Nearest-Neighbor Guided Evaluation of Data Reliability and Its Applications

Boongoen, Tossapon; Shen, Qiang

Published in:
IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)

DOI:
10.1109/TSMCB.2010.2043357

Publication date:
2010

Citation for published version (APA):
Boongoen, T., & Shen, Q. (2010). Nearest-Neighbor Guided Evaluation of Data Reliability and Its Applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *40*(6), 1622-1633. https://doi.org/10.1109/TSMCB.2010.2043357

Nearest-Neighbor Guided Evaluation of Data Reliability and Its Applications

Tossapon Boongoen and Qiang Shen

***Abstract*—The intuition of data reliability has recently been incorporated into the main stream of research on ordered weighted averaging (OWA) operators. Instead of relying on human-guided variables, the aggregation behavior is determined in accordance with the underlying characteristics of the data being aggregated. Data-oriented operators such as the dependent OWA (DOWA) utilize centralized data structures to generate reliable weights, however. Despite their simplicity, the approach taken by these operators neglects entirely any local data structure that represents a strong agreement or consensus. To address this issue, the cluster-based OWA (Clus-DOWA) operator has been proposed. It employs a cluster-based reliability measure that is effective to differentiate the accountability of different input arguments. Yet, its actual application is constrained by the high computational requirement. This paper presents a more efficient nearest-neighbor-based reliability assessment for which an expensive clustering process is not required. The proposed measure can be perceived as a stress function, from which the OWA weights and associated decision-support explanations can be generated. To illustrate the potential of this measure, it is applied to both the problem of information aggregation for alias detection and the problem of unsupervised feature selection (in which unreliable features are excluded from an actual learning process). Experimental results demonstrate that these techniques usually outperform their conventional state-of-the-art counterparts.**

***Index Terms*—Alias detection, data reliability, nearest neighbor, ordered weighted averaging (OWA) aggregation, unsupervised feature selection, weight determination.**

I. Introduction

MANY important aggregation operators have been developed to deliver a reasonable outcome upon which an intelligent decision can be made. These operators range from the simple arithmetic mean to fuzzy-oriented ones, including minimum/maximum, uninorm, and many types of more complex t-norm/t-conorm (further details in [3]). Furthermore, a parameterized mean-like aggregation operator, i.e., ordered weighted averaging (OWA), has been introduced [60] and successfully applied in different areas [65]. Essentially, by selecting an appropriate weight vector, an OWA operator can reflect the uncertain nature of human judgment, with the ability to generate an aggregating result that lies between the two extremes of minimum and maximum. A number of different techniques have been proposed to obtain weights that are appropriate for different operators: maximum entropy [44], weight learning [16], recursive formulation [57], Gaussian distribution [58], and data clustering methods [6] (see [20] for more details).

In the process of combining multiple arguments, a precaution worth noting is that unduly high/low or abnormal values may be given by false or biased judgment. In such cases, a typical OWA operator would suffer drastically from assigning the highest priority to either the highest or the lowest value. As a result, the intuition of data *reliability* has recently been incorporated into the research on OWA operators. Unlike many other conventional weight determination techniques that concentrate on human-guided variables, the reliability-oriented approach models the aggregation behavior in accordance with the characteristics of the data being aggregated. The original technique introduced by following this approach is the dependent OWA (DOWA) operator [58], [59], where a normal distribution of argument values is assumed to determine their reliability degrees and, hence, the weights. In particular, a high weight (i.e., good reliability) is given to the argument whose value is close to the center of all arguments (i.e., mean), whereas lower weights are assigned to those further away. This interpretation has also been generalized in the centered OWA operator [63], where weights are high around the middle and decay symmetrically toward the boundary ends.

Despite their generality, these weight generation methods possess a common drawback, which originates from the underlying centralized assumption. Conceptually, argument values are viewed as members of one large cluster (i.e., a global consensus of decision makers' opinions), and the arithmetic mean is considered sufficient to grade their reliability. This approach completely discards the significance of any possible trend that emerges from a local data structure as a subset of values that are tightly clustered together. To avoid this problem, a cluster-based distance metric has been introduced in [6] to measure the reliability from which a so-called Clus-DOWA operator and its weighting scheme can be formulated. Effectively, those values that are very far from the group center (i.e., mean) are not necessarily unreliable if they are seemingly indifferent to their local neighbors (i.e., its distance to the nearest cluster is small). In spite of reported effectiveness, such an agglomerative hierarchical clustering technique [14] has significant drawbacks: high time and space complexity of $O(n^3)$ and $O(n^2)$, respectively (where n is the number of input arguments).

To overcome the aforementioned burden, this paper presents an improved scheme of the existing cluster-based reliability assessment, where the distance to the nearest neighbor is

Manuscript received July 24, 2009; revised November 19, 2009; accepted January 18, 2010. Date of publication April 26, 2010; date of current version November 17, 2010. This work was supported by the U.K. Engineering and Physical Sciences Research Council Grant EP/D057086. This paper was recommended by Associate Editor J. Basak.

The authors are with the Department of Computer Science, Aberystwyth University, SY23 3DB Aberystwyth, U.K. (e-mail: tsb@aber.ac.uk; qqs@aber.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2010.2043357

employed rather than that of the closest cluster. Hence, the data clustering process becomes irrelevant, and the resulting time and space complexity is reduced to $O(n^2)$ and $O(n)$, respectively. To demonstrate the effectiveness of this simplified reliability measure, it is applied to two different information processing tasks: 1) determination of the weights of OWA aggregation for alias detection and 2) unsupervised feature selection. The intuition of nearest neighbors is not new. However, its application within the context of *information aggregation* and *unsupervised feature selection* is unique.

The rest of this paper is organized as follows. Section II introduces the main theoretical concepts of the OWA operators, with emphasis on the reliability-based weight determination approach from which the current research is motivated and developed. In Section III, the nearest-neighbor guided reliability evaluation is thoroughly explained, including its advantages over the existing cluster-based method. Section IV describes the exploitation of this data-driven reliability measure as a stress function by which a user can perceive the importance degrees of different arguments and their corresponding contributions toward the aggregated outcome. The resulting DOWA operator is first applied for the task of alias detection in intelligence data, aggregating similarity measures generated by distinct string-matching algorithms. To reflect the generality of this reliability measure, it is further applied to the problem of unsupervised feature selection, details of which are given in Section V. This paper is concluded in Section VI, with a short discussion of future work.

II. Preliminaries

Here, the theoretical basis and common practice regarding the OWA operator and weight determination methods are presented, upon which the current research is established.

A. OWA Operator

The process of information aggregation appears in many decision-support applications. Despite being computationally simplistic, neither minimum nor maximum is appropriate for most of such applications. Accordingly, a new family of aggregation methods termed the OWA operator has been developed [60]. This type of a mean-like operator provides a flexible way to utilize the entire range of operators from the logical conjunction to the logical disjunction (with the two extremes traditionally implemented by minimum and maximum, respectively).

Definition 1: An OWA operator of dimension n is a mapping $R^n \rightarrow R$, which has an associated weighting vector $W = (w_1, w_2, \ldots, w_n)^T$, where $w_j \in [0,1]$ and $\sum_{j=1}^{n} w_j = 1$. An input vector $(a_1, a_2, \ldots, a_n)$, is aggregated as follows:

$$\mathrm{OWA}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \qquad (1)$$

where b_j is the jth largest element in the vector $(a_1, a_2, \ldots, a_n)$ and $b_1 \geq b_2 \geq \cdots \geq b_n$. Prior to the application of weights, the *reordering* process of arguments $(a_1, a_2, \ldots, a_n)$ to $(b_1, b_2, \ldots, b_n)$ is essential. Intuitively, an OWA operator is *order-dependent* since weights are assigned in accordance with the order of argument values (see details of OWA properties in [38] and [65]).

Weight determination is crucial to this family of operators since associated weights dictate the type of aggregation that an OWA exhibits. A number of different techniques have been proposed for obtaining weights used by the OWA operators, for instance, maximal entropy [44], weight learning [16], Gaussian [58], and data clustering methods [6] (more details in [20]). Another important and useful method for weight determination is the functional approach called basic unit-interval monotonic (BUM) function, as introduced in [62].

Definition 2: Let F be a function $F : [0,1] \rightarrow [0,1]$ such that $F(0) = 0$, $F(1) = 1$, and $F(a) \geq F(b)$ given $a \geq b$. With this BUM function, it is possible to derive a weight vector $(w_1, w_2, \ldots, w_n)^T$ as follows:

$$w_i = F\left(\frac{i}{n}\right) - F\left(\frac{i-1}{n}\right), \qquad i = 1, \ldots, n. \qquad (2)$$

For instance, with a BUM function $F(x) = x, \forall x \in [0,1]$, the resulting weight vector is $w_i = (1/n), i = 1, \ldots, n$, which equivalently leads to an averaging weight (see further details in [62]). This approach helps increase the usefulness of OWA operators. In particular, it enables the modeling of linguistically specified aggregation imperatives and the inclusion of importance associated with aggregated arguments [65]. Following this, a simple weight generation mechanism has recently been introduced with stress functions, by which a user can conceptually specify the type of an OWA operator required for a given application problem [64].

Definition 3: A *stress* function is a nonnegative function $s(x)$ defined on the unit interval $x : [0,1] \rightarrow R^+$. Given this, $F(x)$ can be defined as follows, where $\int_0^1 s(y)dy = K$:

$$F(x) = \frac{1}{K} \int_0^x s(y)dy. \qquad (3)$$

According to (2), OWA weights can be derived as

$$w_i = \frac{1}{K} \left(\int_0^{\frac{i}{n}} s(y)dy - \int_0^{\frac{i-1}{n}} s(y)dy \right), \qquad i = 1, \ldots, n. \quad (4)$$

This calculation can be simplified if weights are approximated directly from a stress function s as follows (see proof and further details in [64]):

$$w_i = \frac{s\left(\frac{i}{n}\right)}{\sum_{j=1}^{n} s\left(\frac{j}{n}\right)}. \qquad (5)$$

With this method, a user can easily characterize the nature of aggregation through locations of stress (i.e., significant values).

B. Data Reliability and DOWA Operators

Weight vectors generated by the aforementioned functions are classified as *argument-independent* since they are not related to the aggregates being studied. In contrast, with the *argument-dependent* approach, weights are determined based on the properties of input arguments. Specifically, the DOWA operator in [58] and [59] utilizes a weight vector derived in accordance with the normal distribution of argument values. In essence, with this centralized perspective, arguments whose values are in the middle of the group, i.e., near the group average, are considered more reliable and acquire higher weights when compared with those further away from the center. Note that the *reliability* of an argument can be conceptually defined as the appropriateness of using the argument as the group representative (i.e., the aggregated outcome).

Similarly, the clustered argument DOWA (Clus-DOWA) operator, recently introduced in [6], aims to decrease the effect of false or biased judgment in a group decision making. Here, the intuition of *reliability* is also engaged to differentiate a collection of arguments. An argument whose value is similar to those of others is considered reliable and can be regarded as the group representative. In contrast, an argument that is largely different from the rest is discriminated as the unreliable member.

At the outset, to obtain the cluster-based weight vector for a set of arguments $\{a_1, a_2, \ldots, a_n\}$, the agglomerative hierarchical clustering technique [14] is adopted such that the cluster structure is achieved through iterations of merging the nearest pair of clusters into a larger one. Particularly, the clustering process terminates as soon as all arguments have been merged to their nearest clusters. Following that, for each argument a_i, the distance d_i behind such merging is recorded for the evaluation of its reliability.

Definition 4: For each argument $a_i, i = 1, \ldots, n$, its reliability r_i can be directly estimated from the distance to its nearest cluster d_i recorded during the clustering process, i.e.,

$$r_i = 1 - \frac{d_i}{\sum_{j=1}^{n} d_j}. \tag{6}$$

From this, the weight vector can then be calculated from the vector of reliability measure $(r_1, r_2, \ldots, r_n)$ as follows:

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j}, \qquad i = 1, 2, \ldots, n. \tag{7}$$

Data-dependent weight vectors, generated by this distributed methodology, have proven effective for classification and feature selection problems, with the performance superior to that of the centralized counterpart.

III. Nearest-Neighbor Guided Evaluation of Data Reliability

In spite of reported success, the major drawback of the cluster-based reliability measure is high computational requirement: with the time and space complexity being $O(n^3)$ and $O(n^2)$, respectively (where n is the number of input data). This resource-demanding scenario is caused by the application of the agglomerative hierarchical clustering technique to discovering the reliability of each data argument. To overcome this fundamental drawback and maintain the advantage of the distributed approach, the local neighboring context that has previously been realized as a closest cluster is replaced by a set of k nearest neighbors ($k \in \{1, \ldots, n-1\}$). Fig. 1 depicts this modified approach, in which arguments (a_1 and a_2) very far from the global center are considered reliable if they are close to members of their local neighbor sets (N_{a_1} and N_{a_2}, respectively).

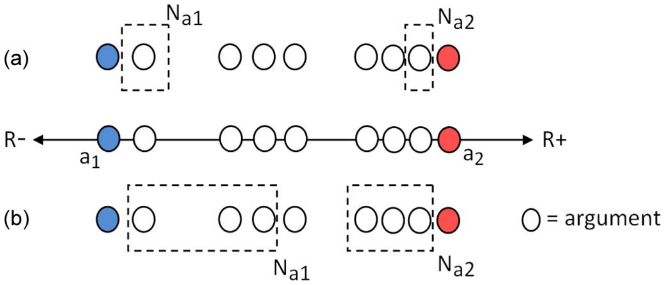

Fig. 1. Different local neighboring sets N_{a_1} and N_{a_2} of arguments a_1 and a_2, respectively, where (a) $k = 1$ and (b) $k = 3$.

```
FindNearestNeighbor($a_i, k, A$)
$A$, the set of arguments, $a_i \in A, i = 1 \ldots n$;
$k$, the number of nearest neighbors, $1 \le k < n$;
$N^k_{a_i}$, the set of $k$ nearest neighbors of $a_i$, $N^k_{a_i} \subset A$;
$|N^k_{a_i}|$, size of the neighbor set, $0 \le |N^k_{a_i}| \le k$;
$d(a_i, a_j)$, the distance between arguments $a_i, a_j \in A$;
$maxN$, the neighbor $n_p \in N^k_{a_i}, d(a_i, n_p) = \max_{\forall n_t \in N^k_{a_i}} d(a_i, n_t)$;

$N^k_{a_i} \leftarrow \emptyset$
for each $a_j \in A$
  if $a_j \neq a_i$
    if $|N^k_{a_i}| < k$
      $N^k_{a_i} \leftarrow N^k_{a_i} \cup a_j$
    else if $d(a_i, a_j) < d(a_i, maxN)$
      $N^k_{a_i} \leftarrow (N^k_{a_i} - maxN) \cup a_j$
return $N^k_{a_i}$
```

Fig. 2. Procedural description of the *FindNearestNeighbor* algorithm.

For a collection of data arguments $A = \{a_1, \ldots, a_n\}$, let $N^k_{a_i}$ be a set of k nearest neighbors of an argument a_i, where $N^k_{a_i} \subset A$, $n_j \in N^k_{a_i}$, $n_j \neq a_i$, $j = 1, \ldots, k$. The reliability of a specific argument can be determined by the distance to members of its nearest neighbor set that can be found using the *FindNearestNeighbor* algorithm given in Fig. 2. The higher this distance is, the less reliable that argument becomes.

Initially, the distance $d(a_i, a_j)$ between any two arguments $a_i, a_j \in A$ is specified simply as

$$d(a_i, a_j) = |a_i - a_j|. \tag{8}$$

This is for computational simplicity. Of course, any other distance metric may be applied if they do not incur too much overheads in computation. Given the distance metric, the reliability $R^k_{a_i}$ of argument a_i depends on the average distance $D^k_{a_i}$ to its k nearest neighbors (i.e., members of $N^k_{a_i}$), which is identified as

$$D^k_{a_i} = \frac{1}{K} \sum_{\forall n_t \in N^k_{a_i}} d(a_i, n_t) = d\left(a_i, \frac{\sum_{\forall n_t \in N^k_{a_i}} n_t}{k}\right). \quad (9)$$

Following this, the reliability measure $R^k_{a_i} \in [0, 1]$, $i = 1, \ldots, n$ can be obtained such that

$$R^k_{a_i} = 1 - \frac{D^k_{a_i}}{D_{\max}} \quad (10)$$

where $D_{\max} = \max_{a_p, a_q \in A, a_p \neq a_q} d(a_p, a_q)$.

Without the data clustering process, this reliability measure is more efficient compared to the existing cluster-based method, with time and space complexity generally decreasing to $O(n^2)$ and $O(n)$, respectively. Note that, in the extreme case of $k = n - 1$, the time complexity becomes a linear function of $O(n)$ as well since the search for nearest neighbors is not required. Essentially, this data-driven measure can be used to determine the weight vector of a DOWA operator (see Section IV), whose efficiency is substantially better than that of the Clus-DOWA counterpart. To further demonstrate the effectiveness of this nearest-neighbor-based reliability measure, it is applied to the task of unsupervised feature selection (see Section V), where the inclusion of a feature in a learning process depends on the overall reliability of its values.

Note that in this research, the underlying reliability measure is parameterized by a user-defined k. Intuitively, k should be small ($k \ll n$) to preserve the locality of data constitution. Empirical results have shown that given a small k, the proposed metric is robust to the setting of this parameter (examples for such empirical investigation are provided later). Alternatively, if representative historical data about the problem domain are available, this parameter may then be acquired using a learning methodology.

IV. Use of Reliability as a Stress Function With Application to Alias Detection

A common pitfall with existing aggregation operators is the inability to provide an explanatory means by which a user can utilize to enhance individual perception of arguments' importance. To resolve this shortcoming, a stress function [64] has recently been introduced as a simple mechanism for attaining interpretability. Accordingly, different types of a stress function can be used to express a weight distribution and, hence, different aggregation behavior. Similar to other argument-independent methods, this approach is practical for the circumstances where human experience is relevant. However, for a reliability-oriented case such as the task of combining several string-matching measures for alias detection, an argument-dependent weight determination technique proves to be particularly effective. As such, here, a novel stress-function-like method to obtain dependent weights and explanatory aggregation is presented, with the application to intelligence data analysis.

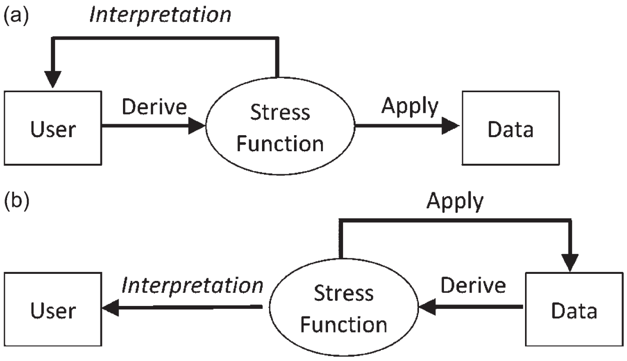

Fig. 3. Stress-function formalism: (a) conventional and (b) its reverse-engineered data-driven methods.

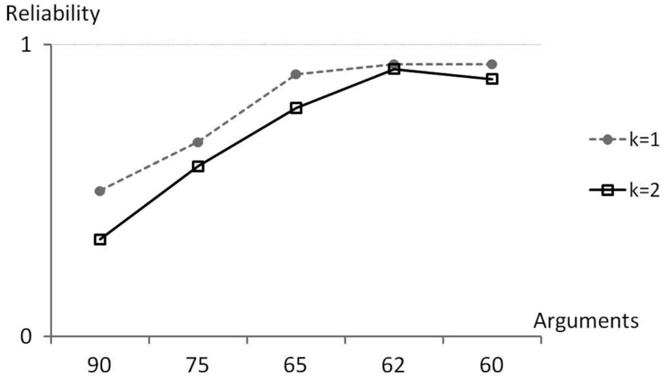

Fig. 4. Example of data-driven stress functions.

A. Explanatory OWA Aggregation With a Stress Function

The nearest-neighbor-based reliability measure, emphasized in Section III, can be regarded as a stress function that describes the significance of each input argument. This perspective is different from but complements the human-directed formalism originally introduced in [64]. As shown in Fig. 3, it can be perceived as the reverse-engineered counterpart of the conventional method. In essence, stress functions are similarly articulated for better interpretation, but they are derived from two distinct sources: human experience and intuition, and the data, respectively.

With the data-driven approach, prior to the actual aggregation process, a set of argument-specific reliability measures $R^k_{a_i}$ for input arguments $A = \{a_1, \ldots, a_n\}$ is generated using the k nearest-neighbor-based method, as formally illustrated in (8)–(10). Having achieved this, the graphical representation of these reliability values (analogous to stress values) can be used to broaden the interpretation of arguments' importance and the underlying data structure. For instance, Fig. 4 presents the resulting stress-like functions obtained from the application of the aforementioned method to arguments $A = \{90, 75, 65, 62, 60\}$, where $k \in \{1, 2\}$.

In addition to the graphical means, it is possible to achieve an enhanced and coherent understanding through a linguistic explanation of reliability measures. This goal is accomplished by

TABLE I
EXAMPLE OF LINGUISTIC EXPLANATION OF RELIABILITY MEASURES, WITH A MEMBERSHIP VALUE OF EACH LINGUISTIC LABEL BRACKETED. NOTE THAT $N^k_{a_i}$ AND $R^k_{a_i}$ DENOTE A k NEAREST-NEIGHBOR SET AND CORRESPONDING RELIABILITY MEASURE OF ARGUMENT a_i

k	a_i	$N^k_{a_i}$	$R^k_{a_i}$	Linguistic Description
1	90	{75}	0.50	Medium (1.0)
	75	{65}	0.67	Medium (0.66), High (0.34)
	65	{62}	0.90	Medium (0.2), High (0.8)
	62	{60}	0.93	Medium (0.14), High (0.86)
	60	{62}	0.93	Medium (0.14), High (0.86)
2	90	{75, 65}	0.33	Low (0.34), Medium (0.66)
	75	{65, 62}	0.58	Medium (0.83), High (0.17)
	65	{62, 60}	0.78	Medium (0.43), High (0.57)
	62	{65, 60}	0.92	Medium (0.17), High (0.83)
	60	{65, 62}	0.88	Medium (0.23), High (0.77)

exploiting descriptive labels with quantitative semantics represented by membership functions [66]. Let L be the set of labels ($l_j, j = 1, \ldots, n^r$, with n^r denoting the number of labels specified for degree of reliability) and S be the set of corresponding fuzzy sets ($s_j, j = 1, \ldots, n^r$) defined over the universe of discourse, $U^r = [0, 1]$. Note that a fuzzy set s_j is herein formally specified as $s_j = \{(x, \mu_{s_j}(x)) | x \in U^r, \mu_{s_j}(x) \in [0, 1]\}$, where $\mu_{s_j}(x) \in [0, 1]$ is the membership function of s_j. In this paper, for computational efficiency, each fuzzy set $s_j, \forall j = 1, \ldots, n^r$, is represented with a triangular membership function that is generally defined as follows:

$$\mu_{s_j}(x) = \begin{cases} 0, & x < x_1 \\ \frac{x - x_1}{x_2 - x_1}, & x_1 \le x \le x_2 \\ \frac{x_3 - x}{x_3 - x_2}, & x_2 \le x \le x_3 \\ 0, & x > x_3 \end{cases} \tag{11}$$

where x_1 and x_3 are the left and right bounds, respectively, x_2 is the mode of the fuzzy set s_j (i.e., $\mu_{s_j}(x_2) = 1$), and $x, x_1, x_2, x_3 \in U^r$.

To be concise, suppose that the label set $L = \{l_1 = \text{Low}, l_2 = \text{Medium}, l_3 = \text{High}\}$, with $n^r = 3$. Thus, the following three membership functions can be defined to represent the quantitative semantics of linguistic labels:

$$\mu_{s_1}(x) = \begin{cases} \frac{0.5 - x}{0.5}, & 0 \le x \le 0.5 \\ 0, & x > 0.5 \end{cases} \tag{12}$$

$$\mu_{s_2}(x) = \begin{cases} \frac{x}{0.5}, & 0 \le x \le 0.5 \\ \frac{1 - x}{0.5}, & 0.5 < x \le 1 \end{cases} \tag{13}$$

$$\mu_{s_3}(x) = \begin{cases} 0, & x < 0.5 \\ \frac{x - 0.5}{0.5}, & 0.5 \le x \le 1 \end{cases} \tag{14}$$

where $S = \{s_1, s_2, s_3\}$ and $x \in U^r$. Of course, more or less labels can be employed for different precision levels required.

Following the previous example, where arguments $A = \{90, 75, 65, 62, 60\}$ and $k \in \{1, 2\}$, Table I presents argument-specific reliability measures in both numerical and linguistic terms. Essentially, this fuzzy linguistic methodology allows the uniform and simple interpretation of arguments' reliability and their contribution toward the final aggregation result. It is effective as the explanatory means, particularly for data analysis or decision-making tasks that involve multiple analysts/experts.

B. Weight Determination for DOWA Aggregation

In addition to the purpose of interpretability, the reliability measure can be directly employed to determine argument-dependent weight vectors. In accordance with the stress-function method [64], for each argument $a_i \in A$, $A = \{a_1, \ldots, a_n\}$, its weight w_i is estimated from the order of its value $\text{Order}(a_i) \in \{1, \ldots, n\}$ within the descending-value list of arguments. Note that $\text{Order}(a_i) = 1$ when $a_i = \max(A)$, and likewise, $\text{Order}(a_i) = n$ when $a_i = \min(A)$. Using any stress function $s(x) \rightarrow R^+$, $x \in [0, 1]$, an argument-specific weight w_i is defined as

$$w_i = \frac{s\left(\frac{\text{Order}(a_i)}{n}\right)}{\sum_{j=1}^{n} s\left(\frac{\text{Order}(a_j)}{n}\right)}. \tag{15}$$

As the reliability measure is *order-independent*, each argument is now assigned with a specific degree of reliability, regardless of its position in the ordered argument list. The previous equation can, therefore, be generalized to the argument-dependent case as follows, where the reliability measure is represented as a stress-like discrete function $r(x) \rightarrow [0, 1]$, $x \in R$:

$$w_i = \frac{r(a_i)}{\sum_{j=1}^{n} r(a_j)}. \tag{16}$$

Note that the analogous formalism has been adopted with both DOWA [59] and Clus-DOWA [6] operators. Particularly to the k nearest-neighbor-based reliability, this definition can be simplified as

$$w_i^k = \frac{R^k_{a_i}}{\sum_{j=1}^{n} R^k_{a_j}} \tag{17}$$

where $R^k_{a_i}$ denotes the reliability measure of argument a_i, estimated from the set of its k nearest neighbors (i.e., arguments). Following that, the resulting argument-dependent operator, denoted as kNN-DOWA, can be specified by

$$\text{kNN-DOWA}(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} a_i w_i^k. \tag{18}$$

Similar to Clus-DOWA [6] and DOWA [59] operators, kNN-DOWA is *neat* (i.e., *order independent*), as it generates the same outcome regardless of the order of argument values [61]. Let $\{c_1, c_2, \ldots, c_n\}$ be any permutation of the argument vector $\{a_1, a_2, \ldots, a_n\}$. Then

$$\text{kNN-DOWA}(a_1, \ldots, a_n) = \text{kNN-DOWA}(c_1, \ldots, c_n). \tag{19}$$

As a continued example, kNN-DOWA is applied to the example of aggregating arguments $A = \{a_1 = 90, a_2 = 75, a_3 = 65, a_4 = 62, a_5 = 60\}$, whose reliability measures are presented in Table I. Accordingly, the weight vectors obtained from $k = 1$ and $k = 2$ nearest-neighbor reliability

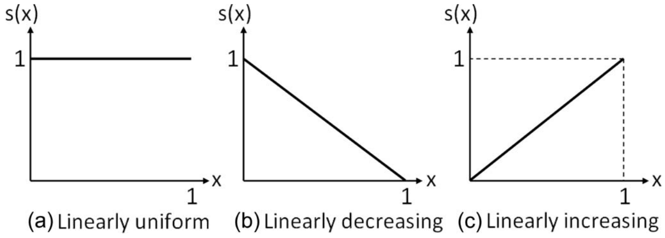

Fig. 5. Stress functions used to formulate (a) OWA-Stress1, (b) OWA-Stress2, and (c) OWA-Stress3 operators, respectively.

assessments are $\{w_1 = 0.127, w_2 = 0.169, w_3 = 0.229, w_4 = 0.237, w_5 = 0.237\}$ and $\{w_1 = 0.095, w_2 = 0.167, w_3 = 0.224, w_4 = 0.262, w_5 = 0.252\}$, respectively.

C. Application to Alias Detection in Intelligence Data

Here, a practical application of kNN-DOWA to alias detection in intelligence data is presented. The performance of the kNN-DOWA and other OWA operators (including both argument dependent and independent) is empirically examined. Note that alias detection is a crucial task to preventing terrorist and criminal activities [7], [19]. Particularly in the case of terrorism, alias and false identities are widely exploited to provide financial and logistical support to terrorist networks that have set up and encourage criminal activities to undermine civil society. Tracking and preventing terrorist activities undoubtedly require authentic identification of criminals and terrorists who typically possess multiple fraud and deceptive names, addresses, bank accounts, and telephone numbers.

A number of string matching techniques [40] have been invented to measure the similarity between a pair of textual entities and can be applied to detecting aliases of named objects. Intuitively, by combining similarity measures of different matching algorithms, superior results may be obtained. To facilitate comparative studies, the kNN-DOWA and other aggregation operators are respectively utilized to combine similarity measures that are derived by the use of different techniques: Levenshtein [40], Q-grams [33], Needleman–Wunsch [41], and Jaro [28].

The aggregation methods are evaluated over the challenging terrorist data set, which is manually extracted from Web pages and news stories related to terrorism [26]. Each entity presented in this link network is the name of a person, place, or organization, while a link denotes an association between objects through reported events. Statistically, this network contains 4088 entities, 5581 links, and 919 alias pairs.

In this evaluation, three linear stress functions (see Fig. 5) are exploited to generate argument-independent weight vectors and their corresponding OWA operators (denoted as OWA-Stress1, OWA-Stress2, and OWA-Stress3, respectively). By following (5) with $n = 4$ (i.e., the number of similarity measures to be aggregated), these weight vectors are $\{w_1 = 0.25, w_2 = 0.25, w_3 = 0.25, w_4 = 0.25\}$, $\{w_1 = 0.4, w_2 = 0.3, w_3 = 0.2, w_4 = 0.1\}$ and $\{w_1 = 0.1, w_2 = 0.2, w_3 = 0.3, w_4 = 0.4\}$, respectively. Note that the OWA-Stress1 operator is equivalent to a simple arithmetic mean, where each argument is allocated with a weight of $1/n$.

TABLE II
NUMBER OF ALIAS PAIRS DISCOVERED BY EACH METHOD USING TOP-β SIMILAR PAIRS

Methods	Disclosed alias pairs from Top-β similar pairs			
	β = 25	β = 50	β = 75	β = 100
Argument Independent				
OWA-Stress1	1	5	11	15
OWA-Stress2	1	6	10	14
OWA-Stress3	2	5	11	16
Argument Dependent				
DOWA	2	6	10	14
1NN-DOWA	3	9	12	18
2NN-DOWA	3	9	13	19
String Similarity				
Jaro	3	6	8	11

Table II shows the number of alias pairs discovered by each method, taking into account top-β name pairs with the highest similarity values, where $\beta \in \{25, 50, 75, 100\}$. Note that the results of the Clus-DOWA operator are not explicitly listed as they are identical to those of the 1NN-DOWA (i.e., kNN-DOWA with k being 1) counterpart, although the latter is more efficient. It is evidently illustrated that the similarity values derived by both 1NN-DOWA and 2NN-DOWA operators are more accurate than those generated by the DOWA and other argument-independent operators (i.e., OWA-Stress1, OWA-Stress2, and OWA-Stress3). In addition, the kNN-DOWA approach also outperforms the best individual string matching technique, i.e., Jaro.

These experimental results reflect well the underlying theoretical ideas. They have shown that the aggregation of string-matching scores generally improves the accuracy that is achievable by any single score alone. In particular, the distributed reliability measure exploited by the kNN-DOWA operators is more robust to extreme values, as compared with the centralized mechanism employed by the DOWA operator. Unlike the kNN-DOWA methods whose behaviors vary in accordance with discovered local consensus (group), existing data-independent operators may deliver inconsistent performance, as their weights are predefined without taking into account the properties of the actual data.

Note that the explanation mechanism (as shown in Table I) can assist data analysts to validate the results generated by the kNN-DOWA approach. This capability helps to reduce the problem of false positives, where innocent individuals have been identified as suspects. Also, the explanatory formalism allows a flexible linguistic-like retrieval of suspected cases.

V. APPLICATION OF DATA RELIABILITY TO UNSUPERVISED FEATURE SELECTION

To further demonstrate the potential of the current research, here, another application of data reliability to the task of unsupervised feature selection is presented. Fundamentally, this application aims to reduce a number of features for more efficient data analysis. The benefits of such work include minimizing the measurement and storage requirements, reducing training and run time, and defying the curse of dimensionality to improve prediction performance [29], [30], [36]. The proposed method uses the reliability measure to justify the relevance (or importance) of each feature and, hence, the possibility of being included in the selected feature subset. Its performance is assessed, over a number of benchmark data sets from the UCI Machine Learning Repository [2], against typical unsupervised methods introduced in the literature.

A. Unsupervised Feature Selection

Feature selection is one of the most significant developments in machine learning [5], [32] and data mining [10], [36]. In particular, it has been applied to a variety of domain applications such as text categorization [35], intrusion detection [34], and customer relationship management [42]. Much of the work in feature selection has followed the supervised approach where invented methods rely on the class or decision labels and their correlation with feature values [37]. However, as argued in [22], *unsupervised feature selection* algorithms prove to be extremely useful with real-world data analysis. These techniques base their judgments on particular characteristics of data values such as entropy [9] and locality preserving ability [68]. In general, when decision labels are available, supervised feature selection methods usually outperform their unsupervised counterparts [4]. Despite this, in many cases where the thorough interpretation of a large data is infeasible, the amount of labeled training samples is often limited. In such circumstances, most conventional supervised techniques may fail on the "small labeled-sample problem" [27].

Unsupervised feature selection algorithms can be categorized into two classes of *wrapper* and *filter* [37]. The former evaluates the candidate feature subsets by the data modeling or clustering algorithm itself. Methods in this category aim to maximize the clustering performance that is gauged using an internal index (e.g., compactness and separability). These include sequential feature selection algorithms [11], expectation–maximization-based methods [13], and neurofuzzy techniques [45]. Unlike these approaches, the filter methodology is exploited as a preprocessing step that is absolutely independent of the learning algorithm used for data generalization. Although the wrapper approach may generate feature subsets of a better quality given a particular learning task, it is less efficient than the filter approach, where the repeated executions of a clustering algorithm are not required [24]. As such, the filter model is often chosen when the number of features concerned is large.

Example Filter Methods to Unsupervised Feature Selection: The filter approach determines the selection of features on their relevance or dependence. A number of such methods have been introduced in the literature with different feature evaluation measures [4], [9], [39], [68]. An initial technique is based on the concept of feature variance, which is employed to reflect a feature's representative capability. Effectively, those with high variance are selected [4]. Let f_{ir} be the rth feature value of the ith data instance x_i, $i = 1, \ldots, n, r = 1, \ldots, m$. The variance score V_r of the rth feature is defined as follows, where $\mu_r = (1/n)\sum_{i=1}^{n} f_{ir}$:

$$V_r = \frac{1}{n}\sum_{i=1}^{n}(f_{ir} - \mu_r)^2. \tag{20}$$

This intuitive measure has been extended with the Laplacian score [68]. In addition to favoring features with larger variances, the extended method also prefers those with strong locality preserving ability. Here, the Laplacian score L_r of the rth feature, which should be minimized, is estimated as

$$L_r = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(f_{ir} - f_{jr})^2 S_{ij}}{\sum_{i=1}^{n}(f_{ir} - \mu_r)^2 D_{ii}} \tag{21}$$

where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} S_{ij}$, and S_{ij} is defined as the neighborhood relation between samples x_i and x_j such that

$$S_{ij} = \begin{cases} e^{\frac{|x_i - x_j|^2}{t}}, & \text{if } x_i \text{ and } x_j \text{ are neighbors} \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where t is a user-defined constant, and sample x_j is a neighbor of x_i if $x_j \in N_i^k$, i.e., the set of k samples nearest to x_i.

In addition, the entropy-based relevance assessment has also been developed for unsupervised feature selection [9]. Unlike the previous measures, the underlying evaluation is not conducted for individual features in isolation, but for feature subsets each created on a leave-one-out basis. The entropy E_r of the feature subset without the rth feature can be estimated as follows:

$$E_r = -\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}.\log(S_{ij}) + (1 - S_{ij}).\log(1 - S_{ij})) \tag{23}$$

where $S_{ij} = e^{-\alpha D_{ij}}$ and $\alpha = -\log(0.5)/\overline{D}$. Note that D_{ij} is the Euclidean distance between samples x_i and x_j, and $\overline{D}$ is the corresponding mean distance between any two samples in a given feature subspace (i.e., where the rth feature is excluded). Effectively, the higher the E_r is, the more significance the rth feature becomes.

Another approach to developing filter methods for unsupervised feature selection is based on the concept of similarity among features [39]. It aims to reduce the redundancy by partitioning the original feature set into subsets (or clusters). Features in the same cluster are regarded to be highly similar, whereas those in different clusters are dissimilar. One feature is then selected from each cluster (as a representative) to constitute the final selected feature subset. In particular, the new similarity measure, named *maximal information compression index* λ_2, is introduced to clustering the underlying features (see [39] for details).

In addition to these methods, techniques like the principal component analysis [12], Isomap [54] (a nonlinear extension of the multidimensional scaling [56]), and the locally linear embedding mechanism [50] can also be used for unsupervised dimensionality reduction. However, these methods deliver a subset of transformed features, instead of an actual subset of the original features. A more comprehensive overview of approaches for dimensionality reduction via feature transformation can be found in [29].

Search Strategy for the Reduced Feature Subset: For an m-dimensional data set (i.e., with m features), the complete search space of the feature selection problem is of the size 2^m. Accordingly, an exhaustive search becomes impractical, even with a moderate m. For this reason, a heuristic search technique may be employed, including sequential methods (e.g., forward/backward selection [1] and floating search [47]), branch-and-bound [8], and randomized search strategies (e.g., evolutionary algorithms [51]). An alternative is that of feature ranking, where the feature-specific utility is assessed in isolation, and those features with the utility above a certain threshold are selected [22], [24]. Particularly, in [46], a simple threshold-directed method is introduced to discriminate the relevance of the original features, within the unsupervised learning framework of conditional Gaussian networks. Also, a technique for aggregation of simple feature rankings, each created from a specific projection of the underlying data, has recently been put forward with promising results [25].

B. Reliability-Based Filter Method

Inspired by the success with OWA aggregation, the proposed reliability measure is herein also applied to the problem of unsupervised feature selection. It can be regarded as the discriminant factor to justifying the relevance of each data feature. The resulting "filter" method reflects the intuition that a feature is considered reliable (or relevant) if its values are tightly grouped together (i.e., possessing *a rigid value pattern*). In essence, with a data set of n samples $(x_1, \ldots, x_n)$, the reliability FR_r of feature f_r, $r \in \{1, \ldots, m\}$, is estimated from the accumulative reliability measures generated for each of its values f_{ir}, $i = 1, \ldots, n$. This is summarized as follows.

- *Step 1.* Acquire the reliability measure $R^k_{f_{ir}}$ of each feature value $f_{ir}, i = 1, \ldots, n$ [see (9) and (10)], using the set of k nearest neighbors.
- *Step 2.* Calculate the accumulative reliability FR_r of feature $f_r, r = 1, \ldots, m$, by combining the reliability measures of all its values, i.e.,

$$FR_r = \sum_{i=1}^{n} R^k_{f_{ir}}. \tag{24}$$

Effectively, the original features can be ranked in accordance with their reliability degrees. The higher the reliability is, the more relevant the feature becomes. In the current research, for computational efficiency, a simple threshold-directed feature selection method similar to those in [24], [25], and [46] is employed. Principally, a feature $f_r, r \in \{1, \ldots, m\}$, is selected only when its corresponding reliability FR_r exceeds a given threshold. Such a discriminating limit can be subjectively modeled by an analyst. However, a predefined threshold may not be effective for a variety of data with different characteristics. It is better to learn this from the underlying data set. Intuitively, the original features can be divided into two classes ("relevance" and "irrelevance") by using the average reliability of all features FR_{average} as the threshold, i.e.,

$$FR_{\text{average}} = \frac{1}{m} \sum_{r=1}^{m} FR_r. \tag{25}$$

```
ReduceFeatureSet($F, FR$)
$F$, the original feature ser, $F = (f_1 \ldots f_m)$;
$f_i$, the data feature, $i, i = 1 \ldots m$;
$FS$, the reduced feature set;
$FR$, the set of feature reliability, $FR = (FR_i \ldots FR_m)$;
$FR_i$, the reliability of feature $i, i = 1 \ldots m$;
$FR_{average}$, the average reliability of all features;

$FS \leftarrow F$
$FR_{average} = \frac{1}{m} \sum_{i=1}^{m} FR_i$
for each $f_i \in F$
    if $FR_i < FR_{average}$
        $FS \leftarrow FS - f_i$
return $FS$
```

Fig. 6. Procedural description of the *ReduceFeatureSet* algorithm.

From this, a heuristic method can be employed to justify the content of the reduced feature set FS. Let the to-be-reduced feature set FS contain all features $f_1, \ldots, f_m$. Next, for each feature f_r, it is dropped from FS if $FR_r < FR_{\text{average}}$. Fig. 6 formally summarizes this *ReduceFeatureSet* algorithm.

C. Empirical Evaluation

Having defined the reliability-based feature selection algorithm, here, the evaluation of its performance against three other unsupervised feature selection techniques, namely, Laplacian (LPC) [68], entropy (ENT) [9], and feature similarity (FSFS) [39], is presented. To generalize this assessment, two different sizes of the nearest neighbors ($k = 1$ and $k = 2$) are exploited to create two variations of the proposed reliability-based methods, named R-1 and R-2, respectively. Effectively, a feature subset selected by each investigated method is used by three distinct clustering techniques (see next) to generate data partitions (whose qualities are gauged using three standard evaluation indexes, as shown later). Intuitively, the higher the quality of a data partition, the more effective the feature subset, and, hence, the more useful the underlying feature selection method. Note that this evaluation is conducted over six benchmark data sets, where true natural clusters are known but are not explicitly used by an unsupervised feature selection method (except for their involvement in the evaluation of the final results). These data sets are obtained from UCI Machine Learning Repository [2], and their details are summarized in Table III.

Clustering Methods: Three different clustering algorithms are used here to evaluate the quality of a reduced feature set created by each investigated method. For comparison purposes, as with the work in [15], [18], and [21], each clustering method divides data samples into a partition of K (the number of *true classes* for each data set) clusters, which is then assessed

TABLE III
DETAILS OF EXPERIMENTED DATA SETS

Dataset Name	No. of Samples	No. of Features	No. of Classes
Iris	150	4	3
Glass	214	9	6
Cleveland	297	13	5
Heart	270	13	2
Olitos	120	25	4
Ionosphere	230	34	2

against the corresponding true partition using a set of external evaluation indexes as given below.

k-means (KM) [23] first randomly selects (predefined) k samples as initial centroids, to which the remaining samples are assigned. Following that, the centroid of each cluster is updated as the mean of all samples in that cluster. This process is iterated until no changes are made to the centroids (i.e., no reassignment of any data sample from one cluster to another).

Single linkage [12] generates a tree (called a "dendogram") as nested groups of data organized hierarchically. The algorithm begins by considering each data sample as a cluster, and then gradually merges similar clusters until all the clusters are combined into one big group. The resulting dendogram reveals cluster–subcluster relations and the order in which they were merged or split. This is obtained using the distance $D_{C_iC_j}$ between two clusters C_i and C_j, where $d(a,b)$ is the distance between samples a and b, i.e.,

$$D_{C_iC_j} = \min_{\forall a \in C_i, b \in C_j} d(a,b). \qquad (26)$$

Spectral clustering [17] employs the spectrum of data similarity matrices to first reduce the dimensionality of a data set and then applies a basic clustering algorithm, such as KM or a graph cut-based method, on the resulting lower dimension data. In this regard, the method is itself of a hybrid of dimensionality reduction and clustering already. This similarity is typically measured using a Gaussian function [52]. Interestingly, such a method makes no assumptions on the data distribution at hand. It is also able to find clusters that are not in any convex regions of the space.

Evaluation Indices: Similar to the quality assessment of clustering methods used in [18], [53], and [55] (where class labels are assumed available just to perform the evaluation), the data partitions generated by the aforementioned clustering algorithms are here evaluated using three validity indexes: classification accuracy (CA) [43], normalized mutual information (NMI) [53], and Rand index (RI) [48]. These evaluation indexes have been widely exploited for assessing the quality of data partitions generated by a clustering algorithm. In particular, to the task of unsupervised feature selection, such assessment is also employed to justify the quality of established feature selection methods [13], [31], [39]. Note that these indexes assess the degree of agreement between two data partitions, where one of the partitions is obtained from a clustering algorithm (π^*), and the other is taken from the assumed prior information [i.e., the known label of the data (Π') that is not required for the actual clustering process].

Experiment Results: At the outset, for comparison, all the feature selection methods employed in this experimental study are assessed using the same feature-subset size per data set, which is equal to those generated by either of the two proposed reliability-based techniques (i.e., R-1 and R-2). Note that both R-1 and R-2 happen to create feature subsets of the same cardinality for each data set. Thus, the sizes are 1, 4, 8, 8, 10, and 10 for Iris, Glass, Cleveland, Heart, Olitos, and Ionosphere, respectively. Table IV presents the performance of different unsupervised algorithms based on CA, NMI, and RI quality indexes. The two best or equal-best performances (marked as best-2 hereafter) under each setting (excluding the case for the "Unreduced") are highlighted with boldface.

The results indicate that the reliability-based approach usually outperforms other unsupervised feature selection methods across all three distinct clustering techniques. Exceptionally, its performance is competitive to that of FSFS for Glass and Olitos, while being superior over other data sets. Note that since KM is nondeterministic (i.e., different runs may create dissimilar data partitions), the results shown in this evaluation are acquired as the average across 50 runs. In deriving these results, the parameters used by the LPC method are as follows: $t = 1$ [see (21) and (22)] and the number of nearest neighbors k is 2.

To further evaluate the performance of R-1 and R-2, reduced feature sets of a different size $rd, rd = 1, \ldots, m-1$, are similarly assessed using the given clustering techniques and quality indexes, where m is the number of original features. Note that unsupervised methods analogously generate a list of ranked features, where the lowest g features are excluded to obtain the reduced feature set of size $m - g$. Following the assessment framework in [67], using CA, NMI, and RI quality measures, Table V shows the averaged performance [with the corresponding standard deviation (SD) statistics] achieved by each investigated algorithm across all $m - 1$ reduced feature sets. According to these results, the proposed reliability approach systematically provides more consistent and effective performance than the rest. For detailed results, please consult online resources.[1]

These results demonstrate that the proposed feature selection mechanism is, indeed, effective and robust to perform the task of identifying relevant data features. This conforms to the design intention of the underlying approach in that the accumulative reliability measure captures well the "compactness" of hypothesized data clusters, without prior knowledge of the actual ones. Interestingly, the property of compactness has been the typical objective function of many clustering algorithms, including KM. The approach proposed here is conceptually similar to the LPC method that takes into account the local data structure. However, LPC has the difficulty in coping with extreme data values, whereas R-1 and R-2 can minimize the effect of such data upon the clustering process.

In addition, ENT concentrates on pairwise-proximity metric that effectively blends local and broader data structures together. As suggested by the empirical findings, this inability to account for local data property brings about less effective performance, as compared to R-1 and R-2. Different from these

[1] http://users.aber.ac.uk/tsb/.

TABLE IV
PERFORMANCE OF UNSUPERVISED TECHNIQUES USING CA–NMI–RI EVALUATION INDEXES. THE BEST-2 PERFORMANCE OF EACH INDEX IS HIGHLIGHTED IN BOLDFACE

Dataset	Clustering	Unsupervised Method					
	Technique	Unreduced	LPC	ENT	FSFS	R-1	R-2
Iris	SL	0.667—0.742—0.772	0.347—0.050—0.338	0.373—0.101—0.359	0.347—0.050—0.338	**0.667—0.730—0.768**	**0.667—0.730—0.768**
	KM	0.830—0.703—0.834	0.550—0.204—0.602	0.691—0.381—0.713	0.551—0.204—0.600	**0.960—0.868—0.950**	**0.954—0.863—0.946**
	SPT	0.847—0.664—0.837	0.520—0.193—0.555	0.673—0.349—0.679	0.520—0.193—0.555	**0.893—0.791—0.880**	**0.893—0.791—0.880**
Glass	SL	0.379—0.138—0.303	0.379—0.121—**0.325**	0.379—0.118—0.300	**0.425—0.223—0.376**	**0.383—0.147**—0.307	**0.383—0.147**—0.307
	KM	0.543—0.313—0.672	0.502—0.295—0.655	0.478—0.228—0.617	**0.596—0.385—0.716**	**0.562—0.305—0.670**	**0.562—0.305—0.670**
	SPT	0.537—0.318—0.668	0.478—0.287—0.623	0.466—0.229—0.618	**0.567—0.304—0.654**	**0.556—0.312—0.684**	**0.556—0.312—0.684**
Cleveland	SL	0.549—0.067—0.370	0.549—0.056—**0.436**	0.549—0.056—**0.436**	0.539—0.038—0.433	**0.552—0.076**—0.422	**0.552—0.076**—0.422
	KM	0.574—0.161—0.637	0.560—0.123—0.638	0.559—0.121—0.639	0.541—0.068—0.579	**0.568—0.158—0.642**	**0.562—0.153—0.639**
	SPT	0.574—0.130—0.617	0.559—**0.137**—0.613	0.559—0.137—0.624	0.549—0.082—0.614	**0.577—0.139—0.635**	**0.576**—0.137—**0.634**
Heart	SL	0.559—0.023—0.505	0.556—0.001—0.500	0.556—0.001—0.500	0.556—0.000—**0.504**	**0.560—0.023—0.505**	**0.560—0.017—0.504**
	KM	0.756—0.221—0.639	0.645—0.067—0.546	0.648—0.070—0.547	0.594—0.036—0.516	**0.740—0.205—0.629**	**0.764—0.228—0.645**
	SPT	0.800—0.277—0.679	0.682—0.092—0.564	0.682—0.092—0.564	0.589—0.024—0.514	**0.785—0.252—0.661**	**0.774—0.232—0.649**
Olitos	SL	0.442—0.121—0.335	0.425—0.099—0.336	**0.442**—0.121—0.335	0.425—0.089—0.323	**0.442—0.148—0.342**	**0.442—0.148—0.342**
	KM	0.766—0.516—0.780	0.596—**0.308**—0.662	0.580—0.246—0.648	**0.672—0.376—0.714**	**0.629**—0.302—**0.679**	**0.629**—0.302—**0.679**
	SPT	0.742—0.527—0.768	0.579—0.212—0.657	0.620—0.212—0.676	**0.692—0.400—0.721**	**0.659—0.326—0.691**	**0.659—0.326—0.691**
Ionosphere	SL	0.509—0.022—0.498	0.509—0.022—0.498	0.509—0.022—0.498	**0.657—0.222—0.547**	**0.657—0.222—0.547**	**0.657—0.222—0.547**
	KM	0.748—0.186—0.621	0.658—0.150—0.549	0.739—0.174—0.613	0.633—0.113—0.542	**0.745—0.196—0.619**	**0.741—0.190—0.615**
	SPT	0.748—0.185—0.621	0.678—0.102—0.562	0.735—0.168—0.609	0.600—0.032—0.518	**0.752—0.194—0.626**	**0.748—0.186—0.621**

TABLE V
AVERAGE OF CA–NMI–RI MEASURES OBTAINED FROM DIFFERENT FEATURE SUBSETS (OF SIZE $1, \ldots, m-1$). THE CORRESPONDING SDS ARE GIVEN IN BRACKETS, AND THE TWO BEST MEASURES OF EACH EXPERIMENT SETTING ARE SHOWN IN BOLDFACE

Dataset	Clustering	Unsupervised Method				
	Technique	LPC	ENT	FSFS	R-1	R-2
Iris	SL	0.456—0.285—0.484	0.573—0.530—0.637	0.456—0.285—0.484	**0.671—0.740—0.773**	**0.669—0.739—0.772**
		(0.183—0.396—0.249)	(0.173—0.372—0.241)	(0.183—0.396—0.249)	(0.004—0.009—0.005)	(0.004—0.008—0.004)
	KM	0.719—0.486—0.741	0.819—0.613—0.818	0.719—0.490—0.741	**0.935—0.829—0.925**	**0.885—0.745—0.879**
		(0.146—0.245—0.120)	(0.111—0.203—0.091)	(0.146—0.250—0.122)	(0.043—0.064—0.043)	(0.067—0.104—0.062)
	SPT	0.700—0.464—0.720	0.802—0.612—0.798	0.698—0.463—0.718	**0.904—0.806—0.891**	**0.851—0.705—0.845**
		(0.158—0.236—0.143)	(0.112—0.230—0.104)	(0.155—0.235—0.142)	(0.032—0.039—0.031)	(0.043—0.088—0.035)
Glass	SL	0.379—0.130—0.312	0.378—0.122—0.316	**0.400—0.174—0.336**	**0.380—0.138—0.317**	**0.380—0.138—0.317**
		(0.004—0.009—0.012)	(0.003—0.016—0.035)	(0.025—0.051—0.039)	(0.003—0.011—0.007)	(0.003—0.011—0.007)
	KM	0.518—0.283—0.654	0.480—0.229—0.611	**0.561—0.355—0.667**	**0.531—0.291—0.664**	**0.531—0.291—0.664**
		(0.027—0.033—0.010)	(0.052—0.077—0.066)	(0.040—0.030—0.081)	(0.044—0.075—0.019)	(0.044—0.075—0.019)
	SPT	0.525—**0.306—0.660**	0.472—0.234—0.602	**0.540—0.311**—0.635	**0.532**—0.285—**0.650**	**0.532**—0.285—**0.650**
		(0.035—0.058—0.017)	(0.052—0.078—0.057)	(0.042—0.009—0.066)	(0.037—0.042—0.028)	(0.037—0.042—0.028)
Cleveland	SL	0.548—0.062—0.412	0.551—0.072—0.455	0.544—0.060—0.473	**0.557—0.117—0.505**	**0.558—0.119—0.506**
		(0.005—0.020—0.068)	(0.011—0.028—0.098)	(0.006—0.022—0.079)	(0.013—0.056—0.038)	(0.013—0.054—0.037)
	KM	0.559—0.111—0.624	0.563—0.129—0.636	0.551—0.097—0.591	**0.570—0.168—0.648**	**0.570—0.166—0.646**
		(0.013—0.048—0.022)	(0.010—0.040—0.021)	(0.011—0.039—0.032)	(0.010—0.019—0.016)	(0.011—0.019—0.015)
	SPT	0.555—0.107—0.618	0.556—0.119—0.629	0.551—0.099—0.606	**0.571—0.148—0.634**	**0.571—0.148—0.635**
		(0.011—0.049—0.023)	(0.011—0.039—0.027)	(0.011—0.030—0.021)	(0.014—0.025—0.023)	(0.013—0.025—0.022)
Heart	SL	0.560—0.017—0.504	0.561—0.015—0.504	0.583—0.035—0.513	**0.639—0.091—0.553**	**0.625—0.079—0.545**
		(0.010—0.012—0.003)	(0.013—0.017—0.005)	(0.028—0.027—0.010)	(0.094—0.080—0.060)	(0.094—0.083—0.061)
	KM	0.685—0.119—0.578	0.673—0.109—0.572	0.634—0.079—0.545	**0.741—0.194—0.626**	**0.748—0.201—0.630**
		(0.061—0.063—0.040)	(0.079—0.083—0.054)	(0.058—0.065—0.043)	(0.021—0.024—0.016)	(0.014—0.016—0.010)
	SPT	0.707—0.142—0.593	0.689—0.132—0.587	0.655—0.099—0.563	**0.783—0.246—0.659**	**0.777—0.239—0.653**
		(0.068—0.077—0.049)	(0.099—0.113—0.073)	(0.097—0.112—0.073)	(0.017—0.029—0.019)	(0.017—0.029—0.019)
Olitos	SL	0.438—0.112—0.336	**0.445—0.133—0.347**	0.434—0.109—0.333	0.441—0.121—0.341	**0.443—0.138—0.349**
		(0.006—0.022—0.004)	(0.007—0.021—0.018)	(0.006—0.024—0.012)	(0.009—0.025—0.038)	(0.008—0.025—0.036)
	KM	0.631—0.325—0.687	0.624—0.334—0.686	**0.677—0.381—0.720**	**0.645—0.343—0.697**	0.636—0.341—0.695
		(0.096—0.145—0.070)	(0.107—0.131—0.066)	(0.084—0.109—0.051)	(0.078—0.107—0.051)	(0.098—0.103—0.062)
	SPT	0.610—0.302—0.673	0.617—0.312—0.674	**0.663—0.385—0.709**	**0.634**—0.320—0.680	0.624—**0.324—0.685**
		(0.093—0.132—0.069)	(0.113—0.143—0.073)	(0.087—0.116—0.057)	(0.078—0.098—0.047)	(0.098—0.134—0.063)
Ionosphere	SL	0.530—0.050—0.505	0.534—0.054—0.506	0.536—**0.058**—0.507	**0.540—0.064—0.508**	**0.540—0.064—0.508**
		(0.051—0.068—0.017)	(0.055—0.073—0.018)	(0.058—0.078—0.019)	(0.061—0.083—0.020)	(0.061—0.083—0.020)
	KM	0.703—0.177—0.588	**0.732**—0.178—**0.610**	0.661—0.132—0.563	0.731—**0.186—0.611**	**0.734—0.180—0.610**
		(0.058—0.046—0.040)	(0.045—0.038—0.030)	(0.065—0.048—0.039)	(0.050—0.034—0.032)	(0.050—0.035—0.032)
	SPT	0.712—0.178—0.592	**0.736**—0.183—**0.612**	0.626—0.087—0.543	**0.737—0.194—0.613**	0.731—**0.190**—0.610
		(0.044—0.037—0.036)	(0.038—0.033—0.027)	(0.081—0.083—0.045)	(0.032—0.018—0.027)	(0.033—0.029—0.027)

techniques, FSFS works by exploiting only the redundancy contained within a given feature set. It does not consider any information on data relevance in determining the resulting feature subset. This mechanism is efficient, but inapplicable to data sets with completely dissimilar features. However, a hybrid method that combines the underlying data redundancy with data reliability may improve the results obtained by any individual method presented herein.

VI. Conclusion

This paper has presented a nearest-neighbor-based reliability measure, which can be efficiently exploited for information aggregation and unsupervised feature selection. The distributed perspective, similar to the existing cluster-based technique (of the Clus-DOWA operator), has been adopted such that the reliability of a data point is determined solely by its distances to the local neighbors. In essence, values that are seemingly indifferent from others in close proximity are considered reliable. This approach is more efficient than the conventional cluster-based counterpart [with time and space complexity decreasing from $O(n^3)$ and $O(n^2)$ to $O(n^2)$ and $O(n)$, respectively]. Although the fundamental concept of nearest neighbors has been well established, its application within the context of *information aggregation* and *unsupervised feature selection* is unique. This paper concentrates on such novel applications.

Technically, the reliability measure can be regarded as a stress-like function to obtain an explanatory and argument-dependent OWA aggregation. Conceptually, the data-driven approach developed herein offers a reverse-engineered approach to the original human-directed stress-function method. While reliability is analogously employed for interpretation toward an aggregation behavior, it is not derived in accordance with human experience and judgment, but from the structural characteristics of the argument values. In addition, reliability values can be mapped onto linguistic descriptors such that a coherent comprehension, particularly among multiple experts or analysts, can be achieved. The distributed approach is able to effectively capture the underlying data characteristics and deliver trustworthy weights. This is illustrated through its superior performance compared with other dependent and independent aggregation methods for alias detection in intelligence data.

Furthermore, the generality of the proposed method has been demonstrated by applying it to the task of unsupervised feature selection. To reduce the size of a feature set to support a more efficient subsequent learning process, the reliability measure has been employed to determine the quality of each feature. Intuitively, a feature is excluded from the final feature subset if its reliability is below the average measure of all features. This heuristic-based method has proven effective in conjunctive use with several clustering algorithms over a number of data sets.

Results obtained are highly promising. However, the generality of the nearest-neighbor guided evaluation of data reliability may be further illustrated by application to other problem domains. One initial such investigation is automated assessment of student academic performance. This is important and very relevant to the present research because such evaluation typically requires the exploitation of multiple criteria and may involve different forms of data [49]. The proposed information aggregation technique can be used to integrate different decision-making criteria, and the unsupervised feature selection method can be applied to choose appropriate data components when assessing a certain aspect of the performance.

Acknowledgment

The authors would like to thank the members of the project team for their contribution, while taking full responsibility for the views expressed here. The authors would also like to thank the anonymous referees for their constructive comments that have helped considerably in revising this paper.

References

[1] D. W. Aha and R. L. Bankert, “A comparative evaluation of sequential feature selection algorithms,” in *Learning From Data*. New York: Springer-Verlag, 1996, pp. 199–206.

[2] A. Asuncion and D. J. Newman, UCI Machine Learning Repository, Irvine, CA: School Inf. Comput. Sci., Univ. California2007. [Online]. Available: www.ics.uci.edu/~mlearn/MLRepository.html

[3] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin, Germany: Springer-Verlag, 2007.

[4] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[5] A. L. Blum and P. Langley, “Selection of relevant features and examples in machine learning,” *Artif. Intell.*, vol. 97, no. 1/2, pp. 245–271, Dec. 1997.

[6] T. Boongoen and Q. Shen, “Clus-DOWA: A new dependent OWA operator,” in *Proc. IEEE Int. Conf. Fuzzy Sets Syst.*, 2008, pp. 1057–1063.

[7] T. Boongoen, Q. Shen, and C. Price, “Disclosing false identity through hybrid link analysis,” *AI and Law*, to be published, DOI: 10.1007/s10506-010-9085-9.

[8] X. Chen, “An improved branch and bound algorithm for feature selection,” *Pattern Recognit. Lett.*, vol. 24, no. 12, pp. 1925–1933, Aug. 2003.

[9] M. Dash, K. Choi, P. Scheuermann, and H. Liu, “Feature selection for clustering: A filter solution,” in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 115–122.

[10] M. Dash and H. Liu, “Feature selection for classification,” *Int. J. Intell. Data Anal.*, vol. 1, no. 3, pp. 131–156, 1997.

[11] M. Dash and H. Liu, “Unsupervised feature selection and ranking,” in *New Trends in Knowledge Discovery for Business Information Systems*. Norwell, MA: Kluwer, 2000.

[12] P. A. Denvijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[13] J. G. Dy and C. E. Brodley, “Feature selection for unsupervised learning,” *J. Mach. Learn. Res.*, vol. 5, pp. 845–889, Dec. 2004.

[14] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, “Cluster analysis and display of genome-wide expression patterns,” *Proc. Nat. Acad. Sci. U.S.A.*, vol. 95, no. 25, pp. 14 863–14 868, Dec. 1998.

[15] X. Z. Fern and C. E. Brodley, “Solving cluster ensemble problems by bipartite graph partitioning,” in *Proc. Int. Conf. Mach. Learn.*, 2004, pp. 36–43.

[16] D. Filev and R. R. Yager, “On the issue of obtaining OWA operator weights,” *Fuzzy Sets Syst.*, vol. 94, no. 2, pp. 157–169, Mar. 1998.

[17] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, “A survey of kernel and spectral methods for clustering,” *Pattern Recognit.*, vol. 41, no. 1, pp. 176–190, Jan. 2008.

[18] A. L. N. Fred and A. K. Jain, “Combining multiple clusterings using evidence accumulation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, Jun. 2005.

[19] X. Fu, T. Boongoen, and Q. Shen, “Evidence directed generation of plausible crime scenarios with identity resolution,” *Appl. Artif. Intell.*, to be published.

[20] R. Fuller, “On obtaining OWA operator weights: A short survey of recent developments,” in *Proc. IEEE Int. Conf. Comput. Cybern.*, 2007, pp. 241–244.

[21] A. Gionis, H. Mannila, and P. Tsaparas, “Clustering aggregation,” in *Proc. Int. Conf. Data Eng.*, 2005, pp. 341–352.

[22] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[23] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Mateo, CA: Morgan Kaufmann, 2000.

[24] J. Handl and J. Knowles, "Feature subset selection in unsupervised learning via multiobjective optimization," *Int. J. Comput. Intell. Res.*, vol. 2, no. 3, pp. 217–238, 2006.

[25] Y. Hong, S. Kwong, Y. Chang, and Q. Ren, "Consensus unsupervised feature ranking from multiple views," *Pattern Recognit. Lett.*, vol. 29, no. 5, pp. 595–602, Apr. 2008.

[26] P. Hsiung, A. Moore, D. Neill, and J. Schneider, "Alias detection in link data sets," in *Proc. Int. Conf. Intell. Anal.*, 2005.

[27] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.

[28] M. A. Jaro, "Probabilistic linkage of large public health data files," *Stat. Med.*, vol. 14, no. 5–7, pp. 491–498, Mar./Apr. 1995.

[29] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*. Piscataway, NJ: IEEE Press, 2008.

[30] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Jul. 2009.

[31] Y. Kim, W. N. Street, and F. Menczer, "Evolutionary model selection in unsupervised learning," *Intell. Data Anal.*, vol. 6, no. 6, pp. 531–556, Dec. 2002.

[32] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1/2, pp. 273–324, Dec. 1997.

[33] K. Kukich, "Techniques for automatically correcting words in text," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 377–439, Dec. 1992.

[34] W. Lee, S. J. Stolfo, and K. W. Mok, "Adaptive intrusion detection: A data mining approach," *Artif. Intell. Rev.*, vol. 14, no. 6, pp. 533–567, Dec. 2000.

[35] E. Leopold and J. Kindermann, "Text categorization with support vector machines: How to represent texts in input space?" *Mach. Learn.*, vol. 46, no. 1–3, pp. 423–444, Jan. 2002.

[36] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Norwell, MA: Kluwer, 1998.

[37] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[38] X. Liu, "Some properties of the weighted OWA operator," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 1, pp. 118–127, Feb. 2006.

[39] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.

[40] G. Navarro, "A guided tour to approximate string matching," *ACM Comput. Surv.*, vol. 33, no. 1, pp. 31–88, Mar. 2001.

[41] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, Mar. 1970.

[42] K. S. Ng and H. Liu, "Customer retention via data mining," *Artif. Intell. Rev.*, vol. 14, no. 6, pp. 569–590, Dec. 2000.

[43] N. Nguyen and R. Caruana, "Consensus clusterings," in *Proc. IEEE Int. Conf. Data Mining*, 2007, pp. 607–612.

[44] M. O'Hagan, "Aggregating template rule antecedents in real-time expert systems with fuzzy set logic," in *Proc. Annu. IEEE Conf. Signals, Syst., Comput.*, 1988, pp. 681–689.

[45] S. K. Pal, R. K. De, and J. Basak, "Unsupervised feature evaluation: A neuro-fuzzy approach," *IEEE Trans. Neural Netw.*, vol. 11, no. 2, pp. 366–376, Aug. 2000.

[46] J. M. Pena, J. A. Lozano, P. Larranaga, and I. Inza, "Dimensionality reduction in unsupervised learning of conditional Gaussian networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 590–603, Jun. 2001.

[47] P. Pudil, J. Novovicov'a, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[48] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.

[49] K. A. Rasmani and Q. Shen, "Data-driven fuzzy rule generation and its application for student academic performance evaluation," *Appl. Intell.*, vol. 25, no. 3, pp. 305–319, Dec. 2006.

[50] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[51] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognit. Lett.*, vol. 10, no. 5, pp. 335–347, Nov. 1989.

[52] M. C. P. Souto, I. G. Costa, D. S. A. Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: A comparative study," *BMC Bioinformatics*, vol. 9, p. 497, Nov. 2008.

[53] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Mar. 2003.

[54] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[55] A. P. Topchy, A. K. Jain, and W. F. Punch, "Clustering ensembles: Models of consensus and weak partitions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1866–1881, Oct. 2005.

[56] W. S. Torgenson, "Multidimensional scaling," *Psychometrika*, vol. 17, pp. 401–419, 1952.

[57] L. Troiano and R. R. Yager, "Recursive and iterative OWA operators," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 13, no. 6, pp. 579–599, Dec. 2005.

[58] Z. Xu, "An overview of methods for determining OWA weights," *Int. J. Intell. Syst.*, vol. 20, no. 8, pp. 843–865, Aug. 2005.

[59] Z. Xu, "Dependent OWA operators," in *Proc. Int. Conf. Model Decisions Artif. Intell.*, 2006, pp. 172–178.

[60] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, Jan./Feb. 1988.

[61] R. R. Yager, "Families of OWA operators," *Fuzzy Sets Syst.*, vol. 59, no. 2, pp. 125–148, 1993.

[62] R. R. Yager, "Quantifier guided aggregation using OWA operators," *Int. J. Intell. Syst.*, vol. 11, no. 1, pp. 49–73, 1996.

[63] R. R. Yager, "Centered OWA operators," *Soft Comput.*, vol. 11, no. 7, pp. 631–639, Feb. 2007.

[64] R. R. Yager, "Using stress functions to obtain OWA operators," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 6, pp. 1122–1129, Dec. 2007.

[65] R. R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell, MA: Kluwer, 1997.

[66] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.

[67] D. Zhang, S. Chen, and Z. H. Zhou, "Constraint score: A new filter method for feature selection with pairwise constraints," *Pattern Recognit.*, vol. 41, no. 5, pp. 1440–1451, May 2008.

[68] J. Zhao, K. Lu, and X. He, "Locality sensitive semi-supervised feature selection," *Neurocomputing*, vol. 71, no. 10–12, pp. 1842–1849, Jun. 2008.

Tossapon Boongoen received the Ph.D. degree in artificial intelligence from Cranfield University, MK43 0AL Cranfield, U.K.

He was a Lecturer with the Royal Thai Air Force Academy, Thailand. He is currently a Postdoctoral Research Associate with the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. His research interests include data mining, pattern recognition, and natural language processing. His current research focuses on link analysis and fuzzy aggregation, and their applications to effective intelligence modeling and reasoning for antiterrorism.

Qiang Shen received the Ph.D. degree in knowledge-based systems from Heriot-Watt University, Edinburgh, U.K.

He holds the established Chair in Computer Science and is the Director of Research with the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. He is also an Honorary Fellow with the University of Edinburgh, Edinburgh, U.K. He has authored two research monographs and over 250 peer-reviewed papers, including one that received an Outstanding Transactions Paper Award from the IEEE. His research interests include computational intelligence, fuzzy and qualitative modeling, reasoning under uncertainty, pattern recognition, data mining, and real-world applications of such techniques for decision support (e.g., crime detection, consumer profiling, systems monitoring, and medical diagnosis).

Dr. Shen is currently an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B AND THE IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is also an editorial board member of several other leading international journals. He has chaired and given keynote lectures at many international conferences.