation with b", and $r_{\text{pess}}(a, b)$ as the degree of truth of the statement that "a holds the relation with b for all significant criteria" [24]. It has proved that $R_{\text{opt}} \subseteq R_{\text{pess}}$, that is, $r_{\text{opt}(a,b)} \geq r_{\text{pess}(a,b)}$. If the minimum, maximum and the standard negation $N(x) = 1 - x$ are selected as the T-norm, T-conorm and negation in definition 3 respectively, then

$$r_{\text{opt}}(a, b) = \max_{j=1,\cdots,m} \min(w_j, r_j(a, b)) \tag{4}$$

and

$$r_{\text{pess}}(a, b) = \min_{j=1,\cdots,m} \max(1 - w_j, r_j(a, b)). \tag{5}$$

Transitivity is considered to be an important property when similarity relations are adopted to handle many real-world problems. Take biological data partitioning as an example, transitivity clustering can outperform typical clustering approaches such as connected component analysis, Markov clustering, and spectral clustering [16]. Therefore, maintaining transitivity in the aggregation of fuzzy relations can be important and sometimes, critical to seeking desired solutions of a given problem. Three theorems about transitivity of fuzzy relation aggregations are introduced below.

Theorem 1. *If $R_1, \cdots, R_m$ are $T_{\min}$-transitive fuzzy relations, and $f_1, \cdots, f_m$ are non-decreasing mappings from $[0, 1]$ into $[0, 1]$, then $R = \min_{j=1,\cdots,m} f_j(R_j)$ is $T_{\min}$-transitive.*

It is easy to conclude from theorem 1 (whose proof can be found in [24]) that if $R_1, \cdots, R_m$ are $T_{\min}$-transitivity, then Eqn. (5) preserves $T_{\min}$-transitivity.

Theorem 2. *[25] Let $R_j = [r_j(a, b)], j = 1, \cdots, m$ be m T_L-transitive fuzzy relations. Then, the weighted average of these relations: $R = [r(a, b)] = [\sum_{j=1}^{m} w_j r_j(a, b)]$ with $w_j \geq 0$ and $\sum_{j=1}^{m} w_j = 1$ is also T_L-transitive.*

Theorem 3. *[14] Let $R_j = [r_j(a, b)], j = 1, \cdots, m$ be m T_L-transitive fuzzy relations. Then, $R = [r(a, b)]$ where $r(a, b) = Agg(r_1(a, b), \cdots, r_m(a, b))$ is T_L-transitive iff the De Morgan's dual of $Agg(x), x \in [0, 1]^m$, defined as $N(x) = 1 - Agg(1 - x_1, \cdots, 1 - x_m)$, satisfies the following condition: $\forall x, y, z \in [0, 1]^m, x = y + z \implies N(x) \leq N(y) + N(z)$.*

Based on the pessimistic aggregated fuzzy relation implemented with the min-max or sum-product weighted aggregation of T_{L}-transitive relations, this paper investigates two types of ordered weighted aggregation of fuzzy relations, as follows.

3. Novel OWA of Fuzzy Relations

As indicated earlier, the associations between fuzzy relations and similarity/distance metrics have been widely studied. According to [26], aggregating operations can also be interpreted in terms of distance measurements. In particular, when appropriate weights are selected, an OWA operator can act as a distance metric, being a positive mapping which satisfies identity, symmetry, and triangle inequality [9, 27]. In essence, OWA operators form a family of aggregation procedures which may be seen as a special type of weighted average based on the ordering of their arguments. The fundamental property of this family of operators is the reordering step in which the arguments are rearranged in descending order and subsequently integrated into a single aggregated value.

Definition 4. *[4] A mapping $A : \mathbb{R}^m \to \mathbb{R}$ is called an OWA operator if*

$$A(x_1, \cdots, x_m) = \sum_{j=1}^{m} w_j x_{\pi(j)}$$

where $x_{\pi(j)}$ is a permutation of $x_j \in \mathbb{R}, j = 1, \cdots, m$, which satisfies that $x_{\pi(j)}$ is the j-th largest amongst all $x_j, j = 1, \ldots, m$, and $w_j \in [0, 1], j = 1, \cdots, m$ is a collection of weights that jointly satisfy $\sum_{j=1}^{m} w_j = 1$.

Three special cases of this type of operator are: arithmetic mean, maximum, and minimum. The mean operator results by setting $w_j = 1/m, j = 1, \cdots, m$, the maximun by $w_1 = 1$ and $w_j = 0$ for $j \neq 1$, and the minimum by $w_m = 1$ and $w_j = 0$ for $j \neq m$. These weighting vectors are denoted as W_{mean}, W_{max} and W_{min} respectively in the remainder of this paper. Obviously, an important feature of the OWA operator is that it is a "mean" operator which satisfies:

$$\min\{x_1, \cdots, x_m\} \leq \sum_{j=1}^{m} w_j x_{\pi(j)} \leq \max\{x_1, \cdots, x_m\}.$$

Such an operator provides aggregation between the maximum and the minimum of the arguments. This boundedness implies that it is idempotent; that is, if all $x_j = constant$ then $A(x_1, \cdots, x_m) = constant$. For presentational simplicity, the weights of an ordered weighted aggregation are hereafter denoted as a weighting vector $W = (w_1, \cdots, w_m)$, in which the j-th component is w_j. Different choices of the weighting vector W can lead to different aggregation results. The ordering of the arguments normally implies the nonlinear nature of the OWA operators.

3.1. Ordered Weighted Aggregations of Fuzzy Similarity Relations

It has proven [9] that if an OWA weighting vector satisfies the buoyancy property (e.g., a special case of Choquet integral where a symmetric submodular fuzzy measure is used), $w_i \geq w_j$ for $i < j$ $(i, j = 1, \cdots, m)$, the corrsponding OWA operator manifests the properties of a norm and hence, it can be used to form a distance metric. The resulting metric has found successful applications in group decision making [28] and in semi-supervised clustering [27]. However, there are many other applications where fuzzy similarities are adopted instead of distance metrics [29, 30, 31]. Thus, further development of OWA operators by exploiting the aggregation of fuzzy relations would benefit such applications. Inspired by this observation, two types of ordered weighted aggregation are introduced here, using the min-max and sum-product operators respectively.

Definition 5. *Let $R_j = [r_j(a, b)], j = 1, \cdots, m$ be m fuzzy relations. Then, the ordered weighted aggregation $R^{min} = [r^{min}(a, b)], a, b \in X$, of component relations $R_1, \cdots, R_m$, implemented with the weighting vector $(w_1, \cdots, w_m)$ and min-max operator is a relation $R^{\min}$ over $X \times X$ such that*

$$r^{\min}(a, b) = \min_{j=1,\cdots,m} \max(1 - w_j, r_{\pi(j)}(a, b)) \tag{6}$$

where $r_{\pi(j)}(a, b)$ is a permutation of $r_j(a, b), j = 1, \cdots, m$, which satisfies that $R_{\pi(j)}(a, b)$ is the j-th largest of the $r_j(a, b), j = 1, \ldots, m$, and $w_j \in [0, 1], j = 1, \cdots, m$ is a collection of weights that jointly satisfy $\max_{j=1,\cdots,m} w_j = 1$.

Definition 6. *Let $R_j = [r_j(a, b)], j = 1, \cdots, m$ be m fuzzy relations. Then, the ordered weighted aggregation $R^L = [r^L(a, b)], a, b \in X$, of component relations $R_1, \cdots, R_m$, implemented with the weighting vector $(w_1, \cdots, w_m)$ and sum-product operator is a relation R^L over $X \times X$ such that*

$$r^L(a, b) = A(r_1(a, b), \cdots, r_m(a, b)) = \sum_{j=1}^{m} w_j r_{\pi(j)}(a, b) \tag{7}$$

where $r_{\pi(j)}(a, b)$ is a permutation of $r_j(a, b), j = 1, \cdots, m$, which satisfies that $r_{\pi(j)}(a, b)$ is the j-th largest of the $r_j(a, b), j = 1, \ldots, m$, and $w_j \in [0, 1], j = 1, \cdots, m$ is a collection of weights that jointly satisfy $\sum_{j=1}^{m} w_j = 1$.

The above two aggregated relations $R^{\min}$ and R^{L} implement two different mappings from multiple similarity relations onto one relation: $R^m \to R$. The constraint over the weights in $R^{\min}$ is different from that over those in R^{L}. This is due to the requirement of Definition 2 that when all the component relations $R_1, \cdots, R_m$ are zero, the constraint $\max_{j=1,\cdots,m} w_j = 1$ must ensure $R^{\min} = 0$. Obviously, both $R^{\min}$ and R^{L} are "mean" operators which satisfy the boundedness

$$\min_{j=1,\cdots,m} r_j(a,b) \le r^{\min}(a,b), r^{\mathrm{L}}(a,b) \le \max_{j=1,\cdots,m} r_j(a,b).$$

Importantly, as these two aggregators are designed for combining fuzzy relations, the reflexivity, symmetry and T-transitivity need to be considered. It is straightforward to prove that the aggregated relations $R^{\min}$ and R^{L} preserve the reflexivity and symmetry if $R_1, \cdots, R_m$ are themselves reflexive and symmetric. However, the aggregated relation does not always display T-transitivity.

The discussion of transitivity and symmetry on fuzzy relations has drawn much attention, especially on their effectiveness and interpretation in real applications. When an inference process is prototype-based, such as generating clusters of patterns that are similar to certain prototypical samples, such properties do not seem to be a necessary condition. However, there are many other situations that may require homogeneous similarities and clusters/granules whose members satisfy symmetry and transitive property. Clusters of this type are easy to recognise from a practical point of view, and the knowledge extracted from one of such clusters may also be applied in the same fashion to the rest. It is owing to this observation that symmetry and transitivity in similarity relations are considered to be very useful properties for knowledge extraction in many scenarios. Thus, the $T_{\min}$-transitivity and T_{L}-transitivity of the proposed $R^{\min}$ and R^{L} are investigated. Although transitivity is not always preserved in these aggregation procedures, it is proven that transitivity can be retained by the aggregated results if the weighting vectors employed satisfy certain constraints. The relevant theoretical development is summarised in the following two theorems.

Theorem 4. *Let $R_1, \cdots, R_m$ be $T_{\min}$-transitive relations and $(w_1, \cdots, w_m)$ be the weighting vector in $R^{\min}$ such that $w_i \le w_j$ for $i < j$, then $R^{\min}$ is $T_{\min}$-transitive.*

Proof. Without losing generality, suppose that $\pi^1(j)$, $\pi^2(j)$ and $\pi^3(j)$ are three permutations of $j = 1, \cdots, m$, such that $r_{\pi^1(j)}(a,b)$, $r_{\pi^2(j)}(a,c)$ and $r_{\pi^3(j)}(c,b)$ are the j-th largest value in $\{r_1(a,b), \cdots, r_m(a,b)\}$, $\{r_1(a,c), \cdots, r_m(a,c)\}$ and $\{r_1(c,b), \cdots, r_m(c,b)\}$, respectively, and that $w'_j = 1 - w_j$. For all $a, b, c \in X$, since $R_1, \cdots, R_m$ are $T_{\min}$-transitive, then

$$r^{\min}(a,b) = \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^1(j)}(a,b)) \ge \min_{j=1,\cdots,m} \max\left(w'_j, \min(r_{\pi^1(j)}(a,c), r_{\pi^1(j)}(c,b))\right).$$

Because of the distributivity of max over the min operator and the associativity of min, the right side of Eqn. (3.1) equals

$$\min_{j=1,\cdots,m} \min\left(\max(w'_j, r_{\pi^1(j)}(a,c)), \max(w'_j, r_{\pi^1(j)}(c,b))\right) = \min\left(\min_{j=1,\cdots,m} \max(w'_j, r_{\pi^1(j)}(a,c)), \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^1(j)}(c,b))\right). \quad (8)$$

Given that $w_i \le w_j \Rightarrow w'_i \ge w'_j$ for $i < j$ and $r_{\pi^2(i)}(a,c) \ge r_{\pi^2(j)}(a,c)$, then $r^{\min}(a,c) = \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^2(j)}(a,c))$ is equal to $\max(w'_m, r_{\pi^2(m)}(a,c))$, which is the minimum value amongst all the permutations of $r_1(a,c), \cdots, r_m(a,c)$ combined with $w'_1, \cdots, w'_m$ where $w'_i \ge w'_j$ for $i < j$. Then,

$$r^{\min}(a,c) = \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^2(j)}(a,c)) \le \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^1(j)}(a,c)),$$

and similarly,

$$r^{\min}(c,b) = \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^3(j)}(c,b)) \le \min_{j=1,\cdots,m} \max(w'_j, r_{\pi^1(j)}(c,b)).$$

Thus, the expression given in Eqn. (8) is greater or equal to

$$\min\big(\min_{j=1,\cdots,m}\max(w'_j, r_{\pi^2(j)}(a,c)), \min_{j=1,\cdots,m}\max(w'_j, r_{\pi^3(j)}(c,b))\big) = \min(r^{\min}(a,c), r^{\min}(c,b)).$$

□

Theorem 5. *Let $R_1,\cdots,R_m$ be T_L-transitive relations and $(w_1,\cdots,w_m)$ be the weighting vector in T_L such that $w_i \leq w_j$ for $i<j$, then R^L is T_L-transitive.*

Proof. In general, suppose that $\pi^1(j)$, $\pi^2(j)$ and $\pi^3(j)$ are three permutations of $j=1,\cdots,m$, such that $r_{\pi^1(j)}(a,b)$, $r_{\pi^2(j)}(a,c)$ and $r_{\pi^3(j)}(c,b)$ are the j-th largest value in $\{r_1(a,b),\cdots,r_m(a,b)\}$, $\{r_1(a,c),\cdots,r_m(a,c)\}$ and $\{r_1(c,b),\cdots,r_m(c,b)\}$, respectively. For all $a,b,c\in X$, since $r_1,\cdots,r_m$ are T_{L}-transitive, then

$$\begin{aligned} r^{\mathrm{L}}(a,b) = \sum_{j=1}^{m} w_j r_{\pi^1(j)}(a,b) \geq \\ \sum_{j=1}^{m} w_j\cdot\max\big(r_{\pi^1(j)}(a,c)+r_{\pi^1(j)}(c,b)-1,0\big) \geq \sum_{j=1}^{m}\max\big(w_j r_{\pi^1(j)}(a,c)+w_j r_{\pi^1(j)}(c,b)-w_j,0\big) \geq \\ \max\left(\sum_{j=1}^{m}\big(w_j r_{\pi^1(j)}(a,c)+w_j r_{\pi^1(j)}(c,b)-w_j\big),0\right) \geq \max\big(\sum_{j=1}^{m} w_j r_{\pi^1(j)}(a,c)+\sum_{j=1}^{m} w_j r_{\pi^1(j)}(c,b)-1,0\big) \end{aligned}$$

e.g.,

$$r^{\mathrm{L}}(a,b) \geq T_{\mathrm{L}}\big(\sum_{j=1}^{m} w_j r_{\pi^1(j)}(a,c), \sum_{j=1}^{m} w_j r_{\pi^1(j)}(c,b)\big) \tag{9}$$

Because the sum-product $\sum_{j=1}^{m} w_j r_{\pi(j)}(a,c)$ with $w_i \leq w_j$ for $i<j$ achieves its minimal value amongst all the permutations of $r_1(a,c),\cdots,r_m(a,c)$ when $\pi(j)=\pi^2(j)$, e.g., when $r_1(a,c),\cdots,r_m(a,c)$ are in descending order and $w_{j=1,\cdots,m}$ are in ascending order. Therefore, $\sum_{j=1}^{m} w_j r_{\pi^1(j)}(a,c) \geq \sum_{j=1}^{m} w_j r_{\pi^2(j)}(a,c)$ and similarly, $\sum_{j=1}^{m} w_j r_{\pi^1(j)}(c,b) \geq \sum_{j=1}^{m} w_j r_{\pi^3(j)}(c,b)$. Then, the right side of Eqn. (9) is greater than or equal to

$$T_{\mathrm{L}}\big(\sum_{j=1}^{m} w_j r_{\pi^2(j)}(a,c), \sum_{j=1}^{m} w_j r_{\pi^3(j)}(c,b)\big) = T_{\mathrm{L}}(r^{\mathrm{L}}(a,c), r^{\mathrm{L}}(c,a))$$

□

3.2. Decision on the Weighting Vector

A common pitfall with existing aggregation operators is the inability to provide an explanatory means by which a user can utilise to enhance the individual perception of arguments' significance. To resolve this short coming, the stress function has been introduced [32] as a simple mechanism for attaining interpretability in OWA, which formalises characterisation (andness/orness) of the resultant OWA operators. This can be accomplished using a function $h:[0,1]\to\mathbb{R}^+$ to stress positions where significant values stand out from the weighting vector. Formally, a weighting vector of OWA can be defined by a stress function h such as that is given below.

Definition 7. *[32] Let $h:[0,1]\to\mathbb{R}^+$ be a non-negative function on the unit interval. The OWA weighting vector $W=(w_1,\cdots,w_j,\cdots,w_m)$ can then be defined by:*

$$w_j = \frac{h(\frac{j}{m})}{\sum_{j=1}^{m} h(\frac{j}{m})} \tag{10}$$

Such a function h is termed a stress function for OWA.

The OWA weighting vector obtained by the use of a stress function has a number of helpful features. For instance, values from a stress function $h(x)$ associated with the lower portion of the left side of $[0,1]$ reflect those weights associated with the larger argument values, while the values associated with the right side of the unit interval reflect the weights associated with the smaller values in the aggregation. Other properties are omitted here but can be found in [27, 32, 33].

Different types of stress function can be used to express, and to impose constraints over, the distribution of weights, thereby resulting in different aggregation behaviours. The overall behaviour of an aggregation operator can be described by the so-called attitudinal character measure. It gives an estimation of whether an aggregation operator behaves similarly to conjunction/andness (influenced by smaller argument values) or disjunction/orness (influenced by larger values) [34]. In particular, a useful attitudinal character measure of an OWA operator with the weighting vector W is:

$$\text{A-C}(W) = \frac{1}{m-1}\sum_{j=1}^{m}((m-j)w_j). \tag{11}$$

Interestingly, note that $\text{A-C}(W_{\text{mean}}) = 0.5$, $\text{A-C}(W_{\max}) = 1$ and $\text{A-C}(W_{\min}) = 0$, where W_{mean}, $W_{\max}$ and $W_{\min}$ denote the weighting vectors used in the conventional aggregation operators that are implemented by the arithmetic average, maximum and minimum, respectively.

The concepts of attitudinal character and stress function can be extended to the proposed aggregations of fuzzy relations. Since the constraints on the weighting vector in $R^{\min}$ is different from Eqn. (10), a modification regarding the normalisation is needed. An intuitive modification is: $w_j = h(j/m)/\max_{j=1,\cdots,m} h(j/m)$. This implies that the measure of attitudinal character for $R^{\min}$ is normalised such that

$$\text{A-C}'(W) = \frac{\text{A-C}(W)}{\sum_{j=1}^{m} w_j}. \tag{12}$$

However, obtaining the weighting vector from a stress function for R^{L} remains the same as with Eqn. (10), this is because the weighting vector of R^{L} satisfies $\sum_{j=1}^{m} w_j = 1$. Hence, $\text{A-C}'(W) = \text{A-C}(W)$ for R^{L}.

As an example, consider three data points $\alpha, \beta, \gamma \in X$. Let $A_{1,\cdots,4}$ denote four fuzzy sets $A_1, \cdots, A_4$ as shorthand, representing four fuzzy terms defined on X. Suppose that $\mu_{1,\cdots,4}(\alpha) = (0.63, 0.94, 0.97, 0.62)$, representing that $\mu_{A_1}(\alpha) = 0.63$, $\mu_{A_2}(\alpha) = 0.94$, $\mu_{A_3}(\alpha) = 0.79$ and $\mu_{A_4}(\alpha) = 0.62$ (with the rest below representing similar information in the same manner), $\mu_{1,\cdots,4}(\beta) = (0.01, 0.49, 0.25, 0.97)$ and $\mu_{1,\cdots,4}(\gamma) = (0.68, 0.91, 0.62, 0.68)$. Denote $T_{\min}$-transitive and T_{L}-transitive fuzzy relations by $T_{\min}$ and T_{L} norms, respectively. Thus, $r_j^{\min}(\alpha,\beta) = \min(\mu_{A_j}(\alpha), \mu_{A_j}(\beta))$ and $r_j^{\text{L}}(\alpha,\beta) = 1 - |\mu_{A_j}(\alpha) - \mu_{A_j}(\beta)|$, $j = 1, \cdots, 4$. Then, the corresponding $T_{\min}$-transitive similarity relations are: $r_{1,\cdots,4}^{\min}(\alpha,\beta) = (0.01, 0.49, 0.25, 0.62)$, $r_{1,\cdots,4}^{\min}(\alpha,\gamma) = (0.63, 0.91, 0.62, 0.62)$, $r_{1,\cdots,4}^{\min}(\gamma,\beta) = (0.01, 0.49, 0.25, 0.68)$, and the T_{L}-transitive ones over the three examples are: $r_{1,\cdots,4}^{\text{L}}(\alpha,\beta) = (0.38, 0.55, 0.28, 0.65)$, $r_{1,\cdots,4}^{\text{L}}(\alpha,\gamma) = (0.95, 0.97, 0.65, 0.94)$, $r_{1,\cdots,4}^{\text{L}}(\gamma,\beta) = (0.33, 0.58, 0.63, 0.71)$.

From the above, Table 1 summarises the resulting weights obtained from the application of the present approach and Table 2 shows the aggregated results of $R_{1,\cdots,4}^{\min}$ and $R_{1,\cdots,4}^{\text{L}}$ using Eqns. (6) and (7), respectively.

It can be seen from the example that generally, the results of aggregated relations are affected by the specification of the stress function. Higher attitudinal character values will result in higher-valued aggregated similarities. The example also demonstrates that $R^{\min}$ dose not preserve $T_{\min}$-transitive in general, as under W_1, $r^{\min}(\alpha,\beta) < \min(r^{\min}(\alpha,\gamma), r^{\min}(\gamma,\beta))$. Similarly, under W_1 and W_2, $r^{\text{L}}(\alpha,\beta) < T_{\text{L}}(r^{\text{L}}(\alpha,\gamma), r^{\text{L}}(\gamma,\beta))$, which means that R^{L} dose not generally preserve T_{L}-transitive either. However, with the assistance of Theorem 3, it can be proven that in general, when the component relations are T_{L}-transitive in R^{L}, the aggregated relation is T_{L}-transitive if and only if the weighting vector satisfies the additional condition that $w_i \leq w_j$ for $i < j$ (see Appendix A).

To further illustrate how the changes in weighting vectors may affect the behaviour of $R^{\min}$ and R^{L}, an example involving a two dimensional dataset is employed. Each dimension represents the membership value of a data point belonging to a certain given fuzzy set. The Z-axises (similarities) in Figs. 1 and 2

Table 1: Examples of Stress Function, where $x \in [0,1]$

	$h(x) =$	Weighting Vector	A-C$'(W)$
W_1	1, for $x = 0$;	$R^{\min}: (1.00, 0.00, 0.00, 0.00)$	1.00
	0, otherwise.	$R^{\mathrm{L}}: (1.00, 0.00, 0.00, 0.00)$	
W_2	$1.25 - x$	$R^{\min}: (1.00, 0.75, 0.50, 0.25)$	0.63
		$R^{\mathrm{L}}: (0.40, 0.30, 0.20, 0.10)$	
W_3	*constant* $\in (0,1]$	$R^{\min}: (1.00, 1.00, 1.00, 1.00)$	0.50
		$R^{\mathrm{L}}: (0.25, 0.25, 0.25, 0.25)$	
W_4	x	$R^{\min}: (0.25, 0.50, 0.75, 1.00)$	0.33
		$R^{\mathrm{L}}: (0.10, 0.20, 0.30, 0.40)$	
W_5	0, for $x = 0$;	$R^{\min}: (0.00, 0.00, 0.00, 1.00)$	0.00
	1, otherwise.	$R^{\mathrm{L}}: (0.00, 0.00, 0.00, 1.00)$	

Table 2: Aggregated Result of Examples

W	W_1	W_2	W_3	W_4	W_5
$r^{\min}(\alpha,\beta)$	0.6200	0.4900	0.0100	0.0100	0.0100
$r^{\min}(\alpha,\gamma)$	0.9100	0.6200	0.6200	0.6200	0.6200
$r^{\min}(\gamma,\beta)$	0.6800	0.4900	0.0100	0.0100	0.0100
$r^{\mathrm{L}}(\alpha,\beta)$	0.6500	0.5290	0.4650	0.4010	0.2800
$r^{\mathrm{L}}(\alpha,\gamma)$	0.9700	0.9260	0.8775	0.8290	0.6500
$r^{\mathrm{L}}(\gamma,\beta)$	0.7100	0.6220	0.5625	0.5030	0.3300

outline the aggregated similarities of every possible point in the problem space, $\beta \in X$, to a certain point α, with $\mu_{A_1}(\alpha) = 0.5$ and $\mu_{A_2}(\alpha) = 0.5$ being the memberships of α belonging to the two fuzzy sets A_1 and A_2 shown. That is, $Z = r^{\min}(\alpha,\beta)$ in Fig. 1 and $Z = r^{\mathrm{L}}(\alpha,\beta)$ in Fig. 2, where $\mu_{1,2}(\alpha) = (0.5, 0.5)$ and $\mu_{1,2}(\beta) \in [0,1] \times [0,1]$.

Note that when non-transitive weighting vectors (A-C$'(W) > 0.5$) are used, both of the surfaces in Figs. 1 and 2 project non-convex contour lines, reflecting the non-transitivity of the aggregation. It can be seen from this example that as the proposed aggregation is applied to such real problems, engineering users will have the degree of freedom to control both the T-transitivity and the behaviour (optimistic or pessimistic) of the $R^{\min}$ and R^{L}, by tuning the stress function, or by tuning the weights in a weighting vector. With the assistance of attitudinal character, engineers can have an intuitive interpretation of how optimistic or pessimistic a certain aggregator is, so that the tuning of the system parameters is not done in a "black-box" manner.

Note also that, due to the fact that aggregations based on min-max operators more easily produce discrete results than the ones based on sum-product operators, $R^{\min}$ is not so sensitive as R^{L} to the change of weights. Besides, if the weighting vectors satisfy the constraint in Theorem 4 in an effort to retain transitivity, the $R^{\min}$ aggregation will behave as the classic min operator and hence, losing the point of developing this type of aggregator (see proof in Appendix B).

4. Examples of Applying R^{L} to Hierarchical Clustering

In order to demonstrate the potential of ordered weighted aggregation of fuzzy relations in problem-solving, this section presents an application of aggregated relations to perform hierarchical data clustering.

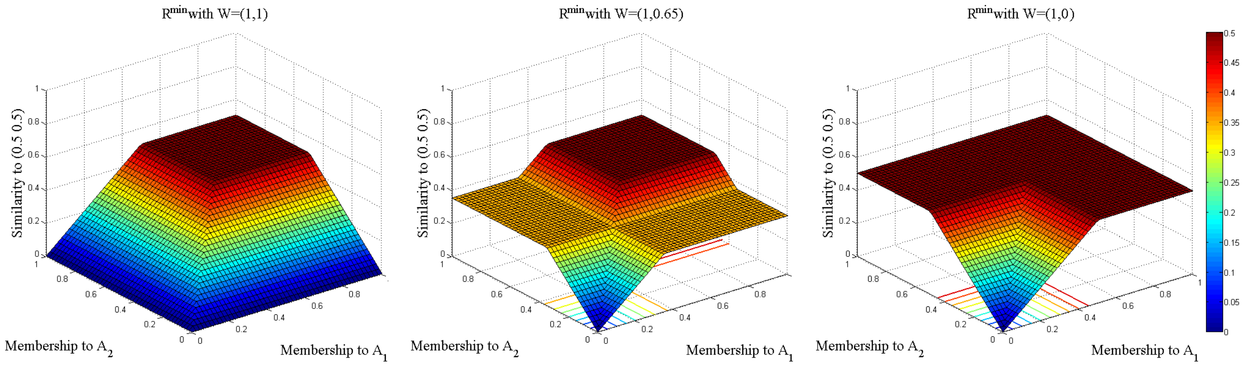


Figure 1: Trend of $R^{\min}$ Aggregated Similarity against Weighting Vector

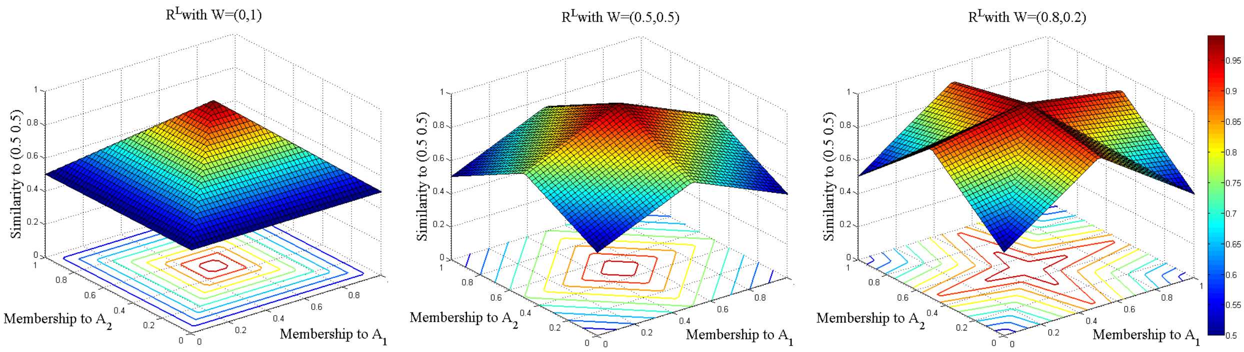


Figure 2: Trend of R^{L} Aggregated Similarity against Weighting Vector

Since similarity is fundamental to the definition of a cluster, a measure of the similarity between two patterns drawn from the same feature space is essential to most clustering procedures. This application therefore, uses the proposed OWA approach to justify the similarity of each pair of patterns, grouping similar patterns into the same cluster using a hierarchical clustering algorithm. Its performance is assessed over synthetic 2-dimensional datasets and also UCI benchmark datasets. In addition, the work further investigates the impact of different weighting vector settings upon the clustering results. Due to the relative insensitiveness of $R^{\min}$ in recognising the changes of the weights, only the results on R^{L} are reported here.

Most hierarchical clustering methods are variants of either the so-called single-link or complete-link algorithm. These two algorithms differ in the way they characterise the similarity between a pair of clusters. In single-link methods, the distance between two clusters is the minimum of the distances between all pairs of patterns separately drawn from the two clusters. In complete-link methods, the distance between two clusters is the maximum of all pairwise distances between patterns in the two clusters. In either case, two clusters are merged to form a larger cluster based on certain minimum distance criterion. Unlike classical k-means clustering, both single-link and compete-link algorithms do not involve random initialisations. The complete-link hierarchical clustering is selected herein to test the performance of the proposed aggregation of fuzzy relations as similarity metrics in clustering analysis. This is because the complete-link algorithm generally produces tightly bound or compact clusters [35].

A component fuzzy relation can be regarded as a similarity/dissimilarity metric amongst data points, based on a certain feature. The OWA-aggregated fuzzy relation reflects the overall estimation of component similarities which are based on each considered feature. Without losing generality, in this work, it is assumed that each feature is described by a number of fuzzy sets, defined on the relevant underlying problem domain. Formally, given a dataset of N points $(p_1, \cdots, p_N)$, each point $p_a, a = 1, \cdots, N$ is described by M features, and the j-th feature is expressed by a set of membership functions $\mu^{jk}, k = 1, \cdots, L_j$. To have a common representation for all features, the number of linguistic labels naming the fuzzy sets for each feature is extended to $max_{j\in\{1,\ldots,L_M\}} L_j$, denoted as L. This does not lose generality as for those features whose original number of fuzzy labels is smaller than L, artificial labels can be created by copying the last label (or the first if so desired) to fill the gap. As such, each data point p_a is characterised by the following: $((\mu_a^{11}, \cdots, \mu_a^{1L}), \cdots, (\mu_a^{j1}, \cdots, \mu_a^{jl}, \cdots, \mu_a^{jL}), \cdots, (\mu_a^{M1}, \cdots, \mu_a^{ML})), j = 1, \ldots, M$.

From this, fuzzy relations between points can be first built with respect to each individual feature. Then, the proposed ordered weighted aggregation can be employed to aggregate the similarities evaluated on the basis of individual features. The hierarchical clustering mechanism using the proposed R^{L} is therefore summarised as follows:

- Step 1. Acquire the fuzzy similarity relations $r_j(p_a, p_b)$ based on the j-th feature, $j = 1, \ldots, M$. According to [14], the T_{L}-transitive similarity relation involving L linguistic labels can be obtained by:

$$r_j^{\mathrm{L}}(p_a, p_b) = \inf_{l\in\{1,\cdots,L\}} (1 - |\mu_a^{jl} - \mu_b^{jl}|) \tag{13}$$

- Step 2. Aggregate r_j^{L}, $j = 1, \cdots, M$ using R^{L}, i.e., Eqn. (7).
- Step 3. Apply complete-link hierarchical clustering based on the aggregated similarities.

4.1. Experimental Setup

To evaluate the performance of utilising R^{L}-aggregated fuzzy relations for clustering, it is experimentally tested over six synthetic 2-dimensional datasets (see Fig. 3) and six UCI standard benchmark datasets whose variables are continuously valued [36], where the underlying true labels of the instances are known but are not explicitly used in the clustering process. Details of these datasets are summarised in Table 3.

Both the normalisation and fuzzification of original UCI datasets are implemented (synthetic datasets are only fuzzified). For a point $p_a, a = 1, \cdots, N$, its j-th feature value $F_j(p_a) \in \mathbb{R}, j = 1, \cdots, M$ is normalised to $F_j'(p_a) \in [0, 1]$ by

$$F_j'(p_a) = \frac{F_j(p_a) - \min\limits_{i=1,\cdots,N}(F_j(p_i))}{\max\limits_{i=1,\cdots,N}(F_j(p_i)) - \min\limits_{i=1,\cdots,N}(F_j(p_i))}. \tag{14}$$

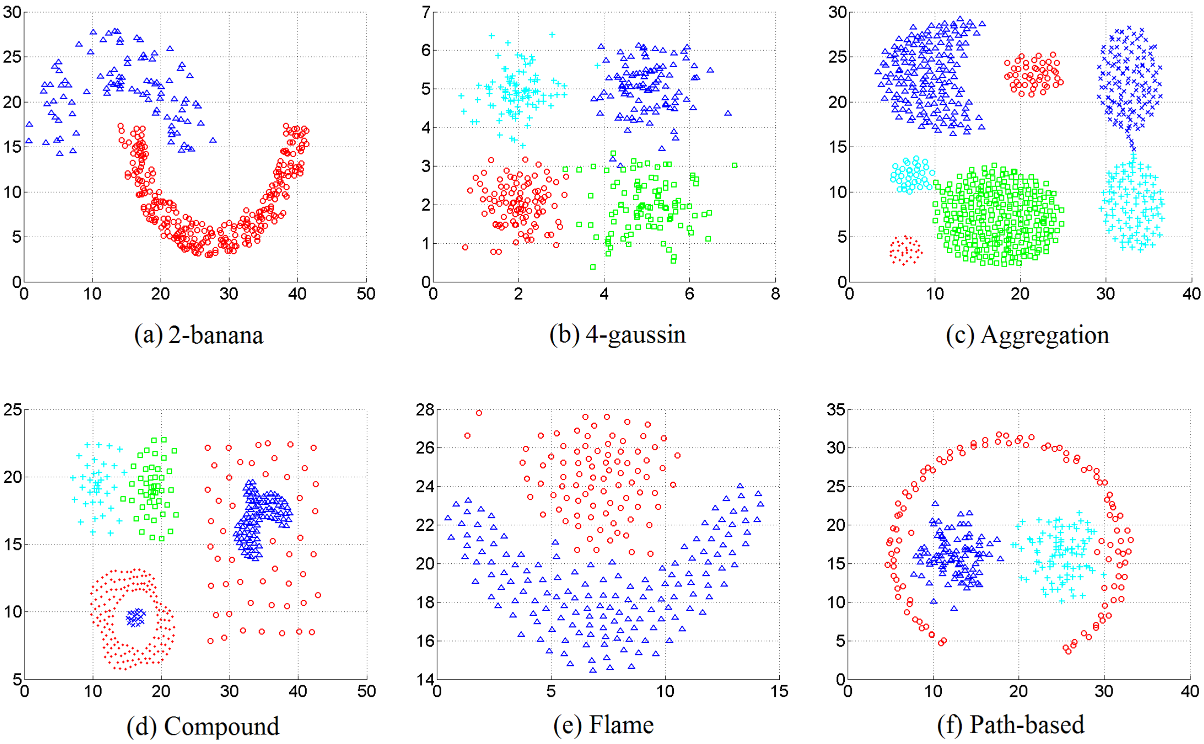


Figure 3: Synthetic 2-dimensional Datasets

Table 3: Summary of Datasets Used

Datasets	#Instances	#Features (M)	#Clusters	Resource
2-banana	373	2	2	[37]
4-gaussin	400	2	4	generated
Aggregation	788	2	7	[38]
Compound	399	2	6	[39]
Flame	240	2	2	[40]
Path-based	300	2	3	[41]
Ecoli	336	7	8	[42]
Glass	214	9	6	[43]
Ionosphere	351	34	2	[44]
Iris	150	4	3	[45]
Thyroid	215	5	3	[46]
Wine	178	13	3	[47]

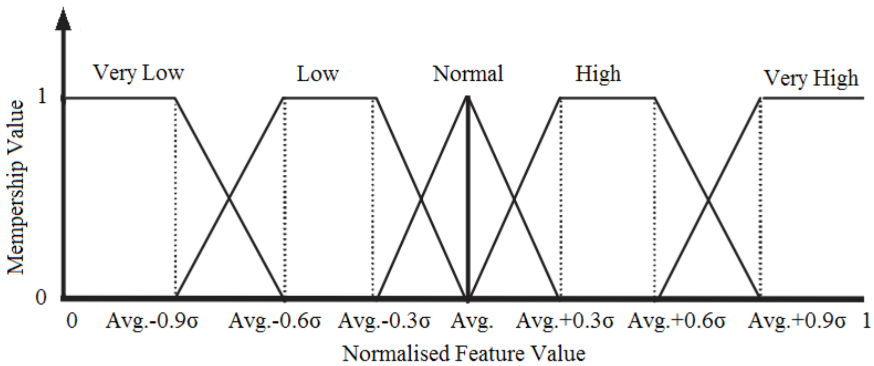


Figure 4: Fuzzification of Feature Values

In fuzzification, every normalised feature value is transformed into a set of five membership functions as depicted in Fig. 4 [14].

To facilitate direct comparison, the number of clusters is set to the the number of given classes per dataset. The clustering results are evaluated in terms of accuracy as the ground truth of the clusters for each dataset is known. To examine the relationship between clustering accuracy and the attitudinal character of the R^{L} weighting vector, 21 weighting vectors are generated using linear stress functions with the attitudinal character values distributed from zero to one. **As it is shown in Fig. 5, an aggregator of minimum can be gained by using SF$_1$ as the stress function in OWA (i.e., $W_1 = (0, \cdots, 0, 1)$), while an aggregator of average can be gained by using SF$_{11}$ (i.e., $W_{11} = (1/m, \cdots, 1/m)$, where m is the number of features). By reshaping the stress function from SF$_1$ to SF$_{11}$, the value of attitudinal character can be tuned from A-C$'(W_1) = 0.0$ to A-C$'(W_{11}) = 0.5$. Weighting vectors with A-C$'(W) \in (0.5, 1]$ are in reverse order of the weights within W_{10} to W_1, respectively (e.g., $W_{21} = (1, 0, \cdots, 0)$ whose weights are in reverse order of $W_1 = (0, \cdots, 0, 1)$).** The weighting vectors with A-C$'(W) \in [0, 0.5]$ preserve transitivity while the others (A-C$'(W) \in (0.5, 1]$) do not. For each weighting vector on each dataset, the clustering algorithm runs only once, since the complete-link hierarchical clustering does not involve any random parameter initialisation and multiple runs would only generate identical outputs.

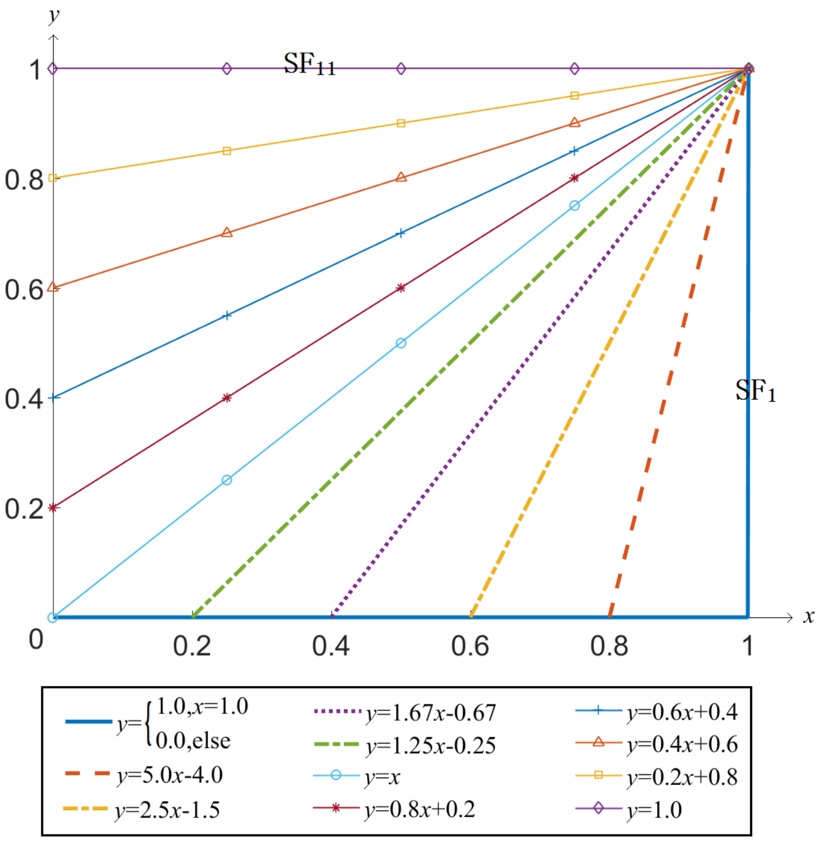
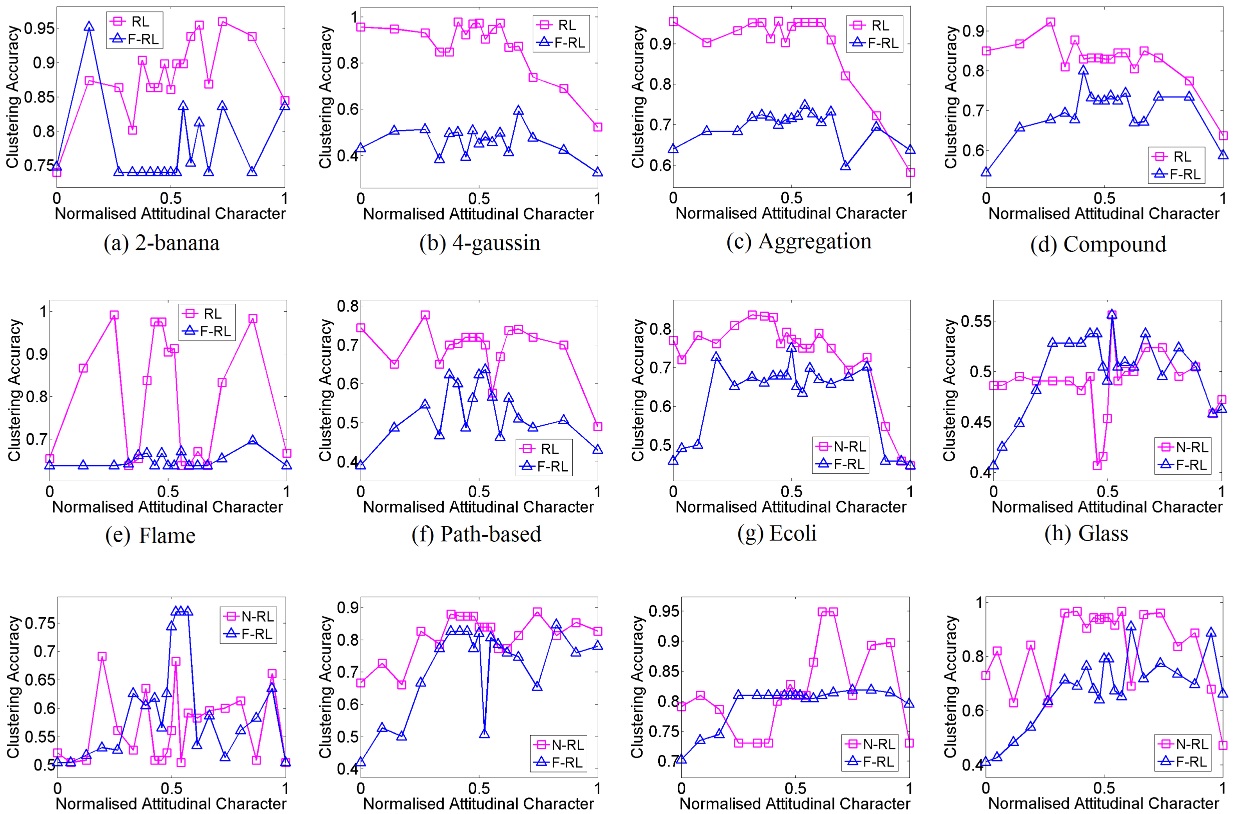


Figure 5: Stress Functions for Weights Generating



(a) 2-banana (b) 4-gaussin (c) Aggregation (d) Compound

(e) Flame (f) Path-based (g) Ecoli (h) Glass

(i) Ionosphere (j) Iris (k) Thyroid (l) Wine

Figure 6: Trend of Accuracy Change against Attitudinal Character

4.2. Results and Discussion

Figure 6 shows the change of accuracy (Y-axis) with respect to the (normalised) attitudinal character value of the weighting vectors (X-axis). "N/F-RL" represents normalisation/fuzzification of the datasets with R^{L} similarity based hierarchical clustering.

It can be seen from Fig. 6 that the final clustering result is sensitive to the A-C$'(W)$ value of a weighting vector. In particular, regarding the 4-gaussin and Aggregation datasets, the accuracy of "RL" generally decreases as A-C$'(W)$ increases. This shows a preference for T_{L}-transitive similarities. A common feature of these two synthetic datasets is that the majority of their data points are in the clusters of convex shapes. In the 2-banana dataset, where the clusters form non-convex shapes, the accuracy of "RL" shows an ascending trend with the increase of A-C$'(W)$. On the Flame dataset, where nearly 50% points are in the cluster of a convex shape, whilst the others are in the cluster of a non-convex shape, the clustering result vibrates drastically when the A-C$'(W)$ increases. It can be concluded from these results that T_{L}-transitive similarities (A-C$'(W) \in [0, 0.5]$) are preferred in the construction of convex clusters (for complete-link hierarchical clustering), while A-C$'(W) \in (0.5, 1]$ are preferred in the construction of non-convex clusters. However, for datasets of both convex clusters and non-convex clusters (especially when their boundaries are close to each other), the selection of an appropriate weighting vector can be a challenge.

To compare the results obtainable by the T_{L}-transitive aggregation against those not T_{L}-transitive, the average accuracies (with standard deviation) and the best achievable accuracies are reported in Tables 4 and 5, for normalised and fuzzified datasets respectively. If the same best accuracy is obtained with more than two weighting vectors, their attitudinal character values are given by intervals. Note that not all the values in the interval are tested in this experiment, but only those discrete points which are shown in Fig.

Table 4: Comparison of Accuracy (%) of N-R^{L}: T_{L}-transitive vs. Not T_{L}-transitive

	Average±Standard Deviation		Best-accuracy (A-C$'(W)$)	
	T_{L}-transitive	Not T_{L}-transitive	T_{L}-transitive	Not T_{L}-transitive
Ecoli	**78.84±3.63**	66.76±13.15	**83.63 (0.33)**	78.87 (0.62)
Glass	47.20±3.22	**50.23±2.76**	49.53 (0.11, 0.42)	**55.61 (0.52)**
Ionosphere	54.98±6.06	**58.43±6.26**	**69.13 (0.20)**	68.26 (0.52)
Iris	78.85±9.17	**82.47±3.44**	88.00 (0.38)	**88.67 (0.75)**
Thyroid	78.18±3.58	**84.09±8.54**	82.79 (0.48)	**94.88 ([0.62,0.67])**
Wine	**84.63±12.91**	83.09±16.50	**96.63 (0.39)**	96.63 (0.57)

Table 5: Comparison of Accuracy (%) of F-R^{L}: T_{L}-transitive vs. Not T_{L}-transitive

	Average±Standard Deviation		Best-accuracy (A-C$'(W)$)	
	T_{L}-transitive	Not T_{L}-transitive	T_{L}-transitive	Not T_{L}-transitive
Ecoli	**63.18±9.99**	60.53±10.62	**75.00 (0.50)**	70.24 (0.82)
Glass	49.24±4.68	**50.56±3.00**	53.74 (0.42,0.46)	**55.61 (0.52)**
Ionosphere	57.87±7.34	**62.26±10.82**	74.35 (0.50)	**76.96 ([0.52, 0.57])**
Iris	67.09±17.09	**74.27±9.66**	82.67 ([0.38,0.45])	**84.67 (0.83)**
Thyroid	77.72±4.62	**80.93±0.72**	80.93 ([0.25,0.5])	**81.86 ([0.75,0.83])**
Wine	61.59±13.22	**75.06±9.07**	79.21 (0.50)	**91.01 (0.61)**

6. Since it is difficult to define a non-transitive counterpart for a transitive aggregation, the paired t-tests are not available for this comparison.

In terms of hierarchical clustering accuracy regarding the selected UCI datasets, the ascending/descending trend with the increase of A-C$'(W)$ is not so obvious. This differs from the cases on the 2-banana, 4-gaussin and Aggregation datasets. Overall, the accuracies of clustering with non T_{L}-transitive similarities are better than those achievable by T_{L}-transitive ones. However, the standard deviations of the results are slightly higher, which shows that the result of hierarchical clustering is sensitive to the A-C$'(W)$ value of weighing vectors. Nevertheless, non T_{L}-transitive results outperform T_{L}-transitive ones on both normalised and fuzzified datasets. This demonstrates that weighting vectors with higher A-C$'(W)$ (orness) are more preferable on problems that exhibit similar properties as those selected UCI datasets.

Note that when W_{mean} is applied to R^{L} aggregation, the resultant hierarchical clustering on normalised datasets is identical to Manhattan distance based complete-link hierarchical clustering (refer to the point on line "RL" and "N-RL" with A-C$'(W) = 0.5$ in Fig. 6). By comparing the positions where the R^{L} aggregated similarity relation obtains its best accuracy, it can be concluded that the proposed approach has the potential of providing better clusters than the classic Manhattan distance metric when they are applied to hierarchical clustering. Furthermore, R^{L} also has the potential of producing better results than the classic max and min operators in the aggregation of T_{L}-based fuzzy similarities on different features.

Clearly, the proposed aggregations are capable of dealing with fuzzy datasets. However, the accuracies achieved using R^{L} on fuzzified datasets are not necessarily better than those achieved on unfuzzified datasets. One possible explanation is that the component similarity relations defined in Eqn. (13) are "pessimistic" due to the use of the min operator. This may result in $r_j^{\mathrm{L}}(p_a, p_b) = 0$ between many data point pairs (of p_a and p_b) when trapezoid membership functions are used. In this case, less information is provided for the following hierarchical clustering than the use of unfuzzified datasets.

5. Application to Water Treatment Plant Monitoring

Environmental performance has become a critical and general issue in human society [22, 48, 49]. Creating and maintaining robust environmental indicators is essential when complex environmental quality factors need to be effectively exploited in developing and communicating environmental public policy [23, 50]. For instance, a poorly maintained industrial water treatment plant can significantly degrade the environmental performance [51]. Such a plant is likely to involve a number of similar water quality measurements, just as pointed out in [52, 53]: "interrelations between attributes are unavoidable as the plant is a single system with interconnections". Having taken notice of this, in this section, the proposed aggregation of fuzzy similarities is applied to clustering the dataset of malfunctions in an urban waste water treatment plant [36]. Also, the interactions (i.e., the T-transitivity and the degree of orness) amongst the aggregation of water quality measurements are analysed.

The water treatment dataset contains a set of historical data obtained over a period of 527 days, with one series of measurements per day. 38 real-valued indicators (attributes) are monitored per day, with one set of such measurements forming one record in the dataset. According to the elements of the water treatment plant, those measurements can be grouped into 5 aspects: input to plant (9 attributes), input to primary settler (6 attributes), input to secondary settler (7 attributes), output from plant (7 attributes), and plant performance (9 attributes). The status of the plant on a certain date is deemed to be in one of 13 different categories, some of which representing normal status (positive labels) and others malfunctions in the plant (negative labels). Note that all malfunctions appear for very short periods (usually in a single day), so there are not many examples with negative labels. For monitoring purpose, all forms of faulty behaviour are regarded as just one type of malfunction. Thus, the dataset is re-labelled into two major categories: 513 samples for positive label, and the remaining 14 samples for negative label. The re-labelled categories are shown in Table 6. For the present investigation, the dataset is normalised for each individual attribute and the missing value is replaced by the means of the values of that attribute.

Table 6: Summary of Re-labelled Water Treatment Dataset

Original Label	Description	New Label
Class 1	Normal situation	Positive
Class 2	Secondary settler problems-1	Negative
Class 3	Secondary settler problems-2	Negative
Class 4	Secondary settler problems-3	Negative
Class 5	Normal situation with performance over the mean	Positive
Class 6	Solids overload-1	Negative
Class 7	Secondary settler problems-4	Negative
Class 8	Storm-1	Negative
Class 9	Normal situation with low influent	Positive
Class 10	Storm-2	Negative
Class 11	Normal situation	Positive
Class 12	Storm-3	Negative
Class 13	Solids overload-2	Negative

Given a historical dataset as described above, the process of utilising the proposed method to detect water treatment plant malfunction can be summarised by the stages of off-line training and on-line monitoring.

5.1. *Off-line Training*

The training stage is aimed to find the best clustering structure which can describe the distribution of the negative and positive instances in the historical dataset. Meanwhile, the interaction between attributes can be quantified by the attitudinal character of weighing vector employed in the best clustering, which helps engineers to understand the formation of these clusters. The steps of off-line training is summarised as follows:

- **Step 1. Define a search space $\mathbb{W}$ of weighting vectors with different attitudinal characters.**
- **Step 2. For each weighting vector $W_k \in \mathbb{W}$, calculate the aggregated similarity matrix $R^{\mathbf{L}}_{W_k}$ amongst instances by using Eqn. (7) and (13).**
- **Step 3. For each $R^{\mathbf{L}}_{W_k}, W_k \in \mathbb{W}$, generate a clustering result of the training dataset C_{W_k} by using complete-linkage hierarchical clustering.**
- **Step 4. Select the best clustering result $C^* \in \{C_{W_k} | W_k \in \mathbb{W}\}$ as the prototype for malfunction detection by analysing the quality of each clustering result, and select the weighting vector with which C^* is generated as W^*.**

The above method starts with the definition of a search space of weighting vectors. As shown in Section 4, such a space consists of 21 linear stress functions whose attitudinal character measures are monotonically increasing with their input. Table 7 (as given in Appendix C) shows the A-C$'(W)$ and the weights of weighting vectors W_1 to W_{11}. The weights within W_{12} to W_{21} are in reverse order of the weights within W_{10} to W_1, respectively (e.g., $W_{21} = (1, 0, \cdots, 0)$,$W_1 = (0, \cdots, 0, 1)$). According to Theorem 5, the weighing vectors whose A-C$'(W)$ are in $[0, 0.5]$ (i.e., W_1 to W_{11}) preserve transitivity, while the others (i.e., W_{12} to W_{21}) do not. By searching for the best weighting vector to support clustering, the proposed method can obtain not only a quality clustering prototype for malfunction detection, but also an intuitive explanation of how the plant features may interact in revealing the working states of the plant.

By adopting **each of the weighting vectors** that are defined within the search space, the R^{L}-aggregated fuzzy similarity is used to support the complete-link hierarchical clustering, forming 13 clusters over the 527 days based on the water quality measurements. In assessing the **clustering result of each weighting vector**, the strategy of majority rule is taken; that is, all the days in a resulting cluster are labelled as positive if the majority of the true labels are positive, otherwise, all the days in that cluster are labelled as negatives. The experimental results **which are generated by all weighting vectors in** $\mathbb{W}$ are plotted in Fig. 7 for **analysis**.

An important point to notice from these results is that in terms of N-RL, the points in the interval of A-C$'(W) \in [0, 0.5]$ are generally higher than those in the interval of A-C$'(W) \in (0.5, 1]$, and its accuracy drops quickly when the A-C$'(W)$ is above 0.7. This indicates that when the R^{L} aggregated fuzzy similarity is applied to the water treatment dataset, the transitive aggregation is better than the non-transitive aggregation. It indicates that in the dataset on water treatment plant, both the positive and negative days form clusters of convex shapes, within the space delimited by the given 38 attributes. To further evaluate the quality of the clustering results generated using different weighting vectors, the precision and recall of the detected negative days are provided in Fig. 8. Here, the *precision* of negative days is calculated as the the number of days correctly labelled as negative days divided by the total number of negative days labelled by the proposed method, and *recall* is defined as the number of days correctly labelled as negative days divided by the total number of negative days that are given in the dataset (i.e., 14 for this dataset).

It is clear from Fig. 8 that the precision values of negative days are generally very high values when A-C$'(W) \in [0, 0.8]$. When the A-C$'(W)$ is above 0.8, the values of precision become invalid since the total number of the negative days labelled by the proposed method is zero, i.e., all the days are labelled as positive days. Such high precision values indicate that if a day is clustered to be negative (by the proposed method), it is very likely that the water treatment plant is indeed malfunctioning on the day. However, the values of recall are not so high as those of precision. In particular, when the A-C$'(W)$ is above 0.7 (or the aggregated fuzzy similarity is non-transitive), the values of recall are lower than 0.4. This implies that many

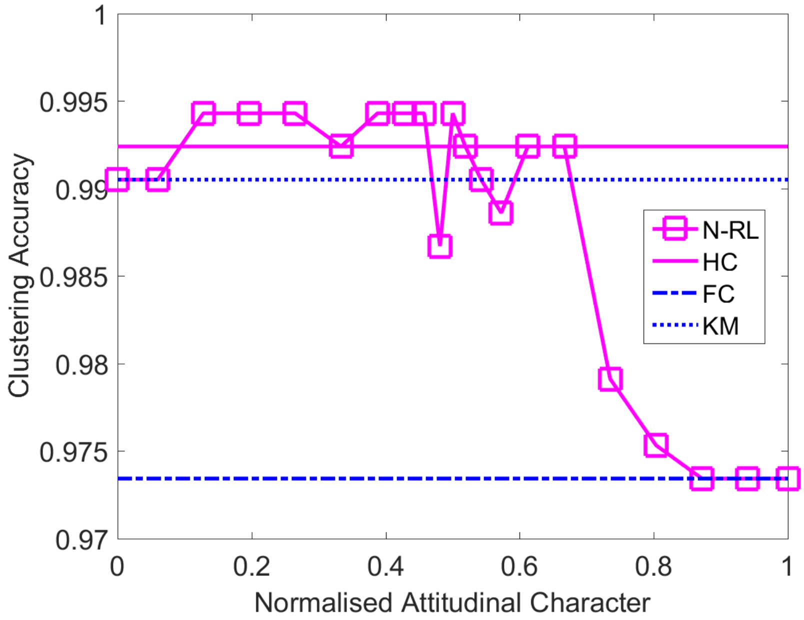


Figure 7: Clustering Accuracy on Water Treatment Dataset

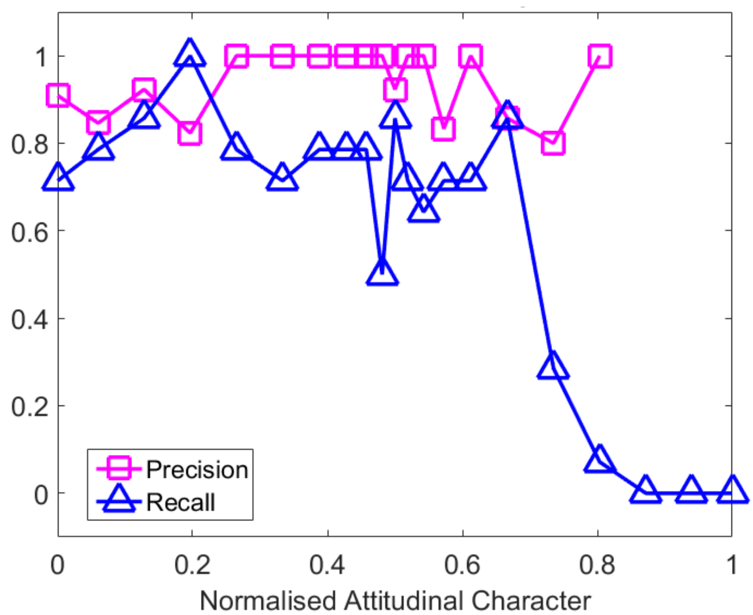


Figure 8: Precision and Recall of Negative Days

negative days are mis-clustered as positive days. If these clustering results were selected as the prototypes to implement the monitoring system, malfunctions occurring on those mis-clustered negative days could not be detected. In this case, if the same malfunctions happen again in future, the system would fail to detect them, which might cause serious consequences. On the other hand, when the A-C$'(W)$ is close to 2.0 (i.e., if a transitive aggregated fuzzy similarity is employed), the value of recall reaches 1.0 and the value of precision at this point is 0.824 (14/17 = 0.824). This shows that all the negative days in the dataset are correctly detected, although 3 positive days are misclassified as negative ones.

Based on the above observation and considering the potential serious environmental consequences that may be caused by any undetected malfunction, the weighting vector W_4 and its associated clustering result are chosen as the prototype to implement the working water treatment plant monitoring system. This is carried out in the implementation despite the fact that there are several weighing vectors which may lead to a seemingly best accuracy as reflected in Fig. 7. Note that according to the definition of attitudinal character, A-C$'(W_4) = 0.197$ can be explained as an andness (pessimistic) behaviour of the aggregation. In other words, when monitoring the water treatment plant (in order to detect any malfunction), the 38 measurements should be conjunctively rather than disjunctively considered, and a day should be classified as normal/abnormal only if the majority measurements of that day are normal/abnormal.

5.2. On-line Monitoring

From this discussion, the best weighting vector is selected through the analysis of the recall and precision of the negative days. Following this, the aggregated similarity can be calculated between a new instance and the resultant 13 clusters. The label of the new instance can therefore be assigned by the 1-NN rule, labelling it with the label of the cluster that is of the highest similarity. **The details of on-line monitoring is shown in Alg. 1.**

Algorithm 1 On-line Monitoring

$C^* = \{c_1, c_2, \cdots, c_{13}\}$, the best clustering result selected from off-line training;
W^*, the weighting vector with which C^* is generated;
r^{L}, the value of R^{L} aggregated similarity, where W^* is employed in the calculation;
Input: $i_o = (i_1, \cdots, i_{38})$, an observed instance which is consist of 38 indicator-values;

```
max_cl = 0;
for k = 1 : 13 do
    complete_linkage =1;
    for each instance i ∈ c_k do
        if r^L(i_0, i) < complete_linkage then
            complete_linkage = r^L(i_0, i)
        end if
    end for
    if complete_linkage > max_cl then
        max_cl = complete_linkage;
        c* = c_k;
    end if
end for
if label of c* is Negative then
    report a malfunction;
end if
```

Algorithm 1 calculates the values of complete-linkage between the new observation and those clusters in C^* one by one, and the cluster with the maximum value of complete-linkage (i.e. max_cl) to the new observation is located. If there are more than one cluster reach the maximum value of complete-linkage, the cluster with the smallest index i is located. Therefore,

in order to detect and report potential malfunction, the negative clusters should be ordered in front of those positive ones in C^* after the off-line training. Also, instead of using the complete-link and 1-NN rule, there are alternative ways of calculating the similarity between an instance-cluster pair, and of deciding on which cluster the instance ought to belong to [55, 56].

5.3. Discussion of Robustness

Robustness is key to engineering applications. For the proposed model of water treatment plant monitoring, the clustering of historical dataset is critical to the off-line training stage and hence, to the whole monitoring system. It is assumed that the historical dataset has collected sufficient amount of instances to represent the distribution of negative and positive clusters. However, in order to increase the robustness of the system, it is wise to redo the off-line training and rebuild the prototype of clusters after new data has been collected.

For a given dataset, different clustering algorithms may lead to different results. The hierarchical clustering is employed in the proposed system amongst many others due to its good performance and also due to the dendrogram is very helpful for engineers to understand the formation of the clusters. In fact, in order to entail comparison of the clustering quality, three other classical clustering methods: Euclidean distance based complete-link hierarchical clustering (HC for shorthand in the result presentation later), fuzzy c-means (FC), and k-means (KM) are employed as well (their clustering results are shown in Fig. 7). The MATLAB default parameter settings are adopted in these algorithms. The results also show that the accuracy achieved by the proposed T_{L}-transitive aggregation outperforms that obtained by Euclidean distance based methods. This demonstrates that if the weights for R^{L} aggregation are appropriately selected, the proposed method may result in very good performance.

Note that apart from analysing the recall and precision of the clustering result, there are many methods that may be employed to decide on the quality of clustering results. For example, the prototype and cluster consistency may be introduced to evaluate the performance of hierarchical clustering [54]. Nevertheless, no matter what techniques to use, interpretability is a very important factor in any application of prototype based methods, in order to assist engineers in detecting malfunctions. By introducing the OWA aggregation of fuzzy similarity relations, the proposed method can provide not only clustering prototypes of a high quality, but also can explain the reason of achieving good clusters through the andness/orness behaviour of the selected aggregator.

6. Conclusion

This paper has presented a novel notion of ordered weighted aggregation of fuzzy relations and its application for hierarchical clustering. Similar to the classic OWA aggregation, the behaviour of the proposed aggregation of fuzzy relations can be controlled by the use of stress functions. The work has also invested the conditions of when the aggregated similarity preserves T-transitivity. In addition to systematic experimental evaluation over conventional benchmark datasets, the realistic application investigation on monitoring malfunctions in an urban waste water treatment plant indicate that the aggregated transitive fuzzy relations generally outperform the non-transitive ones. **Meanwhile, by introducing the OWA aggregation of fuzzy similarity relations into hierarchical clustering, the system can not only provide clustering prototypes with the dendrogram, but also can explain the interaction of attributes in the form of attitudinal character, which is very helpful for engineers to understand the formation of the clustering results in order to assist them in detecting malfunctions.**

Whilst promising, the present work opens up an avenue for further investigation. For instance, the proposed method can be applied on other complex real-world problems such as Mars image clustering for space exploration [57], apart from the detection of water treatment malfunction. It is also useful to examine the performance of the proposed aggregations when more complicated stress functions rather than the linear ones are applied. It would be very interesting to investigate how to learn the weights of the proposed aggregation from datasets, while using stress functions as constraints to control the T_{L}-transitivity of the learned weighting vectors. Last but not least, it would be helpful to consider integrating this work with an effective feature selection (FS) mechanism (e.g., [58]) to effectively reduce the number of attributes required to be measured in order to perform the monitoring task more efficiently. In particular, previous research of applying FS to water treatment plant monitoring as reported in [52, 53] may offer a direct input to such investigations.

Acknowledgments

The authors are very grateful to the reviewers for their constructive comments on the original manuscript which have helped improve this work significantly. The first and third authors would like to thank Aberystwyth University for providing PhD scholarships in support of their research. Part of this work has been supported by the Fundamental Research Funds for the Central Universities, China, under grant number: 2016MS120.

Appendix A

Theorem 6. *Let $R_1, \cdots, R_m$ be T_L-transitive relations, $(w_1, \cdots, w_m)$ be the weighting vector in T_L, R^L is T_L-transitive iff when $w_i \leq w_j$ for $i < j$.*

Proof. It can be concluded from Theorem 3 that R^{L} is T_{L}-transitive $\iff$ the De Morgan's dual of $A(r_1(a,b), \cdots, r_m(a,b))$ satisfies that $\forall x, y, z \in [0,1]^m | x = y + z; N(x) \leq N(y) + N(z)$.

Let $x' = 1 - x$. Then, the De Morgan's dual of $A(x_1, \cdots, x_m)$: $N(x) = 1 - A((1-x_1), \cdots, (1-x_m)) = \sum_{j=1}^{m} w_j x'_{\pi'(j)}$ where $x'_{\pi'(j)}$ is a permutation of $x'_j \in [0,1], j = 1, \cdots, m$, which satisfies that $x'_{\pi'(j)}$ is the j-th largest of x'_j. Since $x' = 1 - x$, the descent permutation of x': $(1-x)_{\pi'(j)}$ can be replaced by an ascent permutation of x: $x_{\pi(j)}$ where $x_{\pi(i)} \leq x_{\pi(j)}$ for $i < j$, so that $(1-x)_{\pi'(j)} = 1 - x_{\pi(j)}$. Thus,

$$\begin{aligned} N(x) =& 1 - \sum_{j=1}^{m} w_j x'_{\pi'(j)} = 1 - \sum_{j=1}^{m} w_j (1-x)_{\pi'(j)} \\ =& 1 - \sum_{j=1}^{m} w_j (1 - x_{\pi(j)}) \\ =& 1 - \sum_{j=1}^{m} (w_j - w_j x_{\pi(j)}) \\ =& 1 - \sum_{j=1}^{m} w_j + \sum_{j=1}^{m} w_j x_{\pi(j)} \\ =& \sum_{j=1}^{m} w_j x_{\pi(j)}. \end{aligned}$$

Hence, the De Morgan's dual of $A(x)$: $N(x)$ can be seen as a 'reversed OWA aggregation' of $x \in [0,1]^m$, where $x_1, \cdots, x_m$ are increasingly ordered. According to [9], an OWA aggregation is a norm (which satisfies the triangle inequality $f(x)+f(y) \geq f(x+y)$) if and only if the OWA weighting vector satisfies the additional condition that $w_i \geq w_j$ for $i < j$. In the case of $N(x)$ with $w_i \leq w_j$ for $i < j$, its arguments are reversely ordered as they were in the original OWA operator, such that it equals to $A(x)$ with $w_i \geq w_j$ for $i < j$. Thus, $N(x)$ satisfies that $\forall x, y, z \in [0,1]^m, x = y + z \implies N(x) \leq N(y) + N(z)$. □

Appendix B

Theorem 7. *Let $R_1, \cdots, R_m$ be m component fuzzy relations, $(w_1, \cdots, w_m)$ be the weighting vector in $R^{\min} = [r^{min}(a,b)]$ such that $w_i \leq w_j$ for $i < j$, then $r^{\min}(a,b) = \min_{j=1,\cdots,m} r_j(a,b)$.*

Proof. From the definition of $R^{\min}$, it can be obtained that $r_{\pi(i)}(a,b) \geq r_{\pi(j)}(a,b)$ for $i < j$ and $r_{\pi(m)}(a,b) = \min\limits_{j=1,\cdots,m} r_j(a,b)$. Assume $w'_j = 1 - w_j$ for $j = 1, \cdots, m$, then $w'_i \geq w'_j$ for $i < j$. Since $r_{\pi(i)}(a,b) \geq r_{\pi(j)}(a,b)$ and $w'_i \geq w'_j$, for $i < j$, then $\max(w'_1, r_{\pi(1)}(a,b)) \geq \cdots \geq \max(w'_m, r_{\pi(m)}(a,b))$. Therefore,

$$
\begin{aligned}
r^{\min}(a,b) &= \min_{j=1,\cdots,m} \max(1 - w_j, r_{\pi(j)}(a,b)) \\
&= \max(w'_m, r_{\pi(m)}(a,b)) \\
&= \max(1 - w_m, r_{\pi(m)}(a,b))
\end{aligned}
$$

According to the definition of $R^{\min}$, $\max\limits_{j=1,\cdots,m} w_j = 1$ and since $w_i \leq w_j$ for $i < j$ then $w_m = 1$.

$$
\begin{aligned}
r^{\min}(a,b) &= \max(1 - w_m, r_{\pi(m)}(a,b)) \\
&= \max(0, r_{\pi(m)}(a,b)) \\
&= r_{\pi(m)}(a,b)) = \min_{j=1,\cdots,m} r_j(a,b)
\end{aligned}
$$

□

Note that the purpose of adding the constraint $\max\limits_{j=1,\cdots,m} w_j = 1$ on $R^{\min}$ is to make it satisfy the requirement of $\text{Agg}(0, \cdots, 0, w_1, \cdots, w_m) = 0$ in Definition 2. If the constraint is removed from the definition of $R^{\min}$, Theorem 4 still holds while the result of $r^{min}(a,b)$ will not equal to $\min\limits_{j=1,\cdots,m} r_j(a,b)$, but to $\max(1 - w_m, r_{\pi(m)}(a,b))$.

Appendix C

Table 7: Searching Space of Weighting Vectors

ID	A-C′(W)	Weighting Vector
W_1	0.0000	(.0000 1.0000)
W_2	0.0600	(.0000 .0183 .0488 .0793 .1098 .1402 .1707 .2012 .2317)
W_3	0.1281	(.0000 .0016 .0097 .0179 .0260 .0341 .0422 .0503 .0584 .0666 .0747 .0828 .0909 .0990 .1071 .1153 .1234)
W_4	**0.1965**	(.0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0000 .0029 .0066 .0103 .0140 .0177 .0214 .0251 .0287 .0324 .0361 .0398 .0435 .0472 .0508 .0545 .0582 .0619 .0656 .0693 .0730 .0766 .0803 .0840)
W_5	0.2650	(.0000 .0000 .0000 .0000 .0000 .0000 .0000 .0008 .0029 .0050 .0071 .0092 .0113 .0134 .0155 .0176 .0197 .0218 .0239 .0260 .0281 .0302 .0323 .0344 .0364 .0385 .0406 .0427 .0448 .0469 .0490 .0511 .0532 .0553 .0574 .0595 .0616 .0637)
W_6	0.3333	(.0013 .0027 .0040 .0054 .0067 .0081 .0094 .0108 .0121 .0135 .0148 .0162 .0175 .0189 .0202 .0216 .0229 .0243 .0256 .0270 .0283 .0297 .0310 .0324 .0337 .0351 .0364 .0378 .0391 .0405 .0418 .0432 .0445 .0459 .0472 .0486 .0499 .0513)
W_7	0.3879	(.0095 .0104 .0113 .0123 .0132 .0141 .0150 .0159 .0168 .0177 .0186 .0195 .0204 .0213 .0222 .0231 .0240 .0250 .0259 .0268 .0277 .0286 .0295 .0304 .0313 .0322 .0331 .0340 .0349 .0358 .0368 .0377 .0386 .0395 .0404 .0413 .0422 .0431)
W_8	0.4275	(.0155 .0160 .0166 .0172 .0178 .0184 .0190 .0196 .0202 .0207 .0213 .0219 .0225 .0231 .0237 .0243 .0248 .0254 .0260 .0266 .0272 .0278 .0284 .0290 .0295 .0301 .0307 .0313 .0319 .0325 .0331 .0337 .0342 .0348 .0354 .0360 .0366 .0372)
W_9	0.4575	(.0200 .0203 .0206 .0210 .0213 .0217 .0220 .0224 .0227 .0230 .0234 .0237 .0241 .0244 .0248 .0251 .0255 .0258 .0261 .0265 .0268 .0272 .0275 .0279 .0282 .0286 .0289 .0292 .0296 .0299 .0303 .0306 .0310 .0313 .0316 .0320 .0323 .0327)
W_{10}	0.4810	(.0235 .0236 .0238 .0239 .0241 .0242 .0244 .0246 .0247 .0249 .0250 .0252 .0253 .0255 .0256 .0258 .0259 .0261 .0262 .0264 .0265 .0267 .0269 .0270 .0272 .0273 .0275 .0276 .0278 .0279 .0281 .0282 .0284 .0285 .0287 .0288 .0290 .0292)
W_{11}	0.5000	(.0263)

References

[1] T. Calvo, G. Mayor, R. Mesiar, Aggregation operators: new trends and applications, Vol. 97, Physica, 2012.

[2] G. Beliakov, A. Pradera, T. Calvo, Aggregation functions: A guide for practitioners, Springer Publishing Company, Incorporated, 2008.

[3] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators: properties, classes and construction methods, in: Aggregation operators, Springer, 2002, pp. 3–104.

[4] R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, Systems, Man and Cybernetics, IEEE Transactions on 18 (1) (1988) 183–190.

[5] H. Chen, L. Zhou, An approach to group decision making with interval fuzzy preference relations based on induced generalized continuous ordered weighted averaging operator, Expert Systems with Applications 38 (10) (2011) 13432–13440.

[6] J. M. Merigó, M. Casanovas, Induced aggregation operators in the euclidean distance and its application in financial decision making, Expert Systems with Applications 38 (6) (2011) 7603–7608.

[7] M. Suo, Y. Li, G. H. Huang, Multicriteria decision making under uncertainty: An advanced ordered weighted averaging operator for planning electric power systems, Engineering Applications of Artificial Intelligence 25 (1) (2012) 72–81.

[8] P. Su, C. Shang, T. Chen, Q. Shen, Exploiting data reliability and fuzzy clustering for journal ranking, IEEE Transactions on Fuzzy Systems PP (99) (2016) 1–1. doi:10.1109/TFUZZ.2016.2612265.

[9] R. Yager, Norms induced from owa operators, Fuzzy Systems, IEEE Transactions on 18 (1) (2010) 57–66. doi:10.1109/TFUZZ.2009.2035812.

[10] D. Li, W. Zeng, J. Li, New distance and similarity measures on hesitant fuzzy sets and their applications in multiple criteria decision making, Engineering Applications of Artificial Intelligence 40 (2015) 11–16.

[11] J. Williams, N. Steele, Difference, distance and similarity as a basis for fuzzy decision support based on prototypical decision classes, Fuzzy Sets and Systems 131 (1) (2002) 35–46.

[12] P. Perner, Are case-based reasoning and dissimilarity-based classification two sides of the same coin?, Engineering Applications of Artificial Intelligence 15 (2) (2002) 193–203.

[13] Y. A. Tolias, S. M. Panas, L. H. Tsoukalas, Generalized fuzzy indices for similarity matching, Fuzzy Sets and Systems 120 (2) (2001) 255–270.

[14] J. Fernández Salido, S. Murakami, Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations, Fuzzy Sets and Systems 139 (3) (2003) 635–660.

[15] L. Lifen, Trust derivation and transitivity in a recommendation trust model, in: Computer Science and Software Engineering, 2008 International Conference on, Vol. 3, IEEE, 2008, pp. 770–773.

[16] T. Wittkop, D. Emig, S. Lange, S. Rahmann, M. Albrecht, J. H. Morris, S. Böcker, J. Stoye, J. Baumbach, Partitioning biological data with transitivity clustering, Nature methods 7 (6) (2010) 419–420.

[17] L. Zadeh, Similarity relations and fuzzy orderings, Information Sciences 3 (2) (1971) 177–200.

[18] M. M. Savino, A. S. Sekhari, A quality management system based on fuzzy quality pointers in iso 9000, International Journal of Product Development 8 (4) (2009) 419–430.

[19] P. Su, C. Shang, Q. Shen, Owa aggregation of fuzzy similarity relations for journal ranking, in: Fuzzy Systems (FUZZ), 2013 IEEE International Conference on, 2013, pp. 1–7. doi:10.1109/FUZZ-IEEE.2013.6622376.

[20] M. M. Savino, A. Brun, C. Xiang, A fuzzy-based multi-stage quality control under the iso 9001: 2015 requirements, European Journal of Industrial Engineering 11 (1) (2017) 78–100.

[21] B. Schweizer, A. Sklar, Probabilistic metric spaces, Courier Dover Publications, 2011.

[22] M. M. Savino, S. Apolloni, Environmental plant optimization in small sized enterprises through an operative framework, International Journal of Operations and Quantitative Management 13 (2) (2007) 95.

[23] S.-M. Liou, S.-L. Lo, C.-Y. Hu, Application of two-stage fuzzy set theory to river quality evaluation in taiwan, Water Research 37 (6) (2003) 1406–1416.

[24] P. Fonck, J. Fodor, M. Roubens, An application of aggregation procedures to the definition of measures of similarity between fuzzy sets, Fuzzy Sets and Systems 97 (1) (1998) 67–74.

[25] T. Sudkamp, Similarity, interpolation, and fuzzy rule construction, Fuzzy Sets and Systems 58 (1) (1993) 73–86.

[26] G. Beliakov, Definition of general aggregation operators through similarity relations, Fuzzy Sets and Systems 114 (3) (2000) 437–453.

[27] G. Beliakov, S. James, G. Li, Learning choquet-integral-based metrics for semisupervised clustering, Fuzzy Systems, IEEE Transactions on 19 (3) (2011) 562–574. doi:10.1109/TFUZZ.2011.2123899.

[28] Z. Xu, J. Chen, Ordered weighted distance measure, Journal of Systems Science and Systems Engineering 17 (4) (2008) 432–445.

[29] P. Baraldi, F. D. Maio, D. Genini, E. Zio, Reconstruction of missing data in multidimensional time series by fuzzy similarity, Applied Soft Computing 26 (2015) 1–9.

[30] S. Pramanik, K. Mondal, Weighted fuzzy similarity measure based on tangent function and its application to medical diagnosis, International Journal of Smart & Nano Materials 4 (1) (2015) 158–164.

[31] M. Verma, J. Rajasankar, N. Anandavalli, A. Prakash, N. R. Iyer, Fuzzy similarity approach for ranking and health assessment of towers based on visual inspection, Advances in Structural Engineering 18 (9) (2015) 1399–1414.

[32] R. Yager, Using stress functions to obtain owa operators, Fuzzy Systems, IEEE Transactions on 15 (6) (2007) 1122–1129.

[33] X. Liu, S. Yu, On the stress function-based owa determination method with optimization criteria, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (1) (2012) 246–257.

[34] J. J. Dujmović, Properties of local andness/orness, in: Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, Springer, 2007, pp. 54–63.
[35] A. K. Jain, M. N. Murty, P. J. Flynn, Data clustering: a review, ACM Computing Surveys (CSUR) 31 (3) (1999) 264–323.
[36] K. Bache, M. Lichman, UCI machine learning repository (2013).
URL http://archive.ics.uci.edu/ml
[37] A. K. Jain, M. H. Law, Data clustering: A users dilemma, in: Pattern Recognition and Machine Intelligence, Springer, 2005, pp. 1–10.
[38] A. Gionis, H. Mannila, P. Tsaparas, Clustering aggregation, ACM Transactions on Knowledge Discovery from Data (TKDD) 1 (1) (2007) 4.
[39] C. T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters, Computers, IEEE Transactions on 100 (1) (1971) 68–86.
[40] L. Fu, E. Medico, Flame, a novel fuzzy clustering method for the analysis of dna microarray data, BMC bioinformatics 8 (1) (2007) 3.
[41] H. Chang, D.-Y. Yeung, Robust path-based spectral clustering, Pattern Recognition 41 (1) (2008) 191–203.
[42] P. Horton, K. Nakai, A probabilistic classification system for predicting the cellular localization sites of proteins., in: Proceedings of the 4th International Conference on Intelligent Systems for Molecular Biology, Vol. 4, 1996, pp. 109–115.
[43] I. W. Evett, E. J. Spiehler, Rule induction in forensic science, Tech. rep., Central Research Establishment, Home Office Forensic Science Service (1987).
[44] V. G. Sigillito, S. P. Wing, L. V. Hutton, K. B. Baker, Classification of radar returns from the ionosphere using neural networks, Johns Hopkins APL Tech. Dig 10 (1989) 262–266.
[45] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (2) (1936) 179–188.
[46] D. Coomans, I. Broeckaert, M. Jonckheer, D. L. Massart, Comparison of multivariate discrimination techniques for clinical data–application to the thyroid functional state., Methods Archive 22 (2) (1983) 93–101.
[47] B. Vandeginste, Parvus: An extendable package of programs for data exploration, classification and correlation, Journal of Chemometrics 4 (2) (1990) 191–193. doi:10.1002/cem.1180040210.
[48] M. M. Savino, A. Mazza, G. Neubert, Agent-based flow-shop modelling in dynamic environment, Production Planning & Control 25 (2) (2014) 110–122.
[49] M. M. Savino, A. Mazza, Toward environmental and quality sustainability: An integrated approach for continuous improvement, IEEE Transactions on Engineering Management 61 (1) (2014) 171–181.
[50] M. M. Savino, M. Macchi, A. Mazza, Investigating the impact of social sustainability within maintenance operations, Journal of Quality in Maintenance Engineering 21 (3).
[51] L. Belanche, M. Sànchez, U. Cortés, P. Serra, A knowledge-based system for the diagnosis of waste-water treatment plants, Springer Berlin Heidelberg, Berlin, Heidelberg, 1992, pp. 324–336. doi:10.1007/BFb0024984.
URL http://dx.doi.org/10.1007/BFb0024984
[52] Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, Pattern Recognition 35 (11) (2002) 2425–2438.
[53] Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition 37 (7) (2004) 1351–1363.
[54] R. A. Mollineda, F. J. Ferri, E. Vidal, An efficient prototype merging strategy for the condensed 1-nn rule through class-conditional hierarchical clustering, Pattern Recognition 35 (12) (2002) 2771–2782.
[55] R. R. Yager, D. P. Filev, Generation of fuzzy rules by mountain clustering, Journal of Intelligent & Fuzzy Systems Applications in Engineering & Technology 2 (3) (1992) 209–219.
[56] S. L. Chiu, Fuzzy model identification based on cluster estimation, Journal of Intelligent & Fuzzy Systems Applications in Engineering & Technology 2 (3) (1994) 267–278.
[57] C. Shang, D. Barnes, Fuzzy-rough feature selection aided support vector machines for mars image classification, Computer Vision & Image Understanding 117 (3) (2013) 202–213.
[58] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: Rough and fuzzy-rough approaches, IEEE Transactions on Knowledge and Data Engineering, 16 (12) (2004) 1457–1471.