**Aberystwyth University**

*Clus-DOWA: A New Dependent OWA Operator*
Boongoen, Tossapon; Shen, Qiang

# Clus-DOWA: A New Dependent OWA Operator

Tossapon Boongoen and Qiang Shen

*Abstract*— **Aggregation operators are crucial to integrating diverse decision makers' opinion. While minimum and maximum can represent optimistic and pessimistic extremes, an Ordered Weighted Aggregation (OWA) operator is able to reflect varied human attitudes lying between the two using distinct weight vectors. Several weight determination techniques ignore characteristics of data being aggregated. In contrary, data-oriented operators like centered OWA and dependent OWA utilize the centralized data structure to generate reliable weights. Values near the center of a group receive higher weights than those further away. Despite its general applicability, this perspective entirely neglects any local data structures representing strong agreements or consensus. This paper presents a new dependent OWA operator (Clus-DOWA) that applies distributed structure of data or data clusters to determine its weight vector. The reliability of weights created by DOWA and Clus-DOWA operators are experimentally compared in the tasks of classification and unsupervised feature selection.**

## I. INTRODUCTION

To aggregate valuable pieces of information, several 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 like minimum/maximum and t-norm/t-conorm (more details in [1]). In addition, Yager [17] introduced a parameterized mean-like aggregation operator, an ordered weighted aggregation (OWA) operator. Essentially, by selecting an appropriate weight vector, the OWA operators can reflect the uncertain nature of human judgment with the ability to generate an aggregating result lying between two extremes of minimum and maximum. The OWA operators have been applied in different areas [18] such as: information fusion [19], multi-criteria decision making [17] and fuzzy system modelling [20]. Furthermore, several variations of OWA operator are invented for linguistic-oriented environment [8] [12], while others are applicable to a wide range of domains using different weight determining algorithms: maximum-entropy [11], Gaussian distribution [15] [21], recursive formulation [13], and weight learning [6]. Recently, Beliakov et al. [2] emphasized another variant utilizing the absorbent tuples to model situations in which certain decision makers may decide the outcome irrespective of the opinion of the others.

A definite pre-caution, in combining multiple arguments, is the fact that unduly high or low values might be given by false or biased judgments. In such cases, a typical OWA operator would suffer drastically from giving the highest priority to either the highest or the lowest value. To achieve more reliable outcome, Xu [15] [16] introduced the dependent

OWA (DOWA) operator, in which the normal-distribution of argument values is used to determine the weight vector. In particular, a high weight is given to the argument whose value is close to the center of all arguments (i.e. arithmetic mean), while lower weights are assigned to those further away. This centralized interpretation has also been adopted in the centered OWA operator [21], where weights are high around the middle and decayed symmetrically towards boundary ends.

Despite their general applicability, these trustworthy weight generation methods possess an identical drawback, which originates from their underlying centralized assumption. Particularly, argument values are viewed as members of one large cluster (i.e. a global consensus of decision makers' opinion) and their arithmetic mean is considered sufficient to grade their reliability. This approach completely discards the significance of any possible trends emerging from local data structure that can possibly be a subset of values tightly clustered together. Regardless of the cluster density, low weights are to be assigned to its members if it does not situate near the center of the value range. In light of such shortcoming, this paper presents a new cluster-based DOWA operator (Clus-DOWA) whose weight determination is based on the distributed structural interpretation of values being aggregated. Those values very far from the group center (i.e. mean) are not assigned with low weights, if they are seemingly indifferent to their local neighbors. For this purpose, the basic technique of agglomerative hierarchical clustering [5] is applied to create the clustering structure of the studied values. In essence, the distance to the nearest cluster is employed to evaluate the reliability of each argument value and its assigned weight.

The rest of this paper is organized as follows. Section II introduces the main theoretical concepts of the OWA and DOWA operators upon which the present research is developed. Section III presents the clustering-based DOWA operator including its complexity and a worked example. Applications of the DOWA and Clus-DOWA operators to classification and unsupervised feature selection tasks are detailed in the fourth section. Specifically, class-specific fuzzy sets built from weight vectors are used to determine the class to which an unknown instance belongs. In addition, weight vectors are also utilized to justify the reliability of attributes in order to reduce the size of a feature set. The paper is concluded in section V, with the perspective of further works.

## II. DEPENDENT OWA OPERATOR

The process of information aggregation appears in many applications related to the development of decision support

Tossapon Boongoen and Qiang Shen are with the Department of Computer Science, Aberystwyth University, UK (phone: +44 1970 621787; email: {tsb,qqs}@aber.ac.uk).

systems. Despite computationally simplistic, neither minimum nor maximum is appropriate for most applications. Accordingly, Yager [17] pioneered a new set of aggregation techniques called the ordered weighted averaging (OWA) operator. This mean-type operator provides a flexibility to utilize the entire range of *and* to *or* associated with a decision maker's attitude towards aggregation.

### A. OWA Operator

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 $\sum_{j=1}^{n} w_j = 1$. A set of $n$ input parameters, with the input vector $(B_1, B_2, \ldots, B_n)$, is aggregated as follows:

$$OWA(B_1, B_2, \ldots, B_n) = \sum_{j=1}^{n} w_j b_j \qquad (1)$$

where $b_j$ is the jth largest element in the vector $(B_1, B_2, \ldots, B_n)$ and $b_1 \geq b_2 \ldots \geq b_n$. A wide range of OWA operators can be formulated between two extremes of minimum and maximum, through the degree of $orness(\alpha)$ as defined below:

$$\alpha = \frac{1}{n-1} \sum_{i=1}^{n} w_i (n-i) \qquad (2)$$

This measure ranges from 0 to 1, and estimates the degree to which an OWA operator is similar to the logical connective OR (i.e. $maximum$) in terms of its aggregation behavior. When $\alpha = 0$, an OWA operator becomes a minimum type with the weight vector becomes $(0, 0, \ldots, 1)$. In contrast, when $\alpha = 1$, an OWA operator exhibits the maximum connective type with the weight vector of $(1, 0, \ldots, 0)$. Weight distributions at any given degree of orness can be differentiated with the dispersion measurement:

$$disp(w) = -\sum_{i=1}^{n} w_i ln(w_i) \qquad (3)$$

where $0 \leq disp(w) \leq ln(n)$. This measure provides a degree to which information in arguments is used in the aggregation process. The value of dispersion is zero when the orness $(\alpha)$ is either zero or one. In such cases, only one argument with either the lowest or the highest will contribute solely to the ultimate outcome. In contrast, the dispersion is unity when the orness is 0.5, in which all arguments are considered equally important with assigned weight of $1/n$.

### B. Dependent OWA Operator

There are many techniques proposed for obtaining OWA weights. For instance, O'Hagan [11] used the maximal entropy as the primary criteria in formulating a set of weights at a given level of orness. Attempts also exist to generate a weight vector from a set of samples [6]. Weight determining approaches generally fall into two categories of *argument-independent* and *argument-dependent*. Weights derived by the former approach are not related to argument

values. On the other hand, with an argument-dependent approach, weights are determined based on the value of input arguments. Centered OWA [21] and Dependent OWA [15] [16] operators are examples of argument-dependent approaches that particularly employ the centralized weight distribution. Arguments whose values are in the middle of the group, i.e. near to the group average, are more reliable and acquire higher weights comparing to those further away from the center. Specifically, the reliability of an argument reflects the appropriateness of using that argument as a group representative (i.e. aggregating outcome). The following set of equations give an overview of the weight determination process with the DOWA operator.

Let $(a_1, a_2, \ldots, a_n)$ be the argument vector, and $\mu$ be the average value of this argument set, where $\mu = \frac{1}{n} \sum_{j=1}^{n} a_j$. The similarity between any argument $a_j$ and the average value $\mu$ can be calculated as follows:

$$s(a_j, \mu) = 1 - \frac{|a_j - \mu|}{\sum_{i=1}^{n} |a_i - \mu|} \qquad (4)$$

From this, a weight vector $w = (w_1, w_2, \ldots, w_n)^T$ can be generated by applying the following:

$$w_j = \frac{s(a_j, \mu)}{\sum_{i=1}^{n} s(a_i, \mu)}, j = 1, 2, \ldots, n \qquad (5)$$

$$DOWA(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} w_i a_i \qquad (6)$$

The DOWA operator becomes neat in such a way that one weight is provided for a specific argument, regardless of its order of magnitude in the argument vector. Hence, the reordering step, normally pursued for an OWA operator, is entirely irrelevant in this case.

### III. CLUSTER-BASED DOWA OPERATOR

Similar to Centered OWA and DOWA operators, the proposed Clus-DOWA aims to decrease the effect of false or biased judgment in a group decision making. The former approaches utilize the centralized method in such a way that the argument set is viewed as one cluster, whose center is solely used to determine the weight vector. In contrary, the Clus-DOWA operator is based on the distributed clusters of arguments. Instead of interpreting a set of arguments as one large cluster, it is trivial to perceive that local clusters (i.e. local consensus) can appear within the global space. In particular, each local cluster represents agreement or consensus among arguments in close proximity. Intuitively, to evaluate the reliability of one argument is to discover the difference (in terms of distance) between that particular argument and its nearest local cluster. The magnitude of difference from surrounding neighbor dictates the difficulty that one argument can come to agreement with others. Conceptually, greater difference signifies greater difficulty and hence smaller reliability. Reliable arguments are those with small differences from their neighbors. Figure 1 graphically depicts this distributed approach, in which arguments ($a1$ and

$a2$) very far from the global center are reliable provided that they are close to local clusters' center.
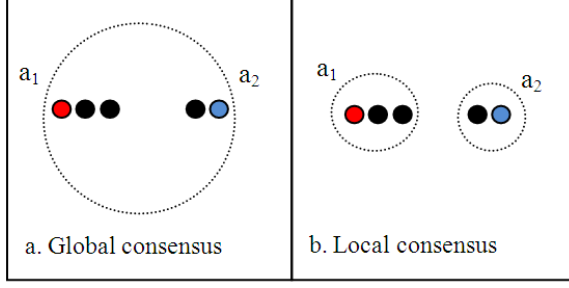


Fig. 1. Centralized and distributed interpretations.

### A. Cluster-Based Algorithm

With this distributed approach, the first task of measuring reliability of arguments is to determine their cluster structure. The agglomerative hierarchical clustering [5] is modified in such a way that its iterative merging process stops as soon as all arguments ($a_i, i = 1, \ldots, n$) have been merged to their nearest clusters. For each argument, the distance ($d_i$) behind such merging as well as the cluster center ($Reference_i$) are recorded for the evaluation of its reliability. With the center-linkage distance measurement (i.e. using the average value of a group as the representative in a distance evaluation), this modified clustering algorithm can be summarized in Figure 2. Note that obviously, a singleton cluster contains only one member.

```
Begin
  Initialize a = n; c = 0; d_i = { }; Reference_i = { }; i = 1 ... n
  Do
  Find nearest elements (argument or cluster) E_x and E_y
    If E_x and E_y are arguments a_x and a_y
      d_x = distance(a_x, a_y); Reference_x = a_y;
      d_y = distance(a_y, a_x); Reference_y = a_x; a = a − 2
    If E_x is argument a_x and E_y is cluster Cl_y
      d_x = distance(a_x, center(Cl_y)); Reference_x = center(Cl_y); a = a − 1
    If E_x is cluster Cl_x and E_y is argument a_y
      d_y = distance(a_y, center(Cl_x)); Reference_y = center(Cl_x); a = a − 1
    Merge E_x and E_y to a new cluster Cl_c, where c = c + 1
    Drop E_x and E_y, and include Cl_c in next iteration
  Until a = 0
  Return d_i, Reference_i, i = 1 ... n
End
```

Fig. 2. Modified agglomerative hierarchical clustering.

Having applied this clustering algorithm to a set of values ($a_1, a_2, \ldots, a_n$), the reliability of each value ($r_i$) can be directly estimated from the distance to its nearest cluster ($d_i$) recorded during the clustering process.

$$r_i = 1 - \frac{d_i}{\sum_{j=1}^{n} d_j} \qquad (7)$$

Similar to Equation (5), the weight vector can then be calculated from the discovered vector of reliability measurement ($r_1, r_2, \ldots, r_n$) as follows:

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j}, i = 1, 2, \ldots, n \qquad (8)$$

Then, the Clus-DOWA operator can be defined as:

$$Clus - DOWA(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j a_j \qquad (9)$$

It is worth noting that with the clustering procedure, the Clus-DOWA operator is fairly computational expensive with the time and space complexity of $O(n^3)$ and $O(n^2)$, where $n$ is the number of arguments. Both requirements are reduced to $O(n)$ using the simple DOWA operator. Despite this disadvantage, weight vectors generated by the Clus-DOWA operator are truly data-oriented and reliable, which will be illustrated later in Section IV.

### B. Worked Example

For illustrating purpose, both Clus-DOWA and DOWA operators are used to aggregate preference values of ten decision makers ($a_1, \ldots, a_{10}$): 60, 62, 63, 66, 70, 75, 79, 85, 89, and 94. With the DOWA operator and the global average of 74.3, the following set of weights can be achieved using Equations (4) and (5): $w_1 = 0.0953, w_2 = 0.0975, w_3 = 0.0986, w_4 = 0.1019, w_5 = 0.1063, w_6 = 0.1103, w_7 = 0.1059, w_8 = 0.0993, w_9 = 0.0949$ and $w_{10} = 0.0894$. With Equation (6), the aggregation outcome is:

$$
\begin{aligned}
DOWA(a_1, \ldots, a_{10}) =& (60 \times 0.0953) + (62 \times 0.0975) + \\
& (63 \times 0.0986) + (66 \times 0.1019) + \\
& (70 \times 0.1063) + (75 \times 0.1103) + \\
& (79 \times 0.1059) + (85 \times 0.0993) + \\
& (89 \times 0.0949) + (94 \times 0.0894) +
\end{aligned}
$$

$$DOWA(a_1, \ldots, a_{10}) = 74.1125$$

In order to find the result of this example with the Clus-DOWA operator, the clustering algorithm shown in Figure 2 is firstly applied to these preference values using the center-linkage distance measurement. Figure 3 presents the distance and nearest cluster's center of each value after being clustered. According to Equations (7) and (8), the following set of weights can be estimated: $w_1 = 0.1033, w_2 = 0.1080, w_3 = 0.1080, w_4 = 0.0985, w_5 = 0.0985, w_6 = 0.0985, w_7 = 0.0985, w_8 = 0.0985, w_9 = 0.0985$ and $w_{10} = 0.0892$. With Equation (9), the aggregation result is the summation of:

$$
\begin{aligned}
Clus - DOWA(a_1, \ldots, a_{10}) =& (60 \times 0.1033) + \\
& (62 \times 0.1080) + (63 \times 0.1080) + (66 \times 0.0985) + \\
& (70 \times 0.0985) + (75 \times 0.0985) + (79 \times 0.0985) + \\
& (85 \times 0.0985) + (89 \times 0.0985) + (94 \times 0.0892)
\end{aligned}
$$

$$Clus - DOWA(a_1, \ldots, a_{10}) = 73.8263$$

It is obviously depicted in Figure 4 that weights given by the DOWA approach are high in the middle and decayed towards both ends. On the other hand, the weights estimated using the Clus-DOWA method reflect the closeness of each value to its neighbors, hence the difficulty to agree with others. Evidently, the weights given to the values 60, 62 and 63 are higher with the Clus-DOWA than the DOWA approach, due to the fact that they are members of the rigid local cluster and their distances to agreement ($d_i$) are small comparing to those of the others. In contrast, the value 94 is assigned with similarly lowest weights by both methods, simply because it is greatly different from neighbors with $d_i = 7$, the highest of the entire group. Note that those values in the middle (70, 75 and 79), which are normally assigned with high weights by the DOWA approach, do not necessarily receive the same treatment with the new method.
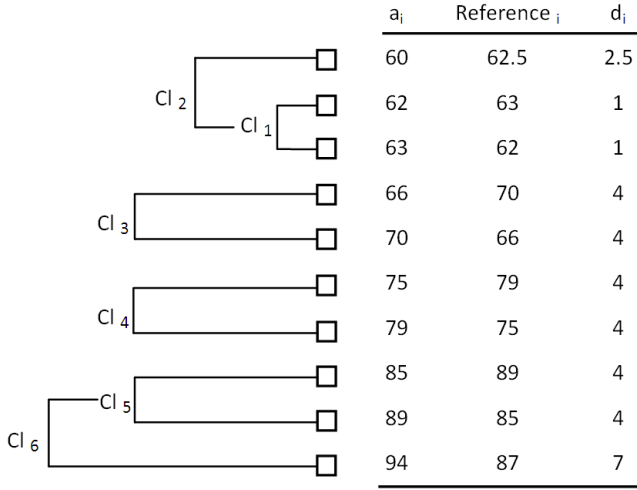


| $a_i$ | Reference $_i$ | $d_i$ |
|---|---|---|
| 60 | 62.5 | 2.5 |
| 62 | 63 | 1 |
| 63 | 62 | 1 |
| 66 | 70 | 4 |
| 70 | 66 | 4 |
| 75 | 79 | 4 |
| 79 | 75 | 4 |
| 85 | 89 | 4 |
| 89 | 85 | 4 |
| 94 | 87 | 7 |

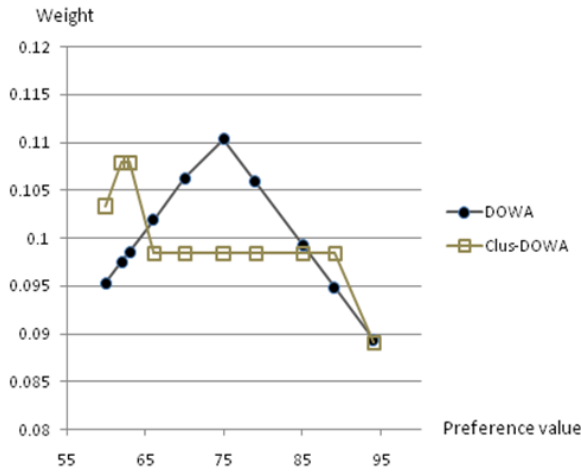Fig. 3.   Results of center-linkage clustering.



Fig. 4.   Weight distribution with DOWA and Clus-DOWA operators.

## IV. APPLICATION STUDIES

This section presents the experimental evaluation of applying the DOWA and Clus-DOWA operators to the tasks of classification and unsupervised feature selection. Their performances are generalized with a collection of well-known datasets obtained from [3].

### A. Classification

This task is to find the correct class, from $n$ possible classes ($C_j, j = 1 \ldots n$), for an unclassified instance that is characterized by $m$ attributes ($a_i, i = 1 \ldots m$). The DOWA operator can be used for the classification by firstly building fuzzy sets of classes in each attribute domain from training instances:

- Values of each attribute in training instances are divided into subsets each belonging to one class.
- The DOWA operator is then applied to these subsets to generate weight vectors.
- The resulting weight vectors are normalized into the range of [0,1] using $(w_i - w_{min})/(w_{max} - w_{min})$, where $w_i$ is any weight, $w_{max}$ and $w_{min}$ are the maximum and minimum weight values in one vector. Essentially, weight vectors are now interpreted as fuzzy sets representing properties of one specific attribute, see Figure 5 for examples.

For each attribute value ($a_i^k, i = 1 \ldots m$) of a given unclassified instance $k$, it is possible to use the membership vectors similar to those shown in Figure 5(b) to linearly estimate its membership values belonging to each class, $\mu_{C_j}(a_i^k), j = 1 \ldots n$. The membership degree is: (i) zero if the attribute value is outside the range of values presented in training instances, (ii) $\mu_{C_j}(a_i^t)$ when $t$ is a training instance and $a_i^k = a_i^t$, (iii) the linear approximation of membership values of two values $a_i^g$ and $a_i^h$ from training instances that are the most similar to $a_i^k$ and $a_i^g < a_i^k < a_i^h$.

$$\mu_{C_j}(a_i^k) = \left( \frac{|a_i^h - a_i^k|}{|a_i^g - a_i^h|} \times \mu_{C_j}(a_i^g) \right) + \left( \frac{|a_i^g - a_i^k|}{|a_i^g - a_i^h|} \times \mu_{C_j}(a_i^h) \right) \quad (10)$$

Then, the total membership value $\mu_{C_j}$ of each class $j$ can be calculated, as shown below. Ultimately, the given instance will be classified to the class with the highest total membership value.

$$\mu_{C_j} = \sum_{i=1}^{m} \mu_{C_j}(a_i^k) \quad (11)$$

This classification process can be used with the Clus-DOWA operator as well, however with additional adjustment. Firstly, both weight and membership functions are similarly built using the Clus-DOWA operator. Then, class membership values $\mu_{C_j}(a_i^k)$ of the given unclassified instance $k$ are identically estimated for cases (i) and (ii), as previously discussed. Nonetheless, case (iii) and Equation (10) cannot

be directly reused, since the weight and membership values are cluster oriented and unnecessarily decayed to both ends like those generated by the DOWA operator. In order to overcome this native barrier, it is necessary to find an attribute value $a_i^f$ presented in training instances that is the most similar to $a_i^k$. The cluster center $Reference_i^f$, to which the attribute value $a_i^f$ is merged, can be used to calculate $\mu_{C_j}(a_i^k)$ as follows:

$$\mu_{C_j}(a_i^k) = \frac{a_i^f - Reference_i^f}{a_i^k - Reference_i^f} \times \mu_{C_j}(a_i^f) \qquad (12)$$

Equation (11) is employed here again to find the class with the highest total membership value, which will be the preferred class for the given instance $k$.

Both classification methods are evaluated with the three benchmark datasets of UCI repository [1]: glass, iris and wine, respectively. As discussed earlier that the complexity of the Clus-DOWA algorithm is worse than the DOWA approach, the evaluation therefore concentrates solely on the accuracy criterion to further justify their quality. Table I presents average accuracies of classifying these dataset with DOWA and Clus-DOWA methods, using 10-fold cross validation. The performance of the Clus-DOWA approach is consistently superior to that of the other. This explicitly indicates that weights generated with the Clus-DOWA operator are more reliable, however with more expensive computation.

TABLE I
CLASSIFICATION ACCURACIES

| Dataset | Average accuracies (percentage) | |
| | DOWA | Clus-DOWA |
| --- | --- | --- |
| glass | 45.45 | 59.09 |
| iris | 85.71 | 90.47 |
| wine | 61.11 | 77.78 |

### B. Unsupervised Feature Selection

To further justify the reliability and also to reveal the usefulness of weights generated by these two operators, they will be used for the task of feature selection that aims to reduce a number of features (i.e. attributes) for more efficient data analysis. The benefits of this work includes: reducing the measurement and storage requirements, reducing training time, and defying the curse of dimensionality to improve prediction performance. Much of the work in feature selection has been concentrated at supervised category where invented methods rely on the class labels and their correlation with feature values. Guyon and Elisseeff [7] point out that the unsupervised feature selection techniques prove to be extremely useful with real-world data analysis, in which neither class labels are available nor thorough data interpretation is feasible. Works in this category base their judgments on particular characteristics of data values such as entropy and density. Dash and Liu [4] specifically emphasize that the entropy is generally low for data containing tight clusters, and thus is a good criterion to determine feature relevance.
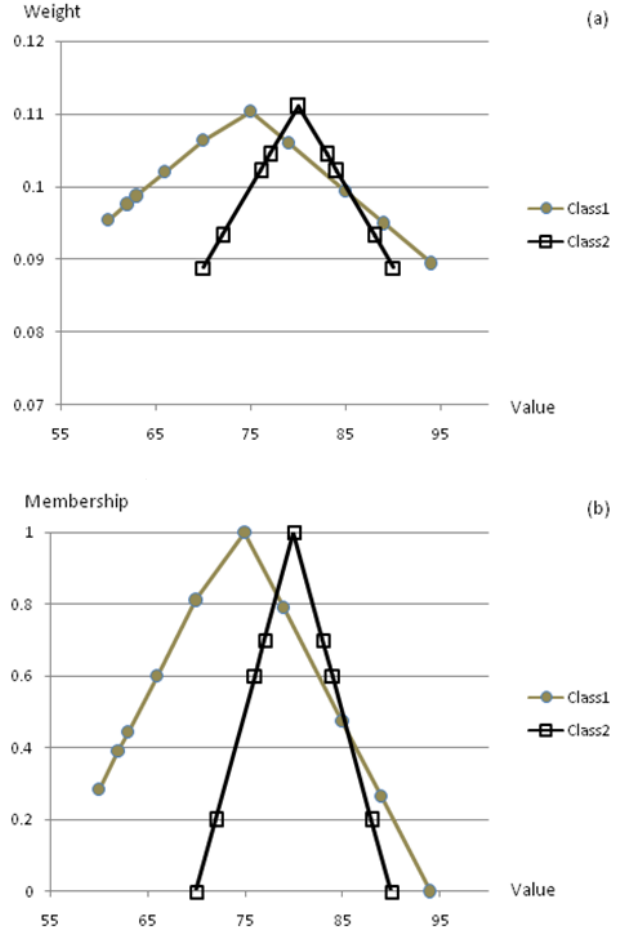


Fig. 5. Representing classes in one attribute domain as (a) weight and (b) membership vectors, using DOWA operator.

Intuitively, the dispersion measure, (see Equation (3)), of weight vectors generated by DOWA and Clus-DOWA operators can be used as evaluating criteria. This is because the higher the dispersion value the more reliable data becomes, as reflected by the fact that data values are reliable when they are close to neighbors (i.e. forming a tight cluster) and their weights are rather indifferent. To estimate the reliability of each feature $(a_i, i = 1 \ldots m)$, DOWA and Clus-DOWA operators are applied to generate weight vectors $(w_i^{DOWA}$ and $w_i^{Clus-DOWA})$ from its values. Then, the dispersion measurements of these weight vectors, $disp(w_i^{DOWA})$ and $disp(w_i^{Clus-DOWA})$, can be found using Equation (3). Note that the most reliable feature has the highest dispersion value of $ln(n)$, where n is the number of dataset instances. A simple heuristic-based algorithm, outlined in Figure 6, is used to deliver a subset of features whose dispersion measurement (i.e. reliability) is relatively competitive with that of the original feature set.

To evaluate performances of both operators upon this task, the dispersion-oriented algorithm shown in Figure 6 is firstly applied to find reduced feature sets of glass, iris and wine dataset, respectively. Table II presents the size of

these reduced feature sets, compared to those reported in [9] [10] using FRFS (Fuzzy-Rough Feature Selection) method. Note that FRFS is a supervised technique with state-of-the-art theory as well as outcomes.

Then reduced feature sets are assessed with three different learning classifiers: J48, JRip and PART (from [14]). J48 generates decision trees by choosing the most informative features and recursively partitioning the data into subtables based on their values. Each node in the tree represents a feature with branches from a node representing the alternative values this feature can take in accordance to the current subtable. Partitioning stops when all data items in the subtable have the same classification. A leaf node is then created, and this classification assigned. JRip learns propositional rules by repeatedly growing rules and pruning them. During the growth phase, features are added greedily until a termination condition is satisfied. Features are then pruned in the next phase subject to a pruning metric. Once the ruleset is generated, a further optimization is performed where classification rules are evaluated and deleted based on their performance on randomized data. PART generates rules by means of repeatedly creating partial decision trees from data. The algorithm adopts a divide-and-conquer strategy such that it removes instances covered by the current ruleset during processing. Essentially, a classification rule is created by building a pruned tree for the current set of instances; the leaf with the highest coverage is promoted to a rule.

Table III summarizes the accuracies of reduced feature sets with the aforementioned classifiers, using 10-fold cross validation. It can be generalized that the reduced feature sets achieved with the Clus-DOWA approach consistently reach better accuracy figures than those generated by the DOWA method. Accordingly, these results signify the fact that a weight vector of the Clus-DOWA operator is more reliable than the other. Moreover, the reliability of the Clus-DOWA operator can be further emphasized with its accuracy results comparing to those of the FRFS method, see Table IV. The Clus-DOWA method performs better than FRFS with glass and wine dataset, while the opposite outcome occurs with the iris dataset. However, it is crucial to note that the FRFS method is unable to reduce the size of feature set of iris dataset, while almost half of the original size is diminished using the Clus-DOWA approach.

TABLE II

SIZE OF REDUCED FEATURE SETS

| Dataset | Instance | Feature | Size of reduced set | | |
|---------|----------|---------|------|-----------|------|
| | | | DOWA | Clus-DOWA | FRFS |
| glass | 214 | 10 | 8 | 7 | 9 |
| iris | 150 | 5 | 3 | 3 | 5 |
| wine | 178 | 14 | 7 | 9 | 10 |

## V. CONCLUSIONS

This paper has presented a new dependent Ordered Weighted Aggregation (OWA) operator whose weight vector

```
Begin
  Initialize A = {a_i, i = 1 ... m}
  Do
    Initialize disp_max = 0; c = 1; a_unreliable
    Find the average dispersion, disp_avg, of all features in A
    Do
      Initialize disp_temp = 0
      Find the average dispersion of features in A, disp_temp, except feature a_c
      If disp_temp > disp_max
        disp_max = disp_temp; a_unreliable = a_c
      c = c + 1
    Until c > size of A
    If disp_max ≥ disp_avg
      A = A − a_unreliable
  Until disp_avg > disp_max
  Return A
End
```

Fig. 6.   Heuristic-based feature selection algorithm.

TABLE III

CLASSIFICATION ACCURACIES OF REDUCED FEATURE SET

| Dataset | Method | Classifier accuracies (percentage) | | |
|---------|--------|-------|-------|-------|
| | | J48 | JRip | PART |
| glass | Unreduced | 67.29 | 71.49 | 67.76 |
| | DOWA | 69.16 | 67.29 | 66.82 |
| | Clus-DOWA | 71.96 | 68.22 | 70.56 |
| iris | Unreduced | 96.00 | 95.33 | 94.00 |
| | DOWA | 72.67 | 74.00 | 73.33 |
| | Clus-DOWA | 96.00 | 92.67 | 95.33 |
| wine | Unreduced | 94.38 | 92.70 | 93.82 |
| | DOWA | 89.89 | 85.39 | 93.26 |
| | Clus-DOWA | 89.89 | 91.57 | 87.64 |

is tightly related to the structural characteristic of values being aggregated. Unlike the centralized assumption used in previous dependent OWA operators, the reliability of a value is not determined solely by its difference to the group average, but rather by its difference to neighbors. Values that are seemingly indifferent from others in a closed proximity are considered reliable and assigned with high weights. This distributed approach is able to better capture the underlying data characteristics and deliver trusty weights, which is experimentally illustrated through its superior performances comparing to the centralized method for both classification and unsupervised feature selection tasks. However, its applicability is to be further examined with other problem domains, especially decision making with multiple experts and criteria. Its complexity has to be improved by utilizing more efficient clustering techniques. In addition, the classification algorithm presented in this paper is to be further developed in such a way that its performance becomes competitive with well-known classification methods like decision tree and nearest neighbor.

TABLE IV

ACCURACIES OF FRFS AND CLUS-DOWA METHODS

| Dataset | Method | Classifier accuracies (percentage) | |
| --- | --- | --- | --- |
| | | JRip | PART |
| glass | FRFS | 67.76 | 68.22 |
| | Clus-DOWA | 68.22 | 70.56 |
| iris | FRFS | 95.33 | 94.00 |
| | Clus-DOWA | 92.67 | 95.33 |
| wine | FRFS | 89.33 | 93.82 |
| | Clus-DOWA | 91.57 | 87.64 |

## REFERENCES

[1] G. Beliakov, A. Pradera and T. Calvo, *Aggregation Functions: A Guide for Practitioners*, Springer: Heidelberg, Berlin, New York, 2007.

[2] G. Beliakov, T. Calvo and A. Pradera, "Absorbent tuples of aggregation operators," *Fuzzy Sets and Systems*, Vol. 158, No. 15 pp. 1675-1691. 2007.

[3] C. L. Blake and C. J. Merz. UCI Repository of machine learning databases. Irvine, University of California, 1998. http://www.ics.uci.edu/ mlearn/.

[4] M. Dash and H. Liu, "Unsupervised Feature Selection and Ranking," *New Trends in Knowledge Discovery for Business Information Systems*, Kluwer Publishers, 2000.

[5] M.B. Eisen, P.T. Spellman, P.O. Brown and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," In *Proceedings of National Academy of Sciences USA*, pp. 14863-14868. 1998.

[6] D. P. Filev and R. R. Yager, "On the issue of obtaining OWA operator weights," *Fuzzy Sets and Systems*, Vol. 94, pp. 157-169. 1998.

[7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182. 2003.

[8] F. Herrera, E. Herrera-Viedma and J.I. Verdegay, "Direct Approach Processes in Group Decision Making Using Linguistic OWA Operators," *Fuzzy Sets and Systems*, Vol. 79, pp. 175-190. 1996.

[9] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization," *Fuzzy Sets and Systems*, Vol. 141, pp. 469-485. 2004.

[10] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," To appear in *IEEE Transactions on Fuzzy Systems*.

[11] M. O'Hagan, "Aggregating template rule antecedents in real-time expert systems with fuzzy set logic," In *Proceedings of Annual IEEE Conference on Signals, Systems, and Computers*, pp.681-689. 1988.

[12] V. Torra, "The Weighted OWA Operator," *International Journal of Intelligent Systems*, Vol. 12, pp. 153-166. 1997.

[13] L. Troiano and R. R. Yager, "Recursive and Iterative OWA Operators," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 13, No. 6, pp. 579-599. 2005.

[14] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools with Java implementations*, Morgan Kaufmann Publishers, San Francisco, 2000.

[15] Z. S. Xu, "An overview of methods for determining OWA weight," *International Journal of Intelligent Systems*, Vol. 20, pp. 843-865, 2005.

[16] Z. S. Xu, "Dependent OWA operators," In *Proceedings of Modeling Decisions for Artificial Intelligence (MDAI2006)*, pp. 172-178. 2006.

[17] R. R. Yager, "Ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Transactions of System, Man and Cybernetics*, Vol. 18, pp. 183-190. 1988.

[18] R. R. Yager and J. Kacprzyk, *The ordered weighted averaging operators: Theory and Applications*, Kluwer Academic Publishers, Boston. 1997.

[19] R. R. Yager, "New Modes of OWA Information Fusion," *International Journal of Intelligence Systems*, Vol. 13, pp. 661-681. 1998.

[20] R. R. Yager, "Including Importances in OWA Aggregations Using Fuzzy Systems Modeling," *IEEE Transactions of Fuzzy Systems*, Vol. 6, No. 2. 1998.

[21] R. R. Yager, "Centered OWA operators," *Soft Computing*, Vol. 11, pp. 631-639. 2007.