Inteligencia Artificial, 25(69) (2022), 42-56
doi: 10.4114/intartif.vol25iss69pp42-56

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Rough-Fuzzy Support Vector Clustering with OWA Operators

Ramiro Saltos[1,*], Richard Weber[2]

1. Facultad de Innovación y Tecnología, Universidad Del Pacífico
Km. 7.5 Vía a la Costa, Guayaquil, Ecuador.
ramiro.saltos@upacifico.edu.ec

2. Department of Industrial Engineering, FCFM, Universidad de Chile
República 701, Santiago de Chile, Chile
richard.weber@uchile.cl

* Corresponding author: Ramiro Saltos

Abstract. Rough-Fuzzy Support Vector Clustering (RFSVC) is a novel soft computing derivative of the classical Support Vector Clustering (SVC) algorithm used successfully in many real-world applications. The strengths of RFSVC are its ability to handle arbitrary cluster shapes, identify the number of clusters, and effectively detect outliers by using the membership degrees. However, its current version uses only the closest support vector of each cluster to calculate outliers' membership degrees, neglecting important information that remaining support vectors can contribute. We present a novel approach based on the ordered weighted average (OWA) operator that aggregates information from all cluster representatives when computing final membership degrees and, at the same time, allows a better interpretation of the cluster structures found. Particularly, we propose the OWA using weights computed by the linguistic and exponential quantifiers. The computational experiments show that our approach obtains comparable results with the current version of RFSVC. However, the former weights all clusters' support vectors in the computation of membership degrees while maintaining their interpretability level for detecting outliers.

Keywords: Ordered Weighted Average; Support Vector Clustering; Uncertainty Modeling; Data Mining.

1 Introduction

Clustering is a well-known data mining task which has been studied extensively in the machine learning and statistics literature [9–11, 28, 32, 33, 46]. As one of the main steps of the Knowledge Discovery in Databases (KDD) process [8], it aims at finding groups of similar objects in a data set. These algorithms use a wide spectrum of search strategies in the data space or in a different feature space. Among the most currently used strategies are those techniques which are based on some kind of prototype representation, such as cluster centers in k-means approaches [9] or certain shell prototypes in shell clustering [41]. However, the growing availability of data according to the basic definition of Big Data (Volume, Variety, Velocity) [46], leads to more complex structures in data sets where the previously mentioned prototypes are not always the most appropriate alternatives for representing clusters.

Support Vector Clustering [3] and its recently introduced soft-computing derivative Rough-Fuzzy Support Vector Clustering (RFSVC) [35], offer an interesting alternative. In both algorithms, clusters

ISSN: 1137-3601 (print), 1988-3064 (on-line)

are represented by observations that are called *support vectors* allowing a more flexible representation of the groups of data. Instead of just one center, these groups are represented by several support vectors reflecting the spread within the cluster structure. SVC as well as RFSVC identify furthermore so-called *inside data points*; each one of these observations belongs completely to exactly one of the clusters found. Objects outside the clusters are called *bounded support vectors*.

Since uncertainty modeling is a relevant issue in clustering, numerous approaches such as fuzzy c-means [4], rough c-means [19, 27], shadowed c-means [24], and hybrid approaches like rough-fuzzy c-means [21] and rough-fuzzy support vector clustering [35], have been proposed in the literature. The main difference between these approaches and the classic (dichotomous) methods is that the latter assigns each object to one cluster whereas, in the former, this fact does not hold necessarily. For example, RFSVC determines the membership degrees for each bounded support vector to all clusters based on its distances to the closest support vector for each cluster. However, since multiple support vectors represent each cluster, important information on the clusters' shapes is lost. This approach is similar to Single Linkage Clustering [12, 37] in conventional hierarchical clustering in which the minimal distance between two clusters is taken into account when deciding which pair of clusters to merge. An alternative would be to consider the average distance to all support vectors, similar to Average Linkage Clustering [40].

In this paper, we propose a more holistic approach, taking into account all support vectors of each cluster at the same time when calculating an observation's membership degree to that cluster. The problem of aggregating information from more than one source is how to perform it. To solve this problem, we use the Ordered Weighted Average (OWA) operator [43] to consider all support vectors simultaneously in the calculation of membership degrees, thus taking all relevant information into account. Among all possible ways to determine the aggregation weights, we propose using the OWA based on linguistic and exponential quantifiers. This particular approach has the advantage that cluster shapes will be taken into account, thus providing a more meaningful interpretation of the results, as will be shown below.

The rest of the paper is arranged as follows: Section 2 provides an overview of the relevant literature. In Section 3 the proposed methodology for Rough-Fuzzy Support Vector Clustering using the OWA operator is presented. Its potential is shown in Section 4 in several computational experiments. In Section 5 we conclude our work and hint at possible future developments.

2 Literature Review

2.1 Ordered Weighted Average and Clustering Applications

Yager [43] proposed an ordered weighted average (OWA) operator which aggregates numbers coming from different sources of information. An OWA operator of dimension n is a mapping from $\mathcal{R}^n$ to $\mathcal{R}$ that has an associated weighting vector $\mathbf{w} \in \mathcal{R}^n$ such that $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$, and is given by:

$$OWA\left(a_1, \ldots, a_n, \mathbf{w}\right) = \sum_{j=1}^{n} w_j b_j \tag{1}$$

where b_j is the j-th largest a_j. A wide range of possible aggregation operators can be obtained when varying the weighting vector. The next ones are worth noting among others [7, 23]:

- If $w_1 = 1$ and $w_j = 0$ for all $j \neq 1$, the OWA becomes the maximum.
- If $w_n = 1$ and $w_j = 0$ for all $j \neq n$, the OWA becomes the minimum.
- If $w_j = \frac{1}{n}$ for all $j = 1, 2, \ldots, n$, we get the arithmetic mean or the simple average.
- The olympic OWA appears if $w_j = \frac{1}{n-2}$ for all $j \neq 1, n$ and $w_1 = w_n = 0$.

There are many approaches in the literature for obtaining the OWA weights [7, 18, 23]. A common approach is to use linguistic quantifiers [20]. The weights are generated by using a regular increasing monotone (RIM) function $Q\colon \mathcal{R} \to \mathcal{R}$ as follows:

$$w_j = Q\left(\frac{j}{n}\right) - Q\left(\frac{j-1}{n}\right) \qquad \forall j = 1, 2, \ldots, n \tag{2}$$

Equation (2) guarantees the weights generated accomplish $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$. The RIM quantifiers used in this study [20, 22] are:

- The basic linguistic quantifier:

$$Q(r) = r^{\alpha} \tag{3}$$

- The quadractic linguistic quantifier:

$$Q(r) = \frac{1}{1 - \alpha \cdot r^{0.5}} \tag{4}$$

- The exponential linguistic quantifier:

$$Q(r) = e^{-\alpha \cdot r} \tag{5}$$

- The trigonometric linguistic quantifier:

$$Q(r) = \arcsin(\alpha \cdot r) \tag{6}$$

where $\alpha > 0$. The OWA operator is monotonic, commutative, bounded, and idempotent [43]. To characterize the aggregation, there are several measures including the degree of orness-andness and the entropy of dispersion [14]. The degree of orness is formulated as follows:

$$\text{orness}(w) = \frac{1}{n-1} \sum_{j=1}^{n} (n-j) w_j \tag{7}$$

The orness measure was introduced in [43] as an appealing property (attitudinal-character) of the OWA operator. The orness quantifies the degree of disjunctive behavior of an aggregation operator. Disjunctive operators combine the values as an "or" operator, by which the aggregation result is high if some (at least one) values are high. On the other hand, conjunctive operators combine the values as an "and" operator so that the final result of aggregation is high, if and only if, all the individuals are high [14]. It is worth noting that the andness is the complement of orness, i.e.:

$$\text{andness}(w) = 1 - \text{orness}(w) \tag{8}$$

Finally, the entropy of dispersion is defined as:

$$E(w) = -\sum_{j=1}^{n} w_j \ln(w_j) \tag{9}$$

where $E(w)$ characterizes how uniformly the input values are being used in the OWA operator [18].

The OWA operator has been applied successfully in many areas such as engineering, medicine, and finance, among others [7, 42, 44]. Many works have recently been proposed in the literature using OWA in the context of clustering, for example [5, 6, 17, 25, 26, 30, 31]. Chakraborty and Chakraborty [6] proposed a new methodology for group decision making based on fuzzy clustering. The clustering algorithm is used to partition the opinion of the experts; then the OWA operator aggregates the clustered opinions to build a ranking. The method is applied in a flight simulator. Nasibov and Kandemir-Cavas [25] integrated the OWA operator within the context of hierarchical clustering to find the distances between clusters. The aggregation operator acts as a generalized case of single linkage, complete linkage, and average linkage algorithms successfully producing results in the case of clustering the phylogenetic tree of protein sequences.

Rahmanimanesh and Jalili [30] developed an anomaly detection method in cluster-based mobile ad hoc networks with an ad hoc on demand distance vector routing protocol. In this context, cluster members periodically send votes to the cluster head and the final decision on attack detection is carried out. In their methodology, an adaptive ordered weighted averaging (OWA) operator is used for aggregating the votes of cluster members in the cluster head to make the final decision. In other line, Ren et al. [31] developed a new customer satisfaction index applied to the tourist industry based on fuzzy clustering. The fuzzy c-means algorithm is used to construct clusters among the customers surveyed in order to obtain weights to build the final index. The OWA operator is used to aggregate the information extracted from clusters and to gain insights. A new class of aggregation operators called majority-clusters DOWA (MC-DOWA) operators, was proposed in [17]. The aim of these operators is to aggregate elements by classification to stress the majority clusters in the aggregation process. It was applied to improve the results in a group decision-making context over already classified data.

Finally, Cena and Gagolewski [5] generalized the well-known single, complete and average linkage schemes by emboding the expert knowledge in the cluster merge process using a three stage OWA operator. Their results show by robustifying the aggregation procedure with the Genie correction, they can obtain a significant performance boost in terms of clustering quality. In [26], based on users' interactions with social networks, the authors developed a method to understand users' life-styles using the OWA operator integrated with hierarchical clustering to find the similarity between users and clusters. Specifically, a two step measure was defined to compare and aggregate clusters. The descriptions of user's lifestyles were obtained from previously reported experiences on social network sites. Community forums can use this approach to identify which individuals on the platform with shared lifestyles are best suited to answer questions from a perspective similar to that of the individual asking the question.

Based on the literature reviewed, it can be concluded that almost all of the related clustering work is focused on using OWA in a pre-processing or post-processing step of the clustering task to obtain a final decision according to the situation studied. The OWA operator was not used to compute membership degrees since the majority of clustering approaches were center-based. The Support Vector Clustering algorithm [3] and its recently introduced soft-computing derivative Rough-Fuzzy SVC [35] report more than just one prototype per cluster found, leaving an open question on how to aggregate the prototype information to compute the final membership degrees. In this paper, we propose a methodology using OWA operators to answer this question.

2.2 Rough-Fuzzy Support Vector Clustering

In 2016, Saltos and Weber [35] introduced a new soft-computing version of the Support Vector Clustering algorithm called Rough-Fuzzy Support Vector Clustering (RFSVC), which is the basic clustering method of the approach introduced in this paper. The contribution made by RFSVC is constructing a rough-fuzzy partition of the dataset using Support Vector Domain Description (SVDD) [39], where outliers can be clearly identified and separated from the clusters found.

The RFSVC algorithm has three stages that will be explained below in more detail. First, there is a *training phase*, in which SVDD is used to obtain a hypersphere (in a higher-dimensional, projected feature space) that encloses most of the data points. All observations that fall outside its boundary are considered outliers. Then, in the subsequent *labeling phase* [3], different clusters within the set of data points enclosed by the hypersphere are distinguished. Finally, a *fuzzification phase* is performed over those objects that were classified as outliers in the first stage. The novelty of RFSVC lies in this step, in which each outlier gets membership degrees to every cluster found. A formal description of the three phases follows.

2.2.1 Training Phase

Let $X = \{\mathbf{x}_i \in \mathcal{R}^d / i = 1, \ldots, N\}$ be the set of N data points of dimension d. The first step projects the data to a reproducing kernel Hilbert space (RKHS), in which a hypersphere with minimal radius that encloses most of the training objects is constructed. The following quadratic optimization problem is solved:

$$Min_{R,\boldsymbol{a},\boldsymbol{\xi}} \; R^2 + C \sum_{i=1}^{N} \xi_i \tag{10}$$

$$s.t.$$

$$\| \phi(\mathbf{x}_i) - \boldsymbol{a} \|^2 \leq R^2 + \xi_i \qquad \forall i = 1, \ldots, N \tag{11}$$

$$\xi_i \geq 0 \qquad \forall i = 1, \ldots, N, \tag{12}$$

where R is the radius of the sphere and $\boldsymbol{a}$ its center; ϕ is a non-linear mapping; $\boldsymbol{\xi}$ is a set of slack variables used to allow some observations falling outside the hypersphere; $\| \cdot \|$ is the Euclidean norm; and $C \in [0, 1]$ is a constant regularization parameter that controls the trade-off between the volume of the sphere and the number of data points it includes. The dual formulation of the previous model is as follows:

$$Max_{\boldsymbol{\beta}} \sum_{i=1}^{N} \beta_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \beta_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{13}$$

$$s.t.$$

$$\sum_{i=1}^{N} \beta_i = 1 \tag{14}$$

$$0 \leq \beta_i \leq C \qquad \forall i = 1, \ldots, N, \tag{15}$$

where $\boldsymbol{\beta}$ are Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ is the kernel function. A widely used kernel function is the Gaussian kernel, which is given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-q\|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{16}$$

where $q > 0$ is a parameter that controls the kernel's width [36].

It can be shown that only objects i with $0 < \beta_i < C$ define the contours of the clusters [3], and are called *support vectors* (SV). Objects with $\beta_i = 0$ lie inside the hypersphere and are called *inside data points* (ID). Finally, objects with $\beta_i = C$ lie outside the hypersphere and are called *bounded support vectors* (BSV) or outliers.

The common understanding is that outliers are very different from the main clusters, they are treated separately. However, the bounded support vectors could lie close to the clusters. In order to be consistent with the literature (See [3, 34]), in this paper we also call them *outliers*.

For a given object $\mathbf{x}$, the distance between its projection and the center of the hypersphere, $\boldsymbol{a}$, can be calculated as:

$$\begin{aligned} R^2(\mathbf{x}) &= \| \phi(\mathbf{x}) - \boldsymbol{a} \|^2 \\ &= K(\mathbf{x}, \mathbf{x}) - 2\sum_{i=1}^{N} \beta_i K(\mathbf{x}_i, \mathbf{x}) + \sum_{i=1}^{N}\sum_{j=1}^{N} \beta_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j) \end{aligned} \tag{17}$$

The radius of the hypersphere follows:

$$R_S = \frac{1}{|SSV|} \sum_{\mathbf{x}_i \in SSV} R(\mathbf{x}_i) \tag{18}$$

where SSV is the set of Support Vectors and $|SSV|$ its cardinality.

The quadratic optimization problem of the training phase can be solved by many efficient algorithms. We used the Generalized Sequential Minimal Optimization (GSMO) algorithm, proposed in [13].

Fig. 1 illustrates a geometrical interpretation of the SVDD algorithm, in which Fig. 1(a) describes the objects' projection to a higher-dimensional space and the construction of the hypersphere, while in Fig. 1(b) the images of data points are projected back to the original space.

2.2.2 Labeling Phase

The training phase outputs the sets of support vectors, bounded support vectors, and inside data points. The main drawback is that many clusters may coexist within the hypersphere without being distinguished. Ben-Hur [3] proposed the following strategy to overcome this problem: Given two data points from different clusters, $\mathbf{x}_i$ and $\mathbf{x}_j$; any path that connects them must exit the hypersphere, i.e., $\exists \lambda \in [0, 1]$, such that $R(y_{i,j}) > R_S$, where $y_{i,j} = y_{i,j}(\lambda) = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{x}_j$. This leads to the definition of the

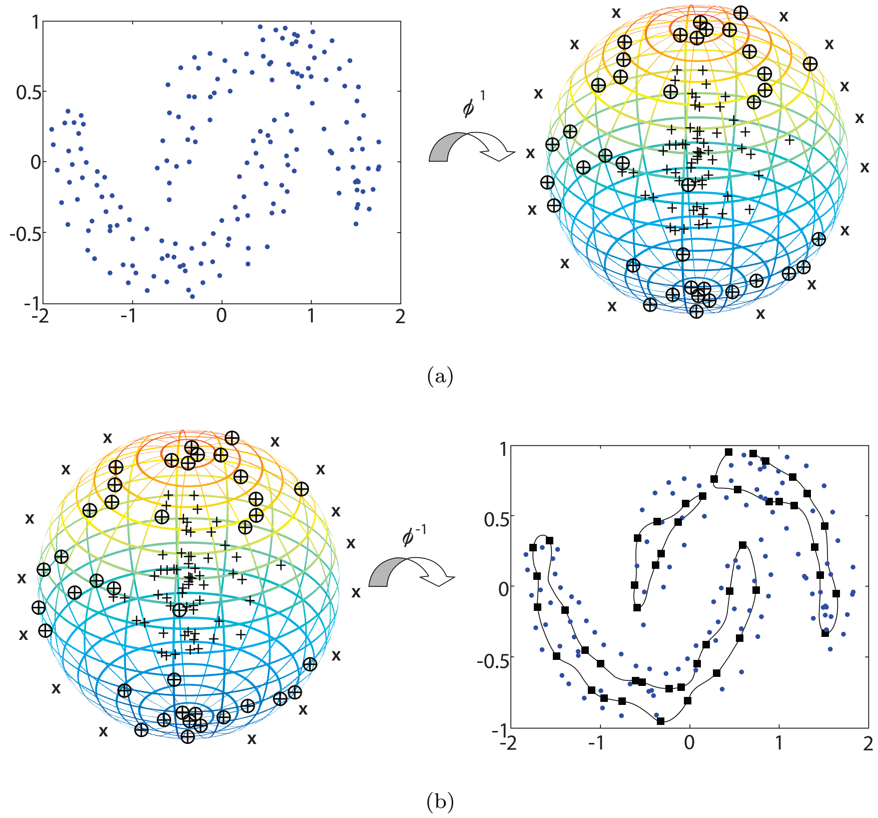

Figure 1: General Idea of RFSVC (+: Inside Data Points, ⊕: Support Vectors, **x**: Bounded Support Vectors). (a) Projection from Original Data Space to Higher-Dimensional Space. (b) Inverse Projection from Enclosing Sphere to Cluster Contours.

adjacency matrix A, whose elements $a_{i,j}$ represent whether or not a pair of points $\mathbf{x}_i$ and $\mathbf{x}_j$ belongs to the same cluster.

$$a_{i,j} = \begin{cases} 1, & \text{if } R(y_{i,j}) \leq R_S, \forall \lambda \in [0,1] \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

Clusters are now defined as the connected components of the graph induced by A. Note that the bounded support vectors remain unclassified since they lie outside the enclosing hypersphere.

This labeling rule is known as Support Vector Graph [3]. One drawback of this strategy is the high computational complexity [29]. As a consequence, more efficient labeling approaches have been proposed in the literature, e.g., Proximity Graph Modeling [45], Cone Clustering Labeling [15], and Fast and Stable Labeling [16], among others. The issue of cluster labeling is beyond the scope of our paper. We used the previously described Support Vector Graph approach in our experiments.

2.2.3 Fuzzification Phase

Saltos and Weber [35] introduced a fuzzification phase to calculate the membership degrees of bounded support vectors to the clusters created in the preceding phases. The strategy is:

1. Cast the hard cluster structure established in the training phase into a rough-fuzzy one based on two components: a lower approximation and a fuzzy boundary.

2. Assign the support vectors and inside data points to the lower approximations of their respective clusters according to the labeling phase.

3. Assign the bounded support vectors to the fuzzy boundaries of all clusters.

4. Calculate the membership degree $\mu_{i,k}$ of bounded support vector i to cluster k using the following equation:

$$\begin{aligned}\mu_{i,k} &= \mu(BSV_i, SV_{k,i}) = K(BSV_i, SV_{k,i}) \\ &= e^{-q\|BSV_i - SV_{k,i}\|^2}\end{aligned} \tag{20}$$

where $SV_{k,i}$ is the closest support vector in cluster k to the bounded support vector i.

Table 1 summarizes the outputs of each algorithm's phase, where $\mu_{i,k}$ is the membership degree of data point i to cluster k.

Table 1: Outputs of the RFSVC Algorithm

Phase	**Output**		
Training	Sets of SV, BSV, and ID		
Labeling	$\mu_{i,k} \in \{0,1\}$	$\forall i \in SV \cup ID$	$\forall k \in C$
	$\mu_{i,k} = 0$	$\forall i \in BSV$	$\forall k \in C$
Fuzzification	$\mu_{i,k} \in [0,1)$	$\forall i \in BSV$	$\forall k \in C$

A comprehensible application of this method using a two-dimensional synthetic data set can be found in [35].

3 Proposed Methodology for RFSVC using OWA Operators

As pointed out in Section 2.2.3, the membership degrees of bounded support vectors (BSV) are calculated using the closest support vector of each cluster at hand. Another option proposed in [35] is calculating the membership degrees using the average distance of BSV to the support vectors that define each cluster. These approaches have advantages and disadvantages. In the first case, the membership degrees offer a good level of interpretability, i.e., if this value is close to one, we know which outside data points are well represented by the current cluster structure while, if it is close to zero, we know which of them are outliers. However, the closest support vector approach does not take all the information provided about the cluster by the remaining support vectors into account. On the other hand, the second approach (i.e. average distance) uses all support vectors to compute the membership degrees of the bounded support vectors, however, the distance between support vectors could be "high" leading to small values of the membership degrees, thus losing interpretability.

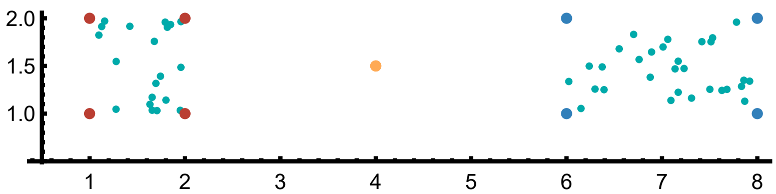

Figure 2: Example of Interpretability (Cluster 1 (left) is given by red SVs while Cluster 2 (right) by blue SVs)

For example, in Figure 2, the orange point (4;1.5) is a BSV whose distance to the nearest SV of both clusters is the same. Using the Closest Support Vector approach, the membership degree will be the same for both clusters, while using the average, it will be higher for Cluster 1 (left) than for Cluster 2 (right).

To overcome the limitations imposed by the closest support vector approach, we introduce a weighted aggregation method to include the information provided by all support vectors of each cluster in the computation of membership degrees. This method is based on the Ordered Weighted Average (OWA) operator [43] where each support vector can be weighted according to one of the following strategies:

- *Nearest Support Vector:* This approach is the same as reported in [35]. We set the weight of the nearest support vector to 1, and the remaining ones to 0.
- *RIM Quantifiers:* This approach computes the weights of each support vector according to the equations (3)-(6).
- *Average/Mean Approach:* This approach assigns equal weights to all support vectors.
- *Farthest Support Vector:* This is like the nearest support vector approach but instead of using the closest SV, it uses the farthest one.

The novel approach to computing the membership degrees is the second one, which induces each support vector's weight using the information provided by the RIM functions defined in Section 2.1. As a consequence, we do not need additional information to implement this method. The Algorithm 1 presents the new OWA-RFSVC.

Algorithm 1: OWA Rough-Fuzzy Support Vector Clustering

Input: Data set X, parameters $q > 0$ and $\upsilon \in (\frac{1}{N}, 1)$
Output: Rough-fuzzy clusters with $[0, 1]$-membership matrix and the number of clusters c
Run the training phase of the SVC algorithm and obtain the set of support vectors (SV), bounded support vectors (BSV), and inside data points (ID).
Run the labeling phase of the SVC algorithm and obtain the crisp cluster partition of the data set.
Assign support vectors and inside data points to the lower approximation of their respective clusters based on the labeling phase solution.
Assign bounded support vectors to the fuzzy boundaries of the clusters generated by the SVC algorithm.
Generate the distance matrix BSV vs. SV to obtain the distance of each data point that is outside of the sphere to each support vector.
Partition the distance matrix by columns according to the labeling phase.
for *each cluster* **do**
    Compute the sets of OWA weights using equations (3)-(6).
    for *each* $x_i \in BSV$ **do**
        for *each cluster* SV **do**
            Compute the preliminary membership degrees using equation (20).
        Compute the final membership degrees using equation (1).

The main advantage of the Algorithm 1 is that it computes a customized set of weights for each cluster, providing a more appropriate way of calculating membership degrees. Another advantage is it does not use external and subjective information since one of the main drawbacks of aggregation operators is the computation of the set of weights which is usually regarded as an external task of the decision-maker. Finally, the use of different OWA weighting methods does not alter the final data partition of RFSVC, but only the value of bounded support vectors' membership degrees.

4 Computational Experiments

In this section, we first introduce the data sets we used to test our procedure and explain how we calibrated the initial parameters. Then, in Section 4.2, we present the results obtained using OWA Rough-Fuzzy Support Vector Clustering and discuss the advantages of the proposed methodology.

4.1 Description of Data Sets and Experimental Set-Up

To compare the proposed approach with previous ones, we used the same data sets introduced in [35]. The main characteristics of the data sets are summarized in Table 2.

Table 2: Data Sets Characteristics

Name	**Type**	**Instances**	**Classes**	**Attributes**
Two Circles	Artificial	4500	2	2
Three Circles	Artificial	6750	3	2
Two Squares	Artificial	4500	2	2
Four Squares	Artificial	9000	4	2
XO	Benchmark	3600	2	2
XOOut	Benchmark	3603	2	2
S1-Gaussian	Benchmark	5000	15	2
Unbalance	Benchmark	6500	8	2
BankNote	Real World	1374	2	4
Glass	Real World	214	6	9
Cancer	Real World	569	2	30
Quake	Real World	2178	NA	4

We set the parameters for Rough-Fuzzy Support Vector Clustering according to [35]. Table 3 shows the parameter sets for each data set tested.

Table 3: Algorithms' Parameters

	RFSVC	
Name	$\boldsymbol{q}$	$\boldsymbol{v}$
Two Circles	5.7	1/3
Three Circles	5.7	1/3
Two Squares	5.44	1/3
Four Squares	8	1/3
XO	7	1/6
XOOut	7	1/6
S1-Gaussian	20	0.4
Unbalance	10	0.05
BankNote	0.25	0.1
Glass	0.1	0.1
Cancer	0.0001	0.20
Quake	1	0.2

The complete clustering results for the datasets used in this paper are available in [35]. All datasets used in this study and the results using the proposed approach can be downloaded from the following link: `https://goo.gl/FmJIAx`. Benchmark and real-world datasets can also be downloaded from the well-known data repositories [1, 2, 38].

4.2 Results

To compare the results obtained using different weights inside the OWA operator, we used Maji's validation indices [21]. In the equations presented below, Lw_j and FB_j are the lower approximation and fuzzy boundary of cluster j, respectively.

- α *index:* This index represents the average accuracy of the c clusters. It captures the average degree of completeness of knowledge about all clusters. A higher value of α indicates a better cluster solution. It is given by:

$$\alpha = \frac{1}{c}\left(\sum_{j=1}^{c}\frac{\sum_{x_i\in Lw_j} w(\mu_{i,j})^m}{\sum_{x_i\in Lw_j} w(\mu_{i,j})^m + \sum_{x_i\in FB_j}(1-w)(\mu_{i,j})^m}\right)$$

- α^* *index:* This index represents the accuracy of the approximation of all clusters. It captures the exactness of approximate clustering. A higher value of α^* indicates a better cluster solution. It is given by:

$$\alpha^* = \frac{\sum_{j=1}^{c} \sum_{x_i \in Lw_j} w(\mu_{i,j})^m}{\sum_{j=1}^{c} \left(\sum_{x_i \in Lw_j} w(\mu_{i,j})^m + \sum_{x_i \in FB_j} (1-w)(\mu_{i,j})^m \right)}$$

In both equations, the value of $w \in (0,1)$ represents the relative importance the lower approximation has compared to the fuzzy boundary. It can be noted that both indices take the value of 1 if, and only if, all the fuzzy boundaries are empty, i.e., the indices favor a crisp partition rather than a fuzzy one. Tables 4 and 5 present the results for Maji's validity indices with different OWA weighting mechanisms: closest support vector (CSV), linguistic quantifier (LQ), quadratic quantifier (QQ), exponential quantifier (EQ), trigonometric quantifier (TQ), average/mean approach (AVG), and farthest support vector (FSV).

Table 4: Results for Maji's α validity index

Instance	CSV	LQ	QQ	EQ	TQ	AVG	FSV
Two Circles	0.9383	0.9618	0.9926	0.9561	0.9982	0.9982	1.0000
Three Circles	0.9237	0.9561	0.9945	0.9519	0.9993	0.9993	1.0000
Two Squares	0.8968	0.9204	0.9683	0.9095	0.9864	0.9864	0.9988
Four Squares	0.9427	0.9631	0.9914	0.9538	0.9973	0.9973	0.9996
XO	0.9785	0.9941	0.9999	0.9975	1.0000	1.0000	1.0000
XO Out	0.9802	0.9944	0.9999	0.9973	1.0000	1.0000	1.0000
S1 Gaussian	0.9478	0.9875	0.9995	0.9778	0.9999	0.9999	1.0000
Unbalance	0.9945	0.9985	1.0000	0.9987	1.0000	1.0000	1.0000
BankNote	0.9827	0.9832	0.9849	0.9837	0.9864	0.9864	0.9933
Glass	0.9932	0.9933	0.9938	0.9934	0.9942	0.9942	0.9967
Cancer	0.9624	0.9624	0.9625	0.9625	0.9625	0.9625	0.9686
Quake	0.7191	0.7363	0.7930	0.7361	0.8415	0.8414	0.9429

Table 5: Results for Maji's α^* validity index

Instance	CSV	LQ	QQ	EQ	TQ	AVG	FSV
Two Circles	0.9382	0.9618	0.9926	0.9560	0.9982	0.9982	1.0000
Three Circles	0.9237	0.9561	0.9945	0.9519	0.9993	0.9993	1.0000
Two Squares	0.8966	0.9203	0.9683	0.9094	0.9864	0.9864	0.9988
Four Squares	0.9425	0.9631	0.9914	0.9538	0.9973	0.9973	0.9996
XO	0.9787	0.9942	0.9999	0.9976	1.0000	1.0000	1.0000
XO Out	0.9803	0.9944	0.9999	0.9973	1.0000	1.0000	1.0000
S1 Gaussian	0.9474	0.9875	0.9995	0.9778	0.9999	0.9999	1.0000
Unbalance	0.9992	0.9998	1.0000	0.9998	1.0000	1.0000	1.0000
BankNote	0.9928	0.9933	0.9951	0.9947	0.9966	0.9966	0.9998
Glass	0.9932	0.9933	0.9938	0.9934	0.9942	0.9942	0.9967
Cancer	0.9624	0.9624	0.9625	0.9625	0.9625	0.9625	0.9686
Quake	0.9677	0.9756	0.9906	0.9860	0.9958	0.9958	0.9992

As can be observed in Tables 4 and 5, the validity indices improve as we reduce the importance given to the closest support vectors. This is congruent with the nature of the validity indices since they favor crisp partitions. However, if membership degrees for bounded support vectors are close to zero, we lose information about the behavior of these possible outliers.

To control the membership degrees' loss of interpretability, we computed the orness measure for each set of OWA's weights. As pointed out in [14], the orness measures how much the OWA behaves as an

Table 6: Orness Measure for different OWA Weigthing Mechanisms

Instance	CSV	LQ	QQ	EQ	TQ	AVG	FSV
Two Circles	1.0000	0.9240	0.6872	0.9558	0.4995	0.5000	0.0000
Three Circles	1.0000	0.9219	0.6835	0.9454	0.4995	0.5000	0.0000
Two Squares	1.0000	0.9266	0.6920	0.9674	0.4995	0.5000	0.0000
Four Squares	1.0000	0.9278	0.6946	0.9712	0.4994	0.5000	0.0000
XO	1.0000	0.9148	0.6720	0.9095	0.4996	0.5000	0.0000
XO Out	1.0000	0.9152	0.6726	0.9114	0.4996	0.5000	0.0000
S1 Gaussian	1.0000	0.9254	0.6901	0.9607	0.4995	0.5000	0.0000
Unbalance	1.0000	0.9199	0.6799	0.9345	0.4995	0.5000	0.0000
BankNote	1.0000	0.9225	0.6865	0.9432	0.4995	0.5000	0.0000
Glass	1.0000	0.9153	0.6727	0.9118	0.4996	0.5000	0.0000
Cancer	1.0000	0.9108	0.6667	0.8936	0.4996	0.5000	0.0000
Quake	1.0000	0.9207	0.6823	0.9395	0.4995	0.5000	0.0000

"or" operator. Since we are interested in the BSV's membership degree being high if it is close to at least one SV and low if it is far from the SVs of all clusters, the set of weights must have an orness close to one. Table 6 shows the average orness of all clusters for each weighting mechanism and dataset. The best options to aggregate the information of all support vectors while computing the BSV's membership degrees are the linguistic (LQ) and the exponential (EQ) quantifiers. Both maintain an orness close to one, thus reducing the interpretability loss.

Since the values of the validity indices obtained using different weighting mechanisms are similar, we performed a one-factor ANOVA test to evaluate whether there is a statistically significant difference between these values. Tables 7 and 8 present the results.

Table 7: Analysis of Variance for α validity index

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Method	6	0.026824	0.004471	1.307366	0.264032
Residuals	77	0.263306	0.00342		

Considering a significance level $\alpha = 0.05$, given the p-value of the ANOVA test in Table 7, we do not have enough statistical evidence to reject the null hypothesis. So, there is no significant difference between the values of the α validity index among the weighting mechanisms. With these results, the best weighting mechanism is the farthest support vector. However, this method does not include all SV in the computation of the membership degrees and has the worst orness measure, losing the membership degrees' interpretability.

Table 8: Analysis of Variance for α^* validity index

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Method	6	0.014083	0.002347	5.99284	3.54E-05
Residuals	77	0.030157	0.000392		

In contrast with the results of Table 7, in Table 8, the p-value of the ANOVA test provides enough statistical evidence to reject the null hypothesis. So, there is a significant difference between the values of the α^* validity index among the weighting mechanisms. With these results, we perform a Tukey multiple comparison test to determine which weighting mechanisms differ from the others. The results are in Table 9.

The corresponding p-values of the weighting mechanisms that differ from the others are shown in **bold** in Table 9. The quadratic, trigonometric, average, and farthest weighting mechanisms are statistically

Table 9: Tukey Multiple Comparison Test for α^* validation index

Method	**Difference**	**Lower Bound**	**Upper Bound**	**p-Value**
LQ-CSV	0.014925	-0.009536	0.039386	0.520989
QQ-CSV	0.030450	0.005989	0.054911	***0.005678***
EQ-CSV	0.013125	-0.011336	0.037586	0.667097
TQ-CSV	0.033958	0.009497	0.058420	***0.001323***
AVG-CSV	0.033958	0.009497	0.058420	***0.001323***
FSV-CSV	0.036667	0.012205	0.061128	***0.000399***
QQ-LQ	0.015525	-0.008936	0.039986	0.472689
EQ-LQ	-0.001800	-0.026261	0.022661	0.999989
TQ-LQ	0.019033	-0.005428	0.043495	0.231797
AVG-LQ	0.019033	-0.005428	0.043495	0.231797
FSV-LQ	0.021742	-0.002720	0.046203	0.114472
EQ-QQ	-0.017325	-0.041786	0.007136	0.337922
TQ-QQ	0.003508	-0.020953	0.027970	0.999458
AVG-QQ	0.003508	-0.020953	0.027970	0.999458
FSV-QQ	0.006217	-0.018245	0.030678	0.987181
TQ-EQ	0.020833	-0.003628	0.045295	0.147080
AVG-EQ	0.020833	-0.003628	0.045295	0.147080
FSV-EQ	0.023542	-0.000920	0.048003	0.066995
AVG-TQ	0.000000	-0.024461	0.024461	1.000000
FSV-TQ	0.002708	-0.021753	0.027170	0.999879
FSV-AVG	0.002708	-0.021753	0.027170	0.999879

different from the closest support vector approach. However, the first ones have the lowest orness measures, thus reducing the membership degrees' interpretability. In contrast, the linguistic and exponential methods are not statistically different from the nearest support vector and, at the same time, maintain a high orness measure and interpretability.

Based on the computational and statistical results, the best options to aggregate all the support vectors in the computation of bounded support vectors' membership degrees are the linguistic and exponential weighting methods. These methods provide statistically similar validity indices results to the closest support vector approach. However, they maintain high levels of interpretability, as shown by the orness measures.

5 Conclusions and Future Work

In this paper, we proposed the OWA Rough Fuzzy Support Vector Clustering. OWA-RFSVC is a novel approach for computing the bounded support vectors' membership degrees considering all support vectors in each cluster. The main advantages of the method are:

1. It uses the information provided by all clusters' prototypes.
2. It maintains a high interpretability level of final membership degrees.
3. It does not require external information for the OWA weights generation.
4. The OWA weights are generated using the linguistic or exponential quantifiers, which only need the number of support vectors in each cluster.
5. There is no need to retrain the RFSVC model, which is usually computationally expensive.

We performed several computational experiments in a diverse set of data to evaluate the effectiveness of our approach. The results showed a statistically significant difference does not exist between the RFSVC and OWA-RFSVC when validating the final partition, which makes sense since OWA-RFSVC does not change the final clustering but only the BSVs' membership degrees. However, the orness measure showed not all weighting methods maintain the same interpretability level. In this case, the best options to aggregate all support vectors' information are the linguistic and exponential quantifiers.

Finally, one key issue in RFSVC is all bounded support vectors equally affect the center location of the hypersphere in the higher dimensional feature space. We are investigating the use of different OWA operators to reduce their influence to improve the robustness of RFSVC against noise. Additionally, the approach presented in this paper can be extended to other multi-prototype clustering algorithms like self-organizing maps.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions for improving the quality of the paper.

Support for both authors from the Institute on Complex Engineering Systems (`www.isci.cl`; ANID PIA/BASAL AFB180003) and Fondecyt (projects 1181036 and 1221562) is also gratefully acknowledged.

References

[1] J. Alcalá, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. Keel data mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(2-3):255–287, 2010.

[2] K. Bache and M. Lichman. UCI Machine Learning Repository, 2013.

[3] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik. Support Vector Clustering. *Journal of Machine Learning Research*, 2(12):125–137, 2001.

[4] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, 1981.

[5] A. Cena and M. Gagolewski. Genie+ OWA: Robustifying hierarchical clustering with OWA-based linkages. *Information Sciences*, 520:324–336, 2020.

[6] C. Chakraborty and D. Chakraborty. A fuzzy clustering methodology for linguistic opinions in group decision making. *Applied Soft Computing*, 7(3):858–869, 2007.

[7] O. Csiszar. Ordered weighted averaging operators: A short review. *IEEE Systems, Man, and Cybernetics Magazine*, 7(2):4–12, 2021.

[8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From Data Mining to Knowledge Discovery in Databases. *IEEE Transactions on Systems, Man and Cybernetics*, 17:37–54, 1996.

[9] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[10] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data.* Prentice-Hall, Inc., 1988.

[11] A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: A Review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[12] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[13] S. S. Keerthi and E. G. Gilbert. Convergence of a Generalized SMO Algorithm for SVM Classifier Design. *Machine Learning*, 46:351–360, 2002.

[14] A. Kishor, A. K. Singh, and N. R. Pal. Orness measure of owa operators: A new approach. *IEEE Transactions on Fuzzy Systems*, 22(4):1039–1045, 2013.

[15] S. H. Lee and K. M. Daniels. Gaussian kernel width exploration and cone cluster labeling for support vector clustering. *Pattern Analysis and Applications*, 15(3):327–344, 2012.

[16] H. Li. A Fast and Stable Cluster Labeling Method for Support Vector Clustering. *Journal of Computers*, 8(12):3251–3256, 2013.

[17] W.-w. Li, P.-t. Yi, and Y.-j. Guo. Majority Clusters-Density Ordered Weighting Averaging: A Family of New Aggregation Operators in Group Decision Making. *International Journal of Intelligent Systems*, 31(12):1166–1180, 2016.

[18] M. Lin, W. Xu, Z. Lin, and R. Chen. Determine owa operator weights using kernel density estimation. *Economic Research-Ekonomska istraživanja*, 33(1):1441–1464, 2020.

[19] P. Lingras and C. West. Interval set clustering of web users with rough k-means. *Journal of Intelligent Information Systems*, 23(1):5–16, 2004.

[20] P. Luukka and O. Kurama. Similarity classifier with ordered weighted averaging operators. *Expert Systems with Applications*, 40(4):995–1002, 2013.

[21] P. Maji and S. Pal. Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices. *IEEE Transactions on Systems, Man and Cybernetics - Part. B*, 37(6):1529–1540, 2007.

[22] S. Maldonado, J. Merigó, and J. Miranda. Redefining support vector machines with the ordered weighted average. *Knowledge-Based Systems*, 148:41–46, 2018.

[23] J. M. Merigó and A. M. Gil-Lafuente. The induced generalized owa operator. *Information Sciences*, 179(6):729–741, 2009.

[24] S. Mitra, W. Pedrycz, and B. Barman. Shadowed c-means: Integrating fuzzy and rough clustering. *Pattern Recognition*, 43(4):1282–1291, 2010.

[25] E. Nasıbov and C. Kandemır-Cavas. OWA-based linkage method in hierarchical clustering: Application on phylogenetic trees. *Expert Systems with Applications*, 38(10):12684–12690, 2011.

[26] J. Nguyen, A. Armisen, G. Sánchez-Hernández, M. Casabayó, and N. Agell. An OWA-based hierarchical clustering approach to understanding users' lifestyles. *Knowledge-Based Systems*, 190, 2020.

[27] G. Peters. Some refinements of rough k-means clustering. *Pattern Recognition*, 39(8):1481–1491, 2006.

[28] G. Peters, F. Crespo, P. Lingras, and R. Weber. Soft clustering - fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning*, 54(2):307–322, 2013.

[29] L. Ping, Z. Chun-Guang, and Z. Xu. Improved Support Vector Clustering. *Engineering Applications of Artificial Intelligence*, 23(4):552–559, 2010.

[30] M. Rahmanimanesh and S. Jalili. Adaptive ordered weighted averaging for anomaly detection in cluster-based mobile ad hoc networks. *Iranian Journal of Fuzzy Systems*, 10(2):83–109, 2013.

[31] J. Ren, Y. He, M. Hu, M. Jin, and H. Gao. A New Satisfaction Evaluation Method Based on Fuzzy Clustering. *International Journal of u-and e-Service, Science and Technology*, 8(9):31–40, 2015.

[32] L. Rokach. *Data Mining and Knowledge Discovery Handbook*, chapter A survey of Clustering Algorithms, pages 269–298. Springer US, Boston, MA, 2010.

[33] T. Sajana, C. S. Rani, and K. Narayana. A survey on clustering techniques for big data mining. *Indian Journal of Science and Technology*, 9(3), 2016.

[34] R. Saltos and R. Weber. Rough-Fuzzy Support Vector Domain Description for Outlier Detection. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, Istanbul - Turkey, 2015. IEEE.

[35] R. Saltos and R. Weber. A Rough-Fuzzy approach for Support Vector Clustering. *Information Sciences*, 339:353–368, 2016.

[36] B. Schölkopf. *Learning with Kernels.* MIT Press, 2002.

[37] P. H. Sneath. The application of computers to taxonomy. *Microbiology*, 17(1):201–226, 1957.

[38] Speech and Image Processing Unit. Clustering Data Sets, 2015.

[39] D. Tax and R. Duin. Support Vector Domain Description. *Pattern Recognition Letters*, 20(11):1191–1199, 1999.

[40] E. M. Voorhees. Implementing agglomerative hierarchic clustering algorithms for use in document retrieval. *Information Processing & Management*, 22(6):465–476, 1986.

[41] T. Wang. Template-Based Shell Clustering Using a Line-Segment Representation of Data. *IEEE Transactions on Fuzzy Systems*, 19(3):575–580, 2011.

[42] Z. Xu and Q.-L. Da. An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18(9):953–969, 2003.

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

[44] R. R. Yager, J. Kacprzyk, and G. Beliakov. *Recent developments in the ordered weighted averaging operators: theory and practice*, volume 265. Springer, 2011.

[45] J. Y. J. Yang, V. E. Castro, and S. Chalup. Support vector clustering through proximity graph modelling. *Proceedings of the 9th International Conference on Neural Information Processing, ICONIP 2002*, 2, 2002.

[46] Q. Yang. Introduction to the IEEE Transactions on Big Data. *IEEE Transactions on Big Data*, 1(1):2–15, 2015.