

WeVoS-ViSOM: An ensemble summarization algorithm for enhanced data visualization

Emilio Corchado^a, Bruno Barúque^{b,*}

^a Departamento de Informática y Automática, Universidad de Salamanca, Facultad de Ciencias, Plaza de los Caídos, s/n. 37008 Salamanca, Spain

^b Department of Civil Engineering, University of Burgos, Escuela Politécnica Superior—Campus Vena (Edif.C), C/ Francisco de Vitoria, s/n. 09006 Burgos, Spain

ARTICLE INFO

Available online 12 August 2011

Keywords:

Topology-preserving maps
Unsupervised learning
Data visualization
Ensembles
Summarization algorithm

ABSTRACT

This study presents a novel version of the Visualization Induced Self-Organizing Map based on the application of a new fusion algorithm for summarizing the results of an ensemble of topology-preserving mapping models. The algorithm is referred to as Weighted Voting Superposition (WeVoS). Its main feature is the preservation of the topology of the map, in order to obtain the most accurate possible visualization of the data sets under study. To do so, a weighted voting process between the units of the maps in the ensemble takes place, in order to determine the characteristics of the units of the resulting map. Several different quality measures are applied to this novel neural architecture known as WeVoS-ViSOM and the results are analyzed, so as to present a thorough study of its capabilities. To complete the study, it has also been compared with the well-know SOM and its fusion version, with the WeVoS-SOM and with two other previously devised fusion Fusion by Euclidean Distance and Fusion by Voronoi Polygon Similarity—based on the analysis of the same quality measures in order to present a complete analysis of its capabilities. All three summarization methods were applied to three widely used data sets from the UCI Repository. A rigorous performance analysis clearly demonstrates that the novel fusion algorithm outperforms the other single and summarization methods in terms of data sets visualization.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Topology-preserving maps are prominent among the great variety of visualization tools for the visualization of multidimensional data sets. The ViSOM [1,2] is a very interesting extension of the well-known Self-Organizing Map (SOM) [3–5] characterized by its capability to generate qualitative representations of the similarities between the data under analysis.

This family of models represents high-dimensional data sets on two-dimensional (2D) maps and facilitates, for the human expert, the interpretation of the internal structure of the data. These models are also characterized by the use of unsupervised and competitive learning.

The main problem of all the neural network algorithms in general is their relative instability [6,7]. Running the same algorithm, even using the same parameters, can lead to quite different results. The use of ensembles is one of the most widespread techniques for increasing the stability of classification models [8,9]. This meta-algorithm trains in a slightly different way, several

algorithms over the same data set and relies on their combined results; rather than in the results of a single execution of the algorithm. This is based on the intuitive idea that a committee of experts working to solve a particular problem would come up with a more reliable solution than a single expert working on the same problem.

The technique is used in a great number of studies that mainly examine classification problems. In this study, however, the desired result is to obtain the most reliable as possible representation of a multidimensional data set on a 2D map. Therefore, the classical ensemble summarization techniques are not directly applicable in this case.

Several algorithms for topographic map summarization have previously been proposed [10–12], although there are some characteristics of the topology-preserving models that have not been taken into account. This research presents and analyses a novel fusion version of the ViSOM called the WeVoS-ViSOM and compares it with the single SOM and ViSOM and its WeVoS fusion version. The study reports the application of these algorithms to three of the most widely used data sets in the UCI web repository [13]: Iris, Wine and Wisconsin Breast Cancer.

The rest of this study is organized as follows: Section 2 introduces the topology-preserving mapping. Section 3 presents five quality measures, previously proposed in the literature, used

* Corresponding author. Tel.: +34 947 259513; fax: +34 901 706775.

E-mail addresses: escorchado@usal.es (E. Corchado), bbaruque@ubu.es (B. Barúque).

to evaluate different properties of topology-preserving mapping algorithms in general. Section 4 includes a brief description of the ensemble meta-algorithms and several previously proposed algorithms for summarizing SOM ensembles. Section 5 describes the novel proposed summarization method in detail: the Weighted Voting Superposition ViSOM (WeVoS-ViSOM). Section 6 describes the evaluation of the properties of the summaries obtained by the WeVoS-ViSOM algorithm and compares them with those calculated for the maps generated by the single models and other summarization methods. Finally, in Section 7, the final conclusions and future lines of research are outlined.

2. Topology-preserving maps

Topology-preserving maps [4,14] comprise a family of techniques that are devised to produce a low-dimensional representation of the training samples while preserving the topological properties of the input space. The best known technique among them is the Self-Organizing Map (SOM) model [5]. It is based on a type of unsupervised learning called competitive learning; an adaptive process in which the units in a neural network gradually become sensitive to different input categories, sets of samples in a specific domain of the input space. The main feature of the SOM algorithm is its topology preservation. When not only the winning unit but also its neighbors on the lattice are allowed to learn, neighboring units gradually specialize to represent similar inputs, and the representations become ordered on the map lattice.

One interesting extension of this algorithm is the Visualization Induced SOM (ViSOM) [1,15], proposed for the direct preservation of the local distance information on the map, along with the topology. The ViSOM constrains the lateral contraction forces between units and hence regularises the inter-unit distances, so that distances between units in the data space are in proportion to those in the input space. The ViSOM does not only take into account the distance between a unit's weights and the input data entry for the update of a unit's weights, but also the distance between that unit and the Best Matching Unit of the whole map (BMU).

Hence, the difference between the SOM and the ViSOM hence lies in the updating of the weights of the neighbors of the winner unit, as can be seen from Eqs. (1) and (2).

Update of neighborhood units in SOM:

$$w_k(t+1) = w_k(t) + \alpha(t)\eta(v,k,t)(x(t) - w_k(t)) \quad (1)$$

Update of neighborhood units in ViSOM:

$$w_k(t+1) = w_k(t) + \alpha(t)\eta(v,k,t) \left[(x(t) - w_v(t)) + (w_v(t) - w_k(t)) \frac{d_{vk} - \Delta_{vk}\lambda}{\Delta_{vk}\lambda} \right] \quad (2)$$

where, x is the input to the network, w_k is the weight vector associated with neuron k , while w_v is the weight vector associated with the winning unit in the lattice or (BMU), $\alpha(t)$ is the learning rate of the algorithm; $\eta(v,k,t)$ is the neighborhood function (usually a Gaussian function), where v represents the position of the BMU for the particular x of time t and k the positions of the units in its neighborhood. λ is a "resolution" parameter, d_{vk} and Δ_{vk} are the respective distances between the units in the data space and in the map space.

3. Features to analyse

Several quality measures have been proposed in the literature to study the reliability and the results of topology-preserving models in representing the data set that have been trained with [16,17]. These quality measures are neither global nor unified, but are a set of complementary ones, as each one assesses a specific

characteristic of the performance of the map in different visual representation areas. Five of them are briefly presented in this section. These measures have been chosen with the objective of measuring as wide a range of these characteristics as possible.

As stated in the introduction, the aim of the novel model (WeVoS-ViSOM) is to obtain a truthful representation of the data set in a map to obtain the best possible visualization of its internal structure. Thus, the most important features to evaluate in this case are the neighboring relationships of the units on the map and the continuity of the map. These features are assessed by topographic error, distortion and to some extent the goodness of map. The two remaining measures (classification accuracy and mean square quantization error) complete the comparison of the models in this research.

Topographic error [18]: Topographic error is calculated by finding the first two BMUs for each entry of the data set and testing whether the second is in or is not the direct neighborhood of the first. This measure, although suitable for an approximation of the quality of a map, is considered somehow simplistic and therefore not completely reliable in some cases by several studies [17].

Distortion [19]: The SOM algorithm can be considered to optimize a particular function, when using a constant radius for the neighborhood function of its learning phase. This function can be used to quantify in a more reliable way than the previous measure the overall topology preservation of a map by means of a Distortion measure in this study. Special attention is paid to this measure in this research due to its relation with visualization properties.

Classification accuracy: Topology-preserving models can be easily adapted for classification of new samples using a semi-supervised procedure [20]. Once the network training is completed, the same data set used in the training stage is fed back into the network. Each unit of the map is labeled with the class it has most consistently recognized. When a new sample is presented to the network, it is classified by the class associated with the unit that is activated at that time. A high value in the classification accuracy rate implies that the units of the map are reacting in a more consistent way to the classes of the samples that are presented. As a consequence, the map should represent the data distribution more precisely.

Mean square quantization error: This measure can be calculated for any algorithm performing vector quantization. In this case, it indicates how well the units of the map approximate the data on the data set. In other words, it measures the closeness of the units composing the map to the different data entries that they recognize (i.e., those considered as the BMUs for that entry); in the input space.

Goodness of map [21]: Finally, this measure combines two of the previous error measures: the square quantization error and the distortion. It takes account of the distance between the input and the BMU as well as the distance between the first BMU and the second BMU along the shortest path between both on the grid map units, calculated solely with units that are direct neighbors in the map. Thus, it measures both the continuity of the mapping from the data set to the map grid and the accuracy of the map in representing the set.

4. Topology-preserving mapping fusion

4.1. Use of ensemble meta-algorithms

The use of an ensemble of similarly trained models or algorithms is intended to improve the performance of classification algorithms [22]. It has been observed that, even though one of the classifiers in an ensemble would yield the best performance,

the sets of patterns misclassified by the different classifiers would not necessarily overlap. In conclusion, different classifier designs potentially offer complementary information on the patterns to be classified and could be harnessed to improve the performance of the selected classifier. The aim is not to rely on a single decision making scheme, but rather to use all the designs or their subsets for decision making, by combining their individual opinions to derive a consensus decision [23,24].

The main problem of competitive-learning-based networks is their inherent instability due to the random nature of their learning algorithms. Central to this research is that the effect of this instability may, however, be minimized by the use of ensembles [25,26]. The learning algorithm of the topology-preserving maps family trains their composing units (or neurons) to specialize during the algorithm iterations in the recognition of certain types of patterns, which also determines the topology of the map. In a similar way to the classification process, it can be inferred that the map regions that do not accurately represent the nature of the data set do not necessarily overlap. Therefore, the visualization of a single map might be improved by adapting each of the composing units of a map in the best possible way to the data set under study by using ensemble techniques, as they offer complementary visualizations of the data set.

Algorithms to combine classifiers can be divided into two broad classes. The simpler variety of algorithms [22] merely combines by averaging the results of each of the classifiers in some way into a final result. More complex types of algorithms [27,28] try to combine not only the results, but the whole set of classifiers; in an attempt to construct a single one that should outperform its individual components. Its main advantage is that it combines an improvement to classification quality with the simplicity of handling only one classifier.

This perspective of a single “summary” or “synthesis” of the patterns stored within the whole ensemble is followed in the present research to improve the performance of the model. The main intention is to obtain a unique map capable of representing the different features contained in the different maps of the ensemble in the clearest and most reliable possible way.

4.2. Summary of the topology-preserving map ensemble

The models used in this study are mainly designed as visualization tools. Constructing ensembles of classifier models is a viable option when trying to boost the classification capabilities of an ensemble, stabilizing its learning algorithm and avoiding overfitting; but when dealing with its visualization feature an ensemble is not directly displayable. Representing all the maps in a simple image can be useful when dealing with only 1D maps [10], but it is unmanageable when visualizing 2D maps. As a part of this research, a novel ensemble combination algorithm has been devised to overcome this problem, by generating a unique map representing the information contained in the different maps composing the ensemble. This combination algorithm is intended to generate an accurate and stable representation of data for visual inspection.

This part of the study encompasses several approaches inspired by previous work on SOM combination [29–31]. The study also includes previously presented methods that center on the generation of a final map summarizing the contents of several maps [11,12] for comparison purposes. Hereafter this process is called “Fusion”. The main characteristics of two of those methods are briefly described in Section 4.2.1 and their performance results are discussed in Section 6.

Then, a novel approach to the fusion of maps is presented in this work (WeVoS-ViSOM). It is fully described in Section 5 and its

performance is compared in Section 6 with previous devised algorithms.

4.2.1. Previous work: fusion of SOMs

The model in this study is compared with two fusion algorithms that were previously devised for similar purposes. These two algorithms [32,12] show great similarity to the one that is presented here, from among those identified in the literature by the authors. Although these algorithms have been developed by different authors and for different tasks, both employ a similar approach to the SOM fusion. Therefore, both approaches can in this case be considered as variants of the same “parent” algorithm, whereas the WeVoS-ViSOM is considered a different approach.

These previous Fusion of SOM meta-algorithms involve comparing the maps unit by unit in the input space. In other words, units that are considered ‘close enough’ to each other are fused to obtain a unit in the final fused map. This is done by calculating the centroid of the weights of the units that will be fused:

$$w_c = \frac{1}{|W_k|} \sum_{w_i \in W_k} w_i \quad (3)$$

being W_k the characteristic vectors of the set of units to fuse. That process is repeated until all units in all trained maps are fused into a unique final one. The criteria to determine which units are ‘near enough’ to be fused is what determines the two variants of the main algorithm.

Criterion 1: Voronoi Polygons [12]: Each unit in a Self-Organizing Map can be associated with a portion of the input data space called the Voronoi polygon [33]. That portion of the multi-dimensional input space contains data for which that precise unit is the BMU of the whole map. It is therefore logical to conclude that units related to similar Voronoi polygons may be considered to share similarities, as they should be situated relatively nearby in the input data space.

To calculate the dissimilarity between two units, a record of which data entries activated each unit as the BMU can be stored by associating a binary vector with the unit, the length of which is the size of the data set. The vector will contain ones (1) in the positions where the unit was the BMU for that sample, and zeros (0) in the rest of positions. The dissimilarity (i.e. the distance) between units can therefore be calculated as in Eq. (4):

$$ds(b_r, b_q) = \frac{\sum_{l=1}^n XOR(b_r(l), b_q(l))}{\sum_{j=1}^n OR(b_r(j), b_q(j))} \quad (4)$$

r and q being the units to determine their dissimilarity and b_r and b_q the binary vectors relating each of the units with the data samples they recognize. n represents the length of the binary vectors, that corresponds to the number of samples used to train the map. Based on what was explained before, it is easily concluded that the OR operation yields the positions where any of the two neurons recognized a data sample; while the XOR operation yields the positions where the same data sample was recognized by both neurons. So the equation represents the number of samples recognized by both units at the same time, in relation to the total number of samples recognized independently by both.

The main problem with this proximity criterion is that it depends on data recognition by the map, rather than on the map itself. This means that a unit that does not react as the BMU for any data could be considered similar to another unit in the same condition, even though they can be relatively far from each other in the input data space. To overcome this problem, all units with a reacting rate lower than a certain threshold are removed before calculating the similarities between the remaining units.

This implies that the neighboring properties of the whole map are no longer considered. The similarity criteria must be used again to maintain a notion of neighboring between the units of the fused map. Any dissimilarity below a given threshold between units will be considered as neighbors in the fused map.

This characteristic can be very useful when the objective of the analysis is to learn and represent the topology of the data set, as the remaining units will approximate the data set in the input space very well, enhancing the vector quantization feature of the SOM. Its drawback is the impossibility of representing that structure in a 2D map, as a lot of neighboring information between units is disregarded. The process is fully described in Algorithm 1.

Algorithm 1. Map Fusion by Voronoi Polygon Similarity.

Input: Set of trained topology-preserving maps: $M_1 \dots M_n$, usage threshold: θ_u , fusion threshold: θ_f , connection threshold: θ_c
Output: A final fused map: M_{fus}
1: Select a training set $S = \langle (x_1, y_1) \dots (x_m, y_m) \rangle$
2: train several networks by using the bagging (re-sampling with replacement) meta-algorithm : M_n
3: let θ_u , θ_f and θ_c be the usage, fusion and connection thresholds respectively
4: **procedure** Fusion ($M_1 \dots M_n$)
5: **for all** $M_i \in M_n$ **do** \triangleright for all maps in the ensemble
6: **for all** $w_j \in W_i$ **do** \triangleright for all neurons in each map
 \triangleright accept neurons with a recognition rate higher than a given threshold
7: $W_{fus} \leftarrow w_i$ if $\sum_i b_r(i) > \theta_u$
8: **end for**
9: **end for**
10: **for all** $w_i \in W_{fus}$ **do**
11: calculate dissimilarity between w_i and ALL neurons in W_{fus} (Eq. (4))
12: $D_i \leftarrow ds(w_i, w_k) \forall w_k \in W_{fus}$
13: **end for**
14: group into different sub-sets (Ws_n) the neurons that satisfy the conditions of $\begin{cases} ds(b_r, b_q) < \theta_f & \forall r, q \in Ws_n \\ ds(b_r, b_q) > \theta_f & \forall r, q \notin Ws_n \end{cases}$
15: **for all** Ws_n **do**
16: calculate the centroid (w_c) of the set by using Eq. (3)
17: add the centroid to the set of nodes of the final map (W_{fus}^*)
18: **end for**
19: **for all** $w_r \in W_{fus}^*$ **do** \triangleright for all neurons in the fused map
20: Connect w_r with any other neuron in W_{fus}^* , if they satisfy $b_r \in Ws_k, b_q \in Ws_l ds(b_r, b_q) < \theta_c$
21: **end for**
22: **end procedure**

Criterion 2: Euclidean Distance [11]: This method involves a unit-by-unit comparison of the maps in the input space, which implies that all the maps in the ensemble must have the same size. Firstly, it searches for the units that are closer in the input space (selecting only one unit in each map of the ensemble), which it then “fuses” to obtain the final unit in the “fused” map (see Eq. (3)). This process is repeated until all the remaining units have been fused. The high computational complexity of the algorithm is approached by using dynamic programming. The final fused map is initialized by calculating the fusion of only two of the maps composing the ensemble. Then, the same calculation is repeated between the resultant fused

map and another one of the maps composing the ensemble. The process continues until all the maps of the ensemble have been included in the calculation of the fused map.

The difference with the previous criteria is that, in this case, a pair-wise match of the units of each map is always possible, so the final fused map has the same size as each of its constituent ones. This also implies that a certain global neighboring structure can be maintained and reconstructed in the fused network. The algorithm that employs this criterion is fully described in Algorithm 2.

Algorithm 2. Map Fusion by Euclidean Distance.

Input: Set of trained topology-preserving maps: $M_1 \dots M_n$
Output: A final fused map: M_{fus}
1: Select a training set $S = \langle (x_1, y_1) \dots (x_m, y_m) \rangle$
2: train several networks by using the bagging (re-sampling with replacement) meta-algorithm : M_n
3: **procedure** FUSION(M_n)
4: initialise M_{fus} with the weight vectors of the first map: $M_{fus} \leftarrow M_1$
5: **for all** $M \in M_n$ **do**
6: **for all** $w'_i \in M_{fus}$ **do**
7: calculate Eucl. Dist. between w'_i and ALL neurons of map M_i
 \triangleright let w^* be the closest neuron in map M_i to the one selected in M_{fus}
8: $w^* \leftarrow \argmin_i (ED(w'_i, w_i))$
9: $w_c \leftarrow w'_i + w^* / 2$ \triangleright applying Eq. (3) to two neurons
10: $w'_i \leftarrow w_c$ replace w_i by the centroid (w_c)
11: **end for**
12: **end for**
13: **end procedure**

5. Weighted voting superposition for ViSOM

The idea behind the novel fusion variant presented in this study – WeVoS-ViSOM – is to obtain the final map on a unit-by-unit basis. However, instead of aiming for the best position for a single unit, as the two previously explained methods, this approach aims to obtain the best position for a unit and their neighbors. As a consequence, the final map obtained keeps one of the most important features of this type of algorithms: its topological ordering. This is an interesting characteristic, taking into account that the principal characteristic of the ViSOM is the enhancement of data visualization. Also, the modified weights update of the ViSOM, which provides the units with more freedom to adapt to the data set, potentially adds instability to the training. The WeVoS-ViSOM appears to be the most suitable fusion algorithm, in order to diminish this effect because the process acknowledges the neighboring of units.

The WeVoS scheme is an improved version of an algorithm presented in several previous works: superposition [31]. It has been applied to the simple SOM in previous works [34] with interesting results.

The first step in this meta-algorithm is to calculate the “quality” of each of the units composing each map, in order to rely on some kind of informed decision for the fusion of neurons. This “quality” (or error) measure can be any of the many quality of map measures existing in the literature on Topology-Preserving Maps, provided that it may be calculated on a unit-by-unit basis.

The final map is obtained again in a unit-by-unit basis. Firstly, the units of the final map are initialized by calculating the centroids of the units in the same position of the map grid in each of the trained maps. Then, the final vector of that unit is recalculated using the information associated to the units in that same position in each map. For each unit, a sort of voting process

is performed, as in Eq. (5):

$$V_{p,m} = \frac{\sum_{i=1}^M b_{p,i} \cdot \sum_{i=1}^M q_{p,i}}{\sum_{i=1}^M b_{p,i}} \quad (5)$$

where, $V_{p,m}$ is the weight of the vote for the unit included in map m of the ensemble, in its position p ; M is the total number of maps in the ensemble, $b_{p,m}$ is the binary vector used for marking the data set entries recognized by unit in position p of map m , and $q_{p,m}$ is the value of the selected quality measure for unit in position p of map m . The weights of the units are fed into the final network as it is done with data inputs during the training phase of a topology-preserving map, considering the homologous unit in the final map as the BMU. The weights of the final unit will be updated toward the weights of the composing unit. The difference in the updating performed for each homologous unit in the composing maps depends on the quality measure calculated for each unit: the higher the quality (or the lower the error) of the unit of the composing map, the stronger the updating of the unit of the summary map toward the weights of that particular unit. A single measure or a linear combination of several quality measures can be used for the determination of the final quality of a unit. The number of data inputs recognized by each unit is also taken into account in the quantization of the “most suitable” unit among those voting for the same position in the final map.

In short, the summarization algorithm considers the “more suitable” weights of a composing unit to be the weights of the unit in the final map according to both the number of inputs recognized and the quality of adaptation of the unit. The steps of this algorithm are fully described in Algorithm 3.

Algorithm 3. Weighted Voting Summarization algorithm.

Input: Set of trained topology-preserving maps: $M_1 \dots M_n$, training data set: S

Output: A final fused map: M_{fus}

1: Select a training set $S = \langle (x_1, y_1) \dots (x_m, y_m) \rangle$

2: train several networks by using the bagging meta-algorithm: $M_1 \dots M_n$

3: **procedure** WeVoS-ViSOM($M_1 \dots M_n$)

4: **for all** map $M_i \in M_n$ **do**

5: calculate the quality/error measure chosen for ALL neurons in the map

6: **end for**

▷ These two values are used in Eq. (5)

7: calculate an accumulated total of the quality/error for each position $Q(p)$

8: calculate recognition rate for each position $B(p)$.

9: **for all** unit position p in M_i **do**

10: initialize the fused map (M_{fus}) by calculating the centroid (w_c) of the neurons of all maps in that position (p) Eq. (3)

11: **end for**

12: **for all** map $M_i \in M_n$ **do**

13: **for all** unit position p in M_i **do**

14: calculate the vote weight (V_{p,M_i}) using Eq. (5).

15: feed the weights vector of neuron w_p into the fused map (M_{fus}) as if it was an input to the network.

The weight of the vote (V_{p,M_i}) is used as the learning rate (α).

The position of that neuron (p) is considered as the position of the BMU (v). ▷ This means that the neuron of the fused map (w_p^*) will approximate the neuron of the composing ensemble ($w_{p,m}$) according to the quality of its adaptation.

16: **end for**

17: **end for**

18: **end procedure**

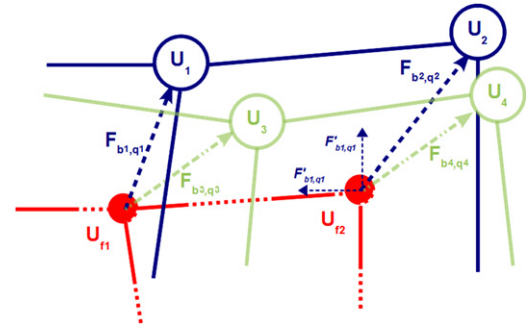


Fig. 1. Final adaptation of units in WeVoS-ViSOM.

It is interesting to note that the WeVoS scheme depends on the base algorithm composing the ensemble to update the final units and their neighborhood on the final map. That means that although very similar, the WeVoS-SOM is different from the WeVoS-ViSOM algorithm. The WeVoS-SOM performs the “fine tuning” of the units on the basis of the composing maps by using the SOM updating (Eq. (1)), while the WeVoS-ViSOM uses the ViSOM updating (Eq. (2)).

A schema representing how the final update of units in the WeVoS-ViSOM is performed is shown in Fig. 1. As can be seen, when the update of the final unit on the left (U_{f1}) is performed according to the homologous unit of the first map (U_1), its neighbor in the final map (U_{f2}) is also updated according to its distance from the unit being updated (U_{f1}) contracting or expanding the grid as is done in the ViSOM algorithm. This is repeated for all units of each composing map. This difference can be empirically appreciated in the experiments presented in Section 6.

6. Experiments and results

Several experiments have been designed and performed to investigate the capabilities of the WeVoS-ViSOM and also to compare it with the two other algorithms for obtaining a fused map from an ensemble. These experiments made use of three of the most popular data sets included in the UCI machine learning repository [13]: Iris, Echo-Cardiogram and Wine data sets. Experiments were performed using both ViSOM and SOM models over the three data sets to train different-sized ensembles, using the classical cross-validation method in order to select testing and training parts of the corresponding data set.

6.1. Test procedure

The procedure is as follows for all the experiments involving this combination of maps: a simple n -fold cross-validation is used in order to employ all data available for training and testing the model and to calculate its average performance. An ensemble of maps is calculated for each step of the cross-validation. The way the ensemble is trained does not affect the way the combination is computed. In this study, the bagging [22] meta-algorithm is used. Each individual map of an ensemble is trained on one of the re-sampled subsets ($n-1$ folds of the whole data set) and it is initialized in the same way, using exactly the same parameters for training. This generates n different trained networks which can be combined into a final network that is expected to outperform each of them individually. The combination of maps is complete once all the maps that compose the ensemble have been trained. Then, the data fold that was omitted from the training re-sampling, is used to test all the trained models: each of the

networks that compose the ensemble as well as the combinations that they generate.

6.2. Visualization results

In this sub-section a few examples of the most interesting visualization results obtained by the different models discussed in this research are presented.

Figs. 2 and 3 represent the adaptation of each map to its structure in representation of the data set under analysis. It depicts the lattices composing the maps embedded in a 2D input space. All the figures represent the Iris data set projected over its first two principal components [35]. The X-axis corresponds to the first PC and Y-axis corresponds to the second PC. Each of the figures also show a map trained over the data set, embedded in the input space formed by the principal components. As before, all maps were trained using the same parameters.

In Figs. 2a and 3a the single model maps displayed over the Iris data set in the input space may be seen. The former displays a SOM grid and the latter a ViSOM grid. In Figs. 2b and 3b the result of performing the algorithm of the Fusion by Euclidean distance is shown. It is easily observed how this algorithm focuses on

distributing the units over the data set in the best possible way, but obtains a map with a lot of twists and folds that does not preserve the topology very well. We display the fusion of the same ensemble in Figs. 2c and 3c, but using the fusion by Voronoi polygon similarity. Again, only the map training algorithm changes in the two figures. Evidently, the topology preservation is completely lost. Finally, Figs. 2d and 3d show the fusion of the ensemble, using the model presented in this work: the WeVoS (WeVoS-SOM and WeVoS-ViSOM respectively). It can be seen that the previously observed problems are not present in this model. What is more interesting is that the grids obtained using the WeVoS algorithm (both SOM and ViSOM) are more widely spread along the data set, covering the input space better than their corresponding single models. This shows visually how the use of ensembles can avoid the overfitting in this case.

The resultant maps that are used for visual inspection of the data corresponding to the same SOM, ViSOM and ensemble variants showed in Figs. 2 and 3 are displayed in Fig. 4. This figure represents the Iris data set in a 2D map and shows the final effects of each way of occupying the input data set by each of the models. As explained before, the Voronoi Polygon Similarity Fusion algorithm does not form a proper lattice, but rather a

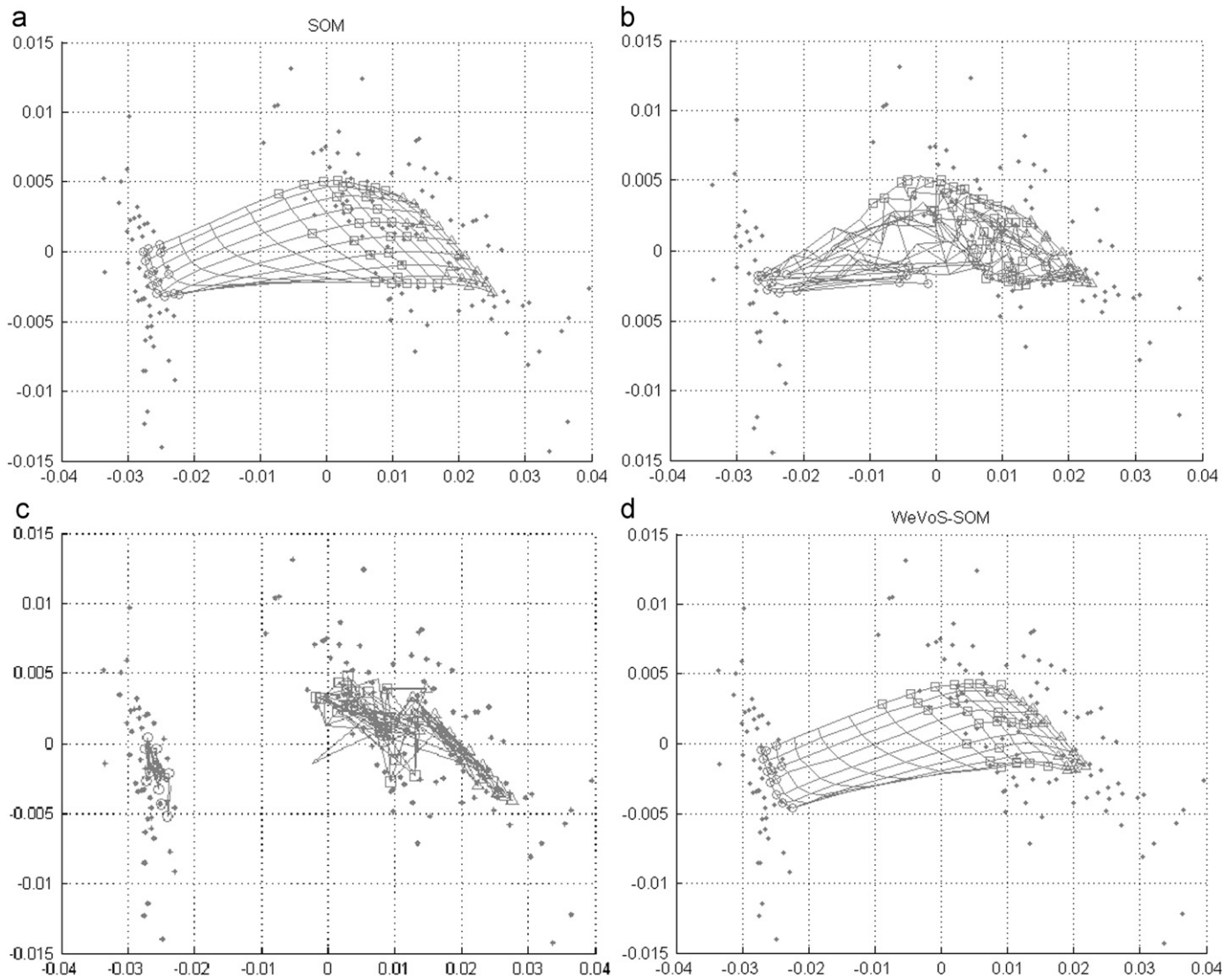


Fig. 2. Comparison of the adaptation the discussed algorithms to the input data set using the SOM as the base component for Iris data set. (a) Single SOM map. (b) Fusion of SOM maps using the Euclidean Distance criterion. (c) Fusion of SOM maps using the Voronoi polygon similarity criterion. (d) Fusion of five SOM maps using the WeVoS algorithm.

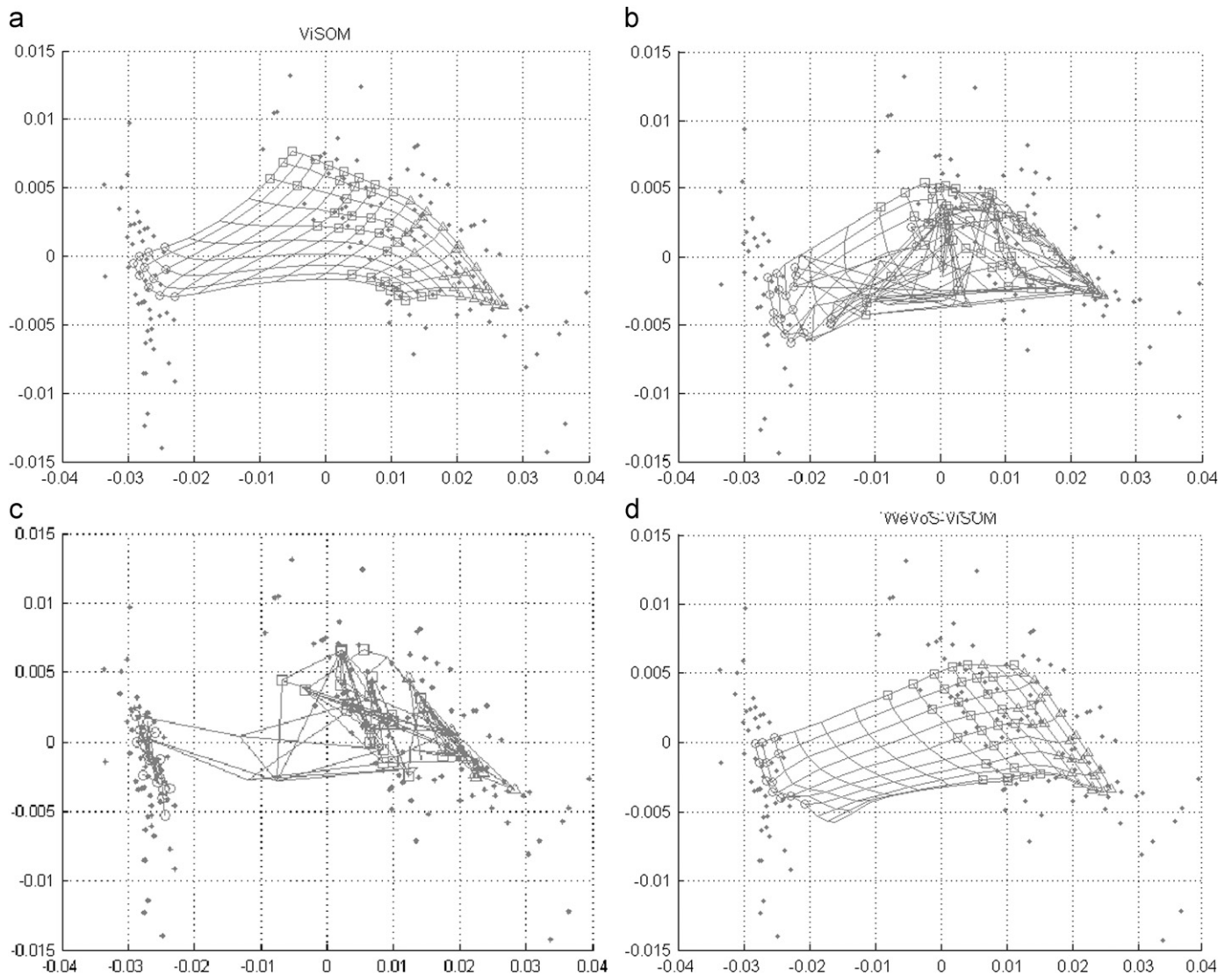


Fig. 3. Comparison of the adaptation the discussed algorithms to the input data set using the ViSOM as the base component for Iris data set. (a) Single ViSOM map. (b) Fusion of ViSOM maps using the Euclidean Distance criterion. (c) Fusion of ViSOM maps using the Voronoi Polygon Similarity criterion. (d) Fusion of ViSOM maps using the WeVoS algorithm.

graph; and unlike the others it is therefore not suitable for data representation. The six final maps shown were trained using the Iris data set. Ensemble fusion maps were obtained from the summarization of seven single maps (except Fig. 4a and 4d that shows the single SOM and ViSOM versions). Each unit of the maps is represented according to the class it has more consistently recognized.

It can be seen from Fig. 4 that the ViSOM algorithm provides in general a smoother map than the classic SOM algorithm. The WeVoS meta-algorithm improves the Single and Fusion by Distance models by obtaining, in general, more compact and more clearly separated groups than the other two. In comparison with the single model (Fig. 4a), the WeVoS-SOM (Fig. 4c) presents a much more separated group for class 1 (circles). In the single map, class 1 appears in a strip on the left corner on the map, leaving a considerable amount of dead units between data and the border of the map. On the other hand, the WeVoS-SOM presents a more separated group covering the top of the map. The Fusion by Distance (Fig. 4b) summarization algorithm does not significantly improve the single map with regard to the data representation feature of the model, as it obtains a map mixing the three classes in the top part of the map. The WeVoS-ViSOM

provides a better visualization compared with the single ViSOM model and the other two summarization algorithms. The single ViSOM (Fig. 4d) represents the Iris data set quite well, with a group of samples corresponding to class 1 clearly separated from the others. Although they also appear in a corner of the map, the cluster of class 1 is separated from the other cluster more than in the classic SOM model (Fig. 4a). The Fusion by Distance (Fig. 4e) map also contains a group separated from the rest, but includes samples of different classes. As explained before, the Fusion by Similarity is not suitable for 2D map representation, as some units are disregarded from the final model and therefore the topology preservation is lost. The WeVoS-ViSOM (Fig. 4f) clearly separates class 1 from the other two in a more compact group in the top of the image. Even comparing this model with the rest of models presented in Fig. 4, the WeVoS-ViSOM is the one that separates more clearly class 1 from the rest. The other two classes, although not so clearly separated as the first one, also appear unmixed between them and in a more compact group than in the single ViSOM map (Fig. 4d).

These results correspond to the main characteristic of the ViSOM: to express the dissimilarity between samples in the input space more clearly by representing distances on the final map

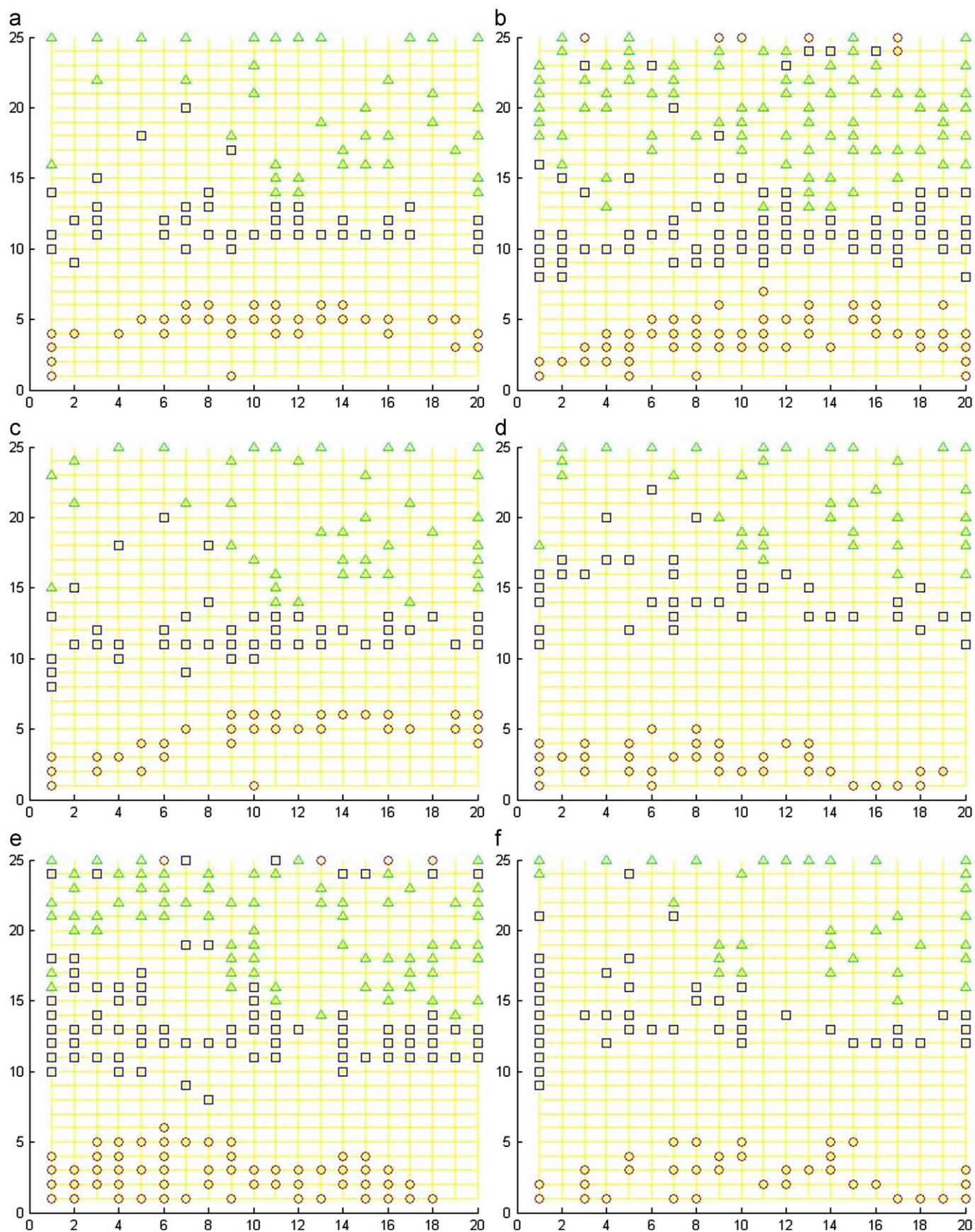


Fig. 4. Maps obtained for the single models and two of the different summarization models, for both the SOM and ViSOM training algorithms, trained over the Iris data set. (a) Single SOM. (b) Fusion by Distance applied to the SOM. (c) WeVoS-SOM. (d) Single ViSOM. (e) Fusion by Distance applied to the ViSOM. (f) WeVoS-ViSOM.

depending on that dissimilarity. The best results of the WeVoS-ViSOM were also expected, as it has as its main objective to emphasize the data structure characteristics that are discovered by the majority of its composing maps. One of the most notable characteristics of the Iris data set is precisely the clear separation between class 1 and classes 2 and 3.

6.3. Analytical results

This sub-section includes complete results for the experiments performed comparing the previously discussed models according to the analytical quality measures presented in Section 3. Two different sets of experiments were performed to compare the performance of the models when varying two different aspects of the training.

All measures presented in this section are error measures, so the desired value is always as close to 0 as possible. The Classification Error is presented in percentage form, normalized between 0 and 1, while the rest of the measures is absolute values. For the sake of clarity, the results for the Fusion by Voronoi Polygon Distances have been left out of the comparative, as they are completely different from the rest and are therefore not comparable.

6.3.1. Experiment 1

The first experiment involves training the ensembles over the complete data set; increasing the number of maps used to construct each one, to assess the effect of the modifications in the number of ensemble components.

Once the ensembles were trained, the fusion of the ensemble was computed by using the two variations explained in Section 4 and the novel algorithm presented in Section 5. In all cases, the weight of the vote for each unit in this latter model was

calculated according to the goodness-of-adaptation measure. All the measures were calculated using the test part of the data set, both for the average measure for the ensemble and for all the variations of the fusions of the ensemble. In all the figures shown in Sections 6.3.1 and 6.3.2, the ordinate axes represent the value of the error measure, while abscissa axes represent the number of composing models that are used by the fusion algorithm.

In the case of the Iris data set (Fig. 5) the different ensemble models do not appear to improve the results the quality measures. The exception to this is the Distortion (Fig. 5b), in which both the WeVoS-SOM and WeVoS-ViSOM obtain better results than the single models and the other fusion algorithms; although it is not a very significant improvement. The best performing algorithms in the Mean Quantization Error (Fig. 5a) and the Goodness of Maps (Fig. 5d) are the single models, the ViSOM being clearly the best of the two.

It is interesting to note that the Iris data set is the simplest data set in this study. This seems to confirm the effect suggested in ensemble theory: it is probable that the ensemble will not outperform the single model when using a base algorithm that performs correctly on a data set to construct the ensemble [22]. This effect comes from the fact that the single algorithm is already behaving in a reasonable stable way and therefore the benefits of using an ensemble meta-algorithm are negligible.

This situation changes for the Wine data set (Fig. 6). Although the best models appear to be the single ones for the quantization error (Fig. 6a), the ensemble algorithms perform best for the other three measures (Fig. 6c–d), especially in the Topographic Error (Fig. 6c) and the Goodness of Map (Fig. 6d). In this last measure, although the WeVoS-ViSOM outperforms the single ViSOM, the best algorithm is clearly the WeVoS-SOM.

In this case, the data set is more complicated than before, having 178 samples and 13 dimensions. This is an example case of

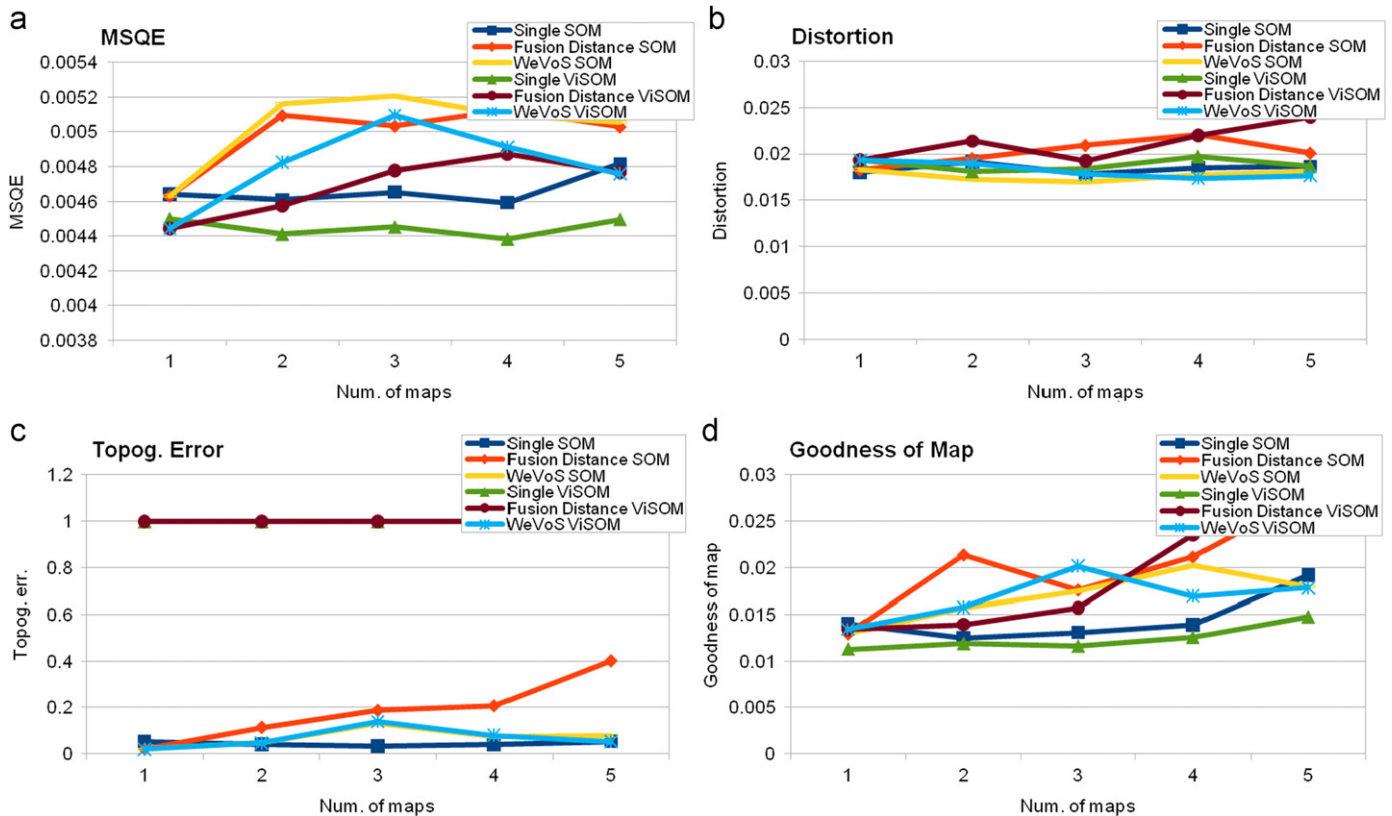


Fig. 5. Evolution of Iris data set results when increasing the number of maps composing the ensemble. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

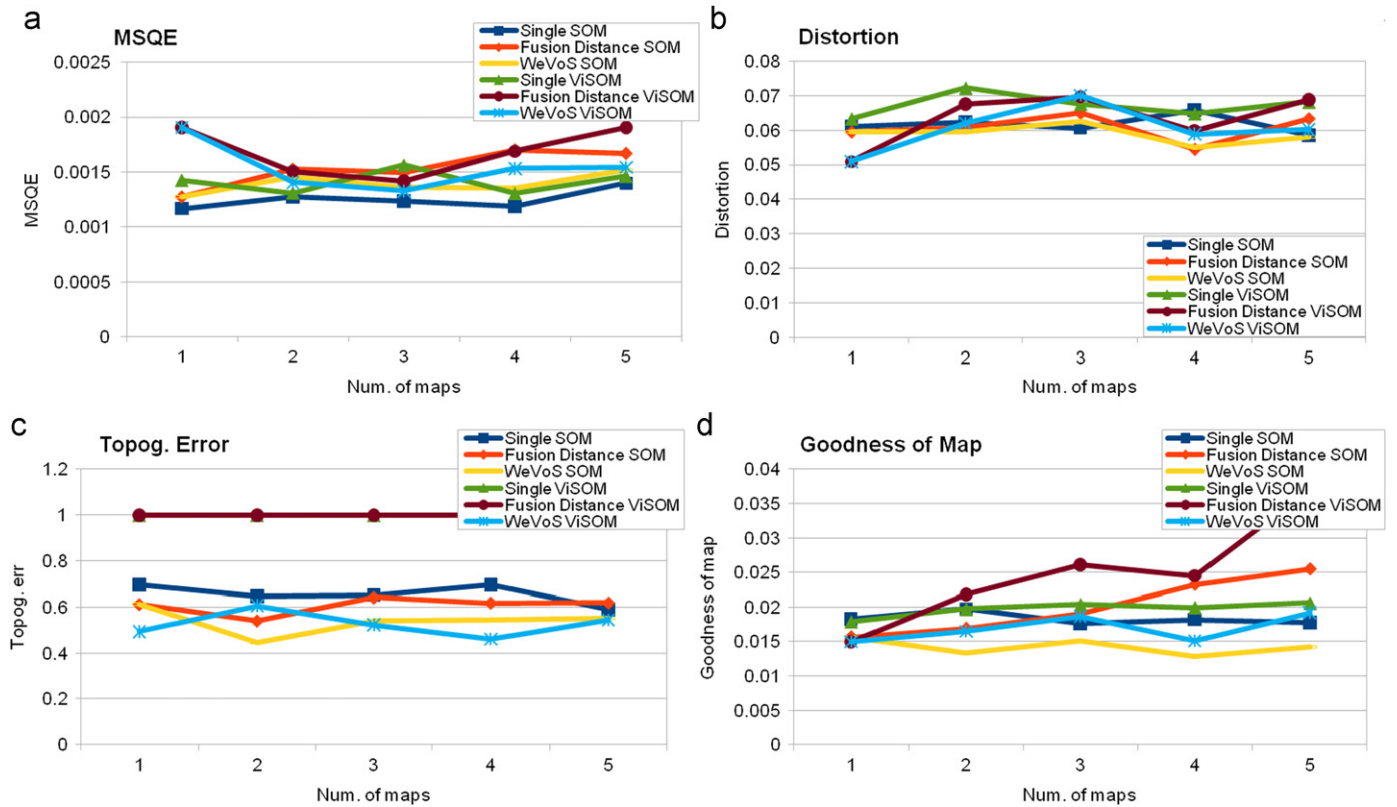


Fig. 6. Evolution of Wine data set results when increasing the number of maps composing the ensemble. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

a situation where the use of ensembles can clearly outperform the results of a single model. The objective of the presented WeVoS-ViSOM algorithm is precisely avoid overfitting – i.e. too close a recognition – of data samples, so the higher Quantization Error it exhibits, is an expected result. In the rest of the cases, where topographic ordering is measured, the results tend to favor the WeVoS-ViSOM presented model and its counterpart the WeVoS-SOM, as a proof of their better organization capabilities.

For the Echo-Cardiogram data set (Fig. 7) the results seem to be similar to those of the Wine data set. The quantization error (Fig. 7a) is higher in the ensemble fusion algorithms. For Distortion (Fig. 7b), ViSOM and WeVoS-ViSOM are very close to each other, although the single ViSOM seems to obtain slightly better results. For the other two measures (Fig. 7c and d) the WeVoS-ViSOM obtains the best results, and also has low variations between results, which indicate the stability of the algorithm.

This is another example of how the WeVoS-ViSOM can improve the results of the single ViSOM when the data set is hard to analyze and represent: the Echo-Cardiogram data set is composed of 105 samples and 9 dimensions. In this case the WeVoS-ViSOM outperforms even the WeVoS-SOM by obtaining even lower error measures for the Goodness of Map or Topographic Error measures, proving to be one of the best models in such types of situations when dealing with a not-very-large data set but composed of a high number of dimensions.

6.3.2. Experiment 2

The second experiment consists in using a moderate number of ensemble components, but modifying the number of data samples used for the training of the models. This emulates the addition of noise or instability to the data sets, as the training

process becomes more difficult when using less amount of data but maintaining its dimensionality.

Results for this experiment confirm the results obtained in the previous one. In the case of the Iris data set (Fig. 8) the WeVoS algorithm is not able to improve Quantization Error results (Fig. 8a), but it outperforms single models in the Distortion (Fig. 8b) and Topographic Error (Fig. 8c) measures, especially the WeVoS-ViSOM which obtains the lowest error. For the Goodness of Map (Fig. 8d) all models – except Fusion by Distance algorithms – behave in a very similar way, none of which outperform each other.

The effect of the added instability in this case is not particularly evident, as the data set analyzed is still quite simple. As a result, the use of the different ensemble does not prove to be of much use.

In the experiment with the Wine data set, the results (Fig. 9) are also better for the ensemble algorithms that use the WeVoS but with one proviso. Although both the WeVoS-SOM and WeVoS-ViSOM algorithms outperform their single homologous algorithm; the one obtaining lower error turns out to be the WeVoS-SOM. This is especially true for the Goodness of Map (Fig. 9d) but is to a lesser extent is similar to the Topographic Error (Fig. 9c). For the Quantization Error (Fig. 9a), the ensemble algorithm still yields results that are not as good as the single models.

As explained for Experiment 1, the effect of the use of the WeVoS meta-algorithm is more evident in this case. Still the WeVoS-SOM is the most useful model, obtaining slightly better results than the WeVoS-ViSOM.

The last experiment, which uses the Echo-Cardiogram data set (Fig. 10), has more pronounced results. In this case, is clear that ensemble models obtain higher Quantization Errors (Fig. 10a) than the single models. This is expected, as it is consistently true

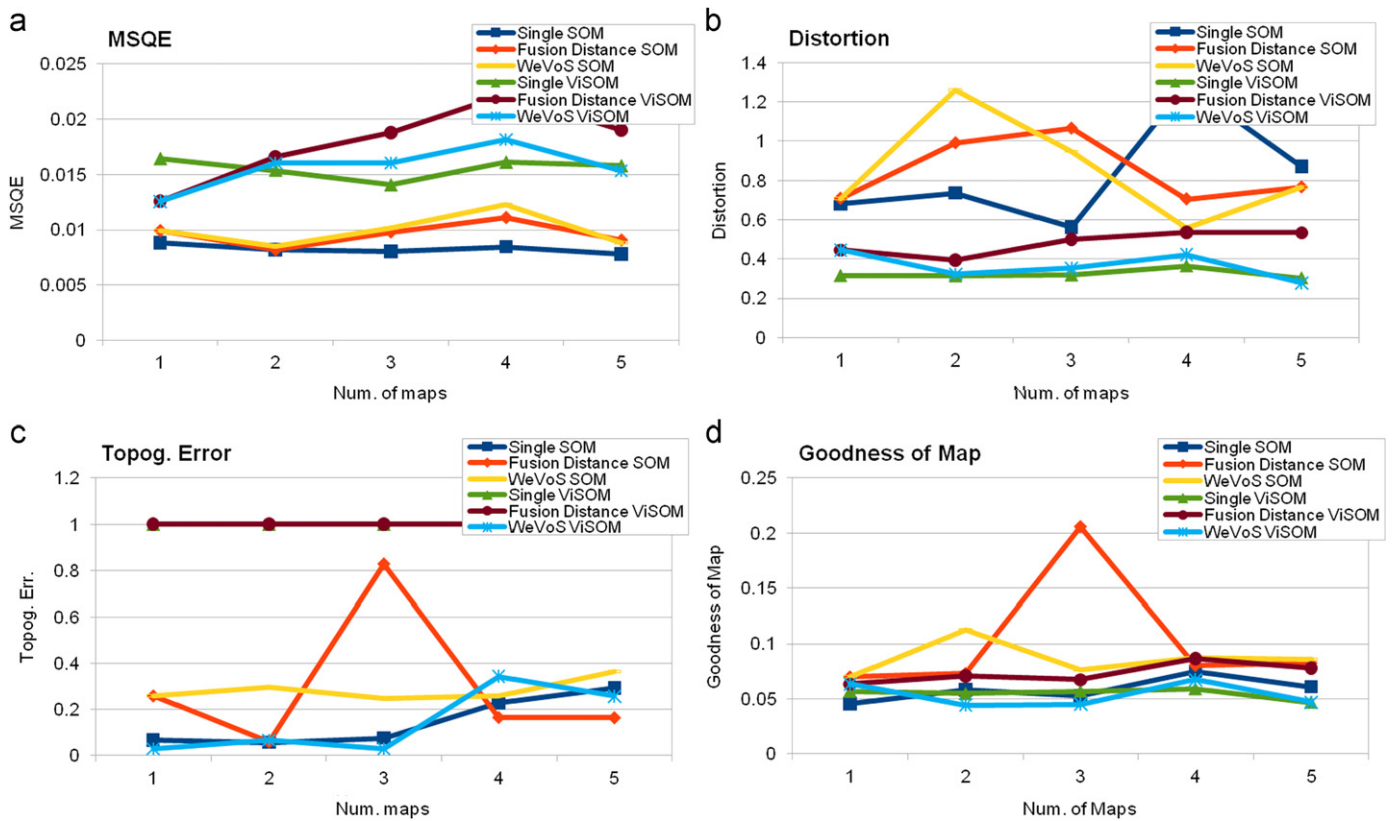


Fig. 7. Evolution of Echo-Cardiogram data set results when increasing the number of maps composing the ensemble. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

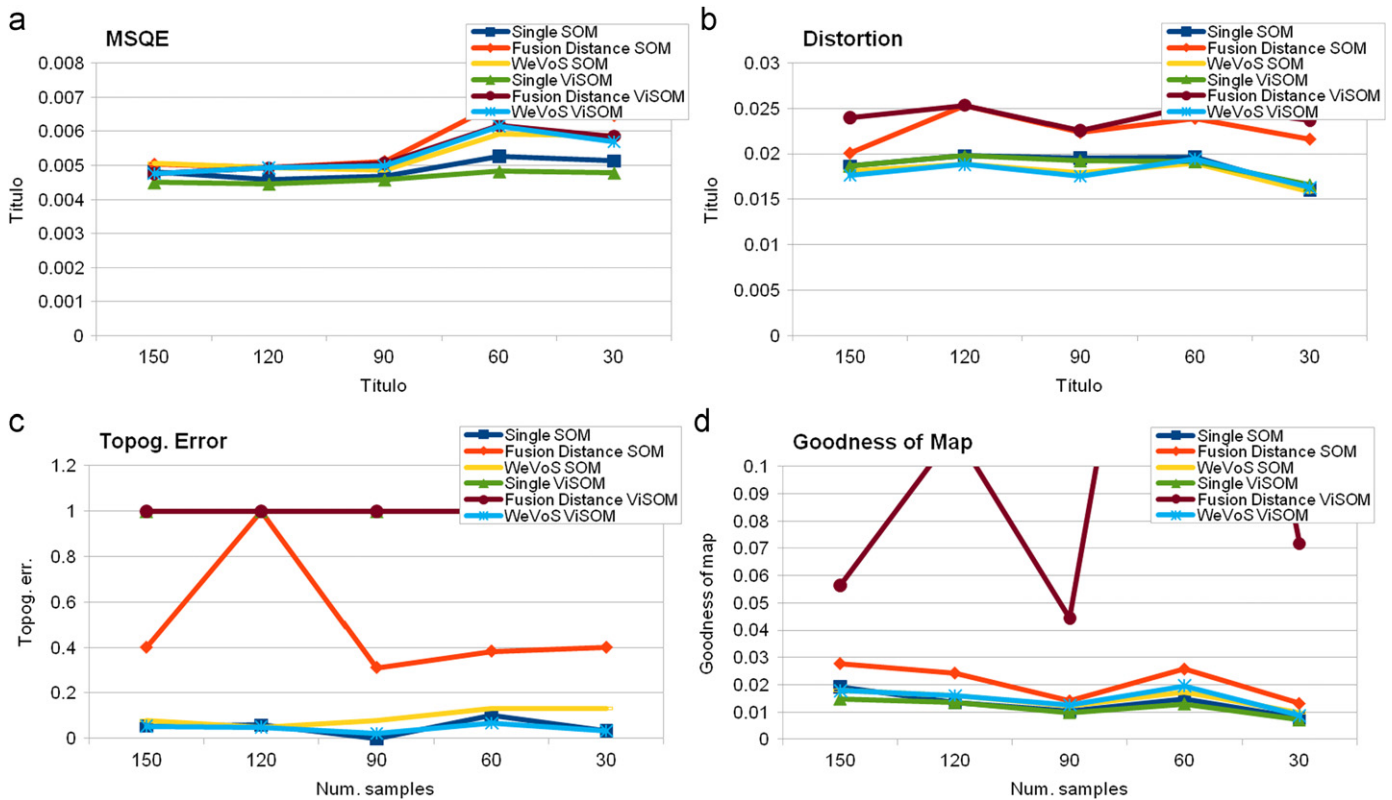


Fig. 8. Evolution of the Iris data set results when decreasing the size of the samples set used of training. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

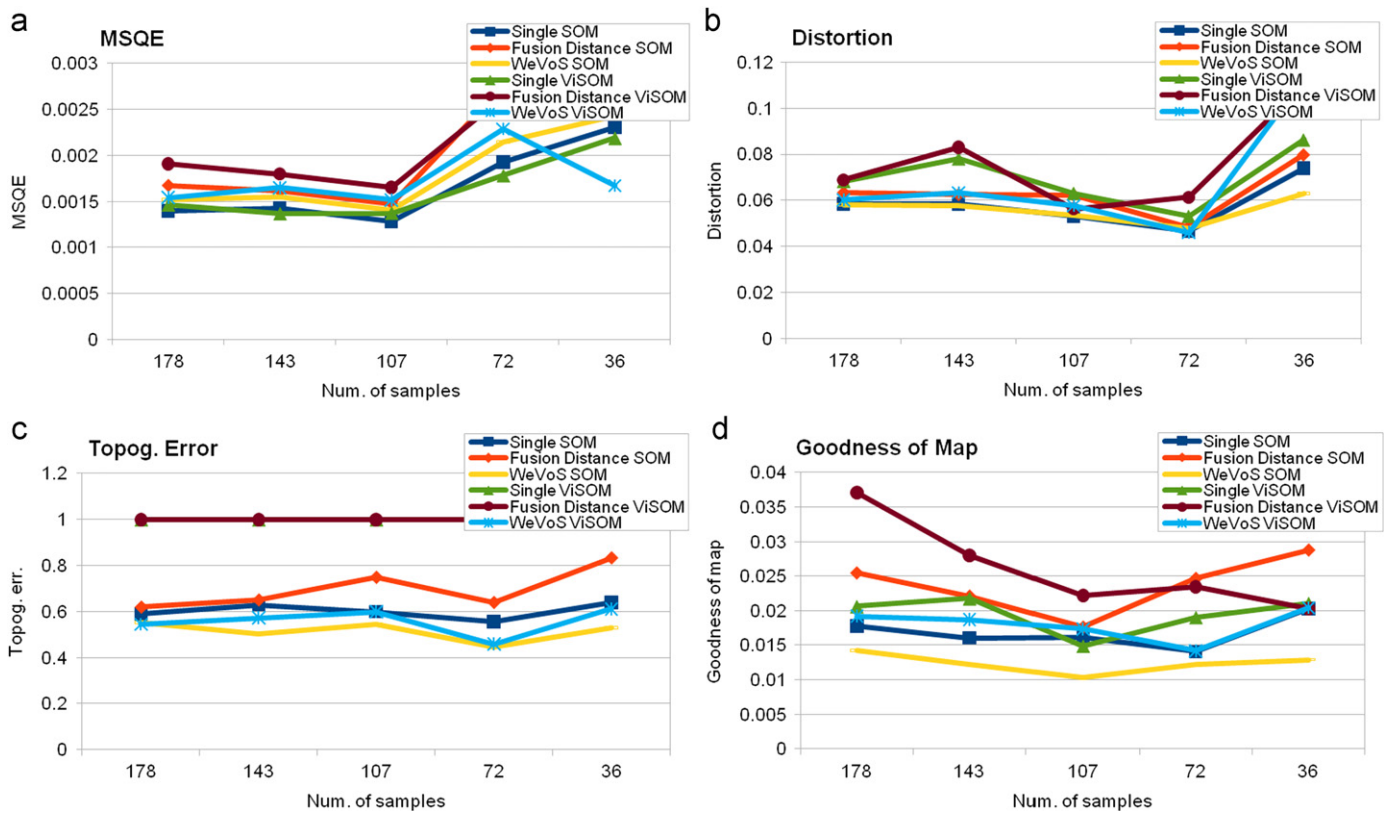


Fig. 9. Evolution of the Wine data set results when decreasing the size of the samples set used of training. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

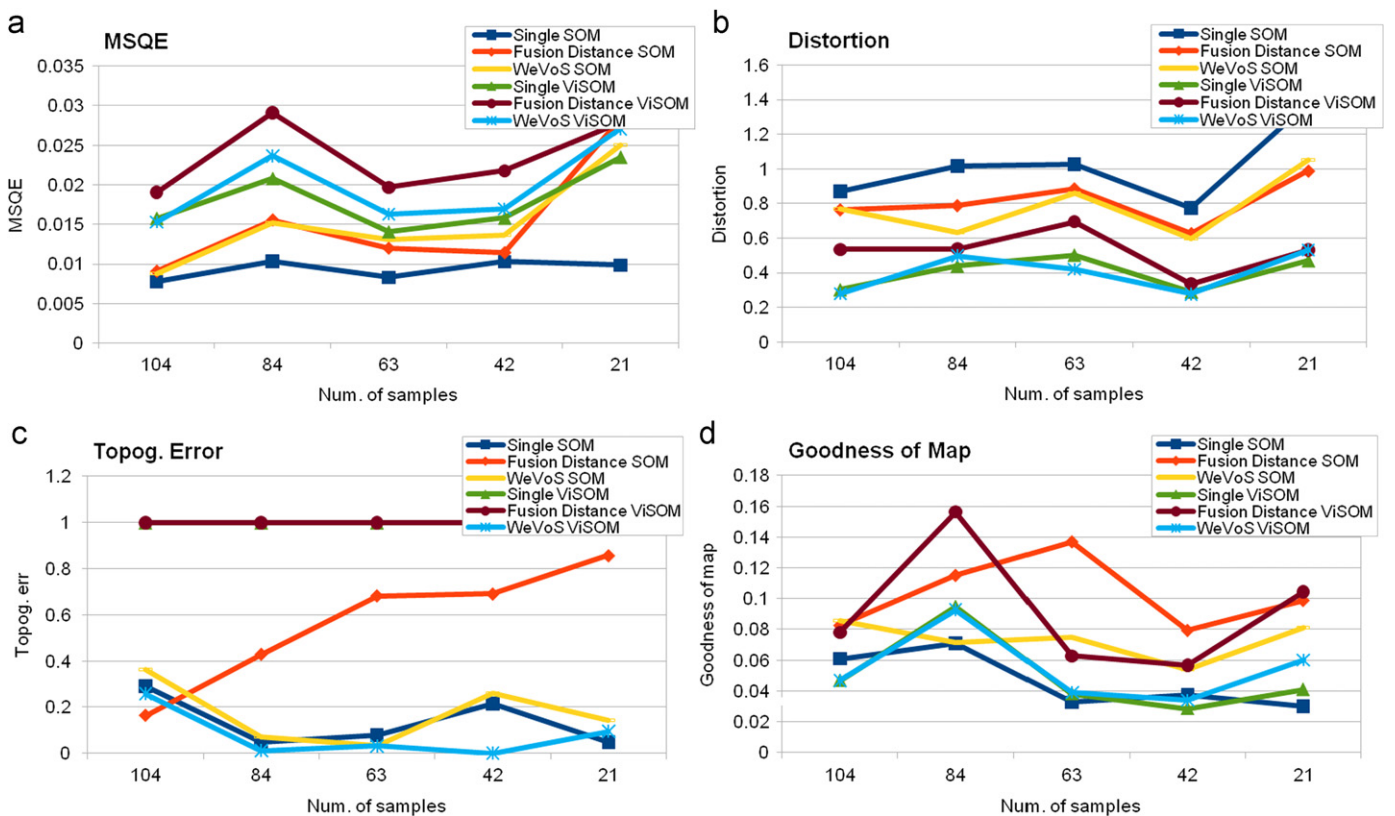


Fig. 10. Evolution of the Echo-Cardiogram data set results when decreasing the size of the samples set used of training. (a) Mean Quantization Error. (b) Distortion. (c) Topographic Error. (d) Goodness of Map.

in all experiments. For the Distortion measure (Fig. 10b), clearly the best model is the WeVoS-ViSOM, although very close to the regular ViSOM. The WeVoS-SOM clearly outperforms the single SOM. For the Topographic Error (Fig. 10c), the WeVoS-ViSOM clearly obtains the best results. The best performing models for the Goodness of Map measure (Fig. 10d) appear to be the single versions, especially the single SOM, which shows in this case a lower error than the WeVoS-SOM. The behavior of the WeVoS-ViSOM and the single ViSOM is not so good, but similar to the SOM; although ViSOM performs slightly better when the size of the data set is reduced to less than half of the original size.

In the case of the Eco-Cardiogram, the WeVoS-ViSOM obtains good results in spite of the added instability due to a reduction in the data set size. It is one of the models with lowest error and also one with the lowest variability in its results, obtaining quite similar results despite the data size reduction. This is a very interesting feature previously observed in other ensemble meta-models [36], derived from a combination of results of several different complementing models.

6.4. Discussion

The results included show some clear conclusions about the Fusion of topology-preserving mapping algorithms. Firstly, although sometimes the Fusion by Euclidean Distance can show a better classification performance than the single SOM (as [11] demonstrates), this could be due to the effect of the re-labeling of neurons rather than the improvement of its topological characteristics. Secondly, as the results show, the final structure obtained by this fusion algorithm is clearly not suitable for the best representation of the data set structure, due to the twists appearing in the map grid. On the contrary, the WeVoS scheme shows a much more regular grid, which as can be seen in the example presented with the Iris (Figs. 2 and 3) can serve to better adjust the grid and distribute its units in the data input space.

All of the analytical results support the idea that although the Fusion by Distance can obtain better classification results, the visualization characteristics of the resultant maps are generally poor.

Among the models compared, the WeVoS-ViSOM is the one that shows the best adaptation to the Iris data set, spreading the grid in a wider way over the data manifold. This translates into a better final visualization of the data set structure – as can be seen in Fig. 4 – due to the enhanced visualization capabilities of the ViSOM and the added improvement of the WeVoS fusion algorithm.

The WeVoS-ViSOM results verify some notable characteristics of the models. First, as all experiments point out, quantization error reduction is not the main purpose of this algorithm. It is evident from Figs. 2 and 3 – but also from each analytical result – that the algorithm tries to spread its units in a better way throughout the input data space, rather than concentrate them to where a larger amount of data is located; in order to get a more informative representation of the data space. The cost is a higher quantization error than other models. Concerning the other quality measures, the most interesting characteristic is that the usefulness of the WeVoS-ViSOM model for data visualization depends on the data set. As can be seen in Fig. 5, results for the Iris data set are not as good as those obtained by single models – with the exception of the Distortion measure. Fig. 6 shows that best performing models are the WeVoS-SOM (Distortion and Goodness of Map measures) and the WeVoS-ViSOM (Topographic Error measure). And, finally, Fig. 7 shows that the WeVoS-ViSOM obtains some of best results (especially Topographic Error and Goodness of Map). These results are similar in the case of the second experiment (Figs. 8–10). This suggests that for the ensemble to be really useful, the data set must have enough complexity from the point of view of an automated learning algorithm. For example: the Iris data set has 150 samples, but only

4 dimensions, while the Wine data set has 178 samples and 13 dimensions and the Echo-Cardiogram data set has 105 samples and 9 dimensions. In this case, as in classical classification ensembles; when a single algorithm performs in a correctly with a given data set, the ensemble fusion algorithms are not able to outperform it; nevertheless, if the data set is complex for the single model, the use of an ensemble meta-algorithm is able to further improve the capabilities of the single model.

7. Conclusions and future work

In this research a novel topology-preserving model known as WeVoS-ViSOM has been presented, analysed and compared with other models. This model has been devised to generate the most accurate visual representation of a multi-variable data set in the form of a 2D map that visually summarizes the principal features of the data set outlined by the different trained maps composing the ensemble. Its main objective is to obtain the most comprehensive visualization as possible, sacrificing as little as possible of the data's topological presentation, which is one of the main qualities of the Self-Organizing Maps. The main characteristic of the model is the smooth adaptation to the input space of the data set, correcting small defects that may arise on a single training; and therefore further improving the visualization capabilities of the ViSOM algorithm. The present study has included detailed descriptions of previously devised summarization algorithms and compared them with the new model. The performance of the summaries obtained by the WeVoS meta-algorithm has been analysed by means of a range of quality measures; and the usefulness of the WeVoS-ViSOM has been empirically proven, thereby demonstrating both a clearer and a smoother representation of the inner structure of the data set under study. Although it does not outperform single models with regard to classification accuracy or quantization error, it succeeds in reducing the distortion error of single models, thus obtaining a more truthful and organized representation of the data set. In the examples given, the WeVoS-ViSOM has obtained lower errors than the WeVoS-SOM, proving to be a very useful tool for data visualization. Future work will be focused on the application of the WeVoS to other topology-preserving models and to other case studies. Also some improvements to the way the ensemble is calculated, taken from the most widely used ensemble meta-algorithms, will be tested in a wider array of real-life problems.

Acknowledgments

This research has been partially supported through projects CIT-020000-2008-2 and CIT-020000-2009-12 of the Spanish Ministry of Education and Innovation and project BU006A08 of the Junta of Castilla and León. The authors would also like to thank the manufacturer of components for vehicle interiors, Grupo Antolín Ingeniería, S.A. within the framework of the MAGNO2008-1028 CENIT project, funded by the Spanish Ministry of Science and Innovation.

References

- [1] H. Yin, ViSOM—a novel method for multivariate data projection and structure visualization, *IEEE Transactions on Neural Networks* 13 (1) (2002) 237–243.
- [2] H. Yin, Data visualisation and manifold mapping using the ViSOM, *Neural Networks* 15 (8–9) (2002) 1005–1016.
- [3] T. Kohonen, Adaptive associative memory principle, *IEEE Transactions on Computers* C 23 (4) (1974) 444–445.
- [4] T. Kohonen, P. Lehtio, J. Rovamo, J. Hyvarinen, K. Bry, L. Vainio, A principle of neural associative memory, *Neuroscience* 2 (6) (1977) 1065–1076.
- [5] T. Kohonen, *Self-Organizing Maps*, vol. 30, Springer, Berlin, Germany, 1995.

- [6] T. Heskes, Balancing between bagging and bumping, in: M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems*, Denver, Colorado, USA, 1997, pp. 466–472.
- [7] B. Bakker, T. Heskes, Clustering ensembles of neural network models, *Neural Networks* 16 (2) (2003) 261–269.
- [8] H. Schwenk, Y. Bengio, Boosting neural networks, *Neural Computation* 12 (8) (2000) 1869–1887.
- [9] U. Johansson, T. Lofstrom, L. Niklasson, Obtaining accurate neural network ensembles, in: M. Mohammadian (Ed.), *International Conference on Computational Intelligence for Modelling, Control & Automation Jointly with International Conference on Intelligent Agents, Web Technologies & Internet Commerce*, vol. 2, Proceedings, 2006, pp. 103–108.
- [10] L. Petrakieva, C. Fyfe, Bagging and Bumping Self Organising Maps Computing and Information Systems Journal, University of the West of Scotland, Paisley, Scotland, 2003, doi:10.1.1.64.6186. URL <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.1385>.
- [11] A. Georgakis, H. Li, M. Gordan, An ensemble of SOM networks for document organization and retrieval, in: *International Conference on Adaptive Knowledge Representation and Reasoning (AKRR'05)*, 2005, pp. 6–141.
- [12] C. Saavedra, R. Salas, S. Moreno, H. Allende, Fusion of self organizing maps, in: *9th International Work-Conference on Artificial Neural Networks (IWANN 2007)*, 2007, pp. 227–234.
- [13] A. Asuncion, D.J. Newman, *UCI Machine Learning Repository*, 2007.
- [14] T. Kohonen, Data management by self-organizing maps, in: *WCCI*, 2008, pp. 309–332.
- [15] H.-T. Gou, J.-S. Xu, L. Wang, A new clustering method based on references, density and neural network, 2006, pp. 1060–1065. doi:10.1109/ICMLC.2006.258560.
- [16] D. Polani, Measures for the organization of self-organizing maps, in: U. Seiffert, L.C. Jain (Eds.), *Self-organizing Neural Networks Recent Advances and Applications Studies in Fuzziness and Soft Computing*, vol. 16, Physica-Verlag, Heidelberg, 2003, pp. 13–44.
- [17] G. Polzbauer, Survey and comparison of quality measures for self-organizing maps, in: J. Paralic, G. Polzbauer, R. Andreas (Eds.), *Fifth Workshop on Data Analysis (WDA'04)*, Elfa Academic Press, 2004, pp. 67–82.
- [18] K. Kiviluoto, Topology preservation in self-organizing maps, *IEEE International Conference on Neural Networks (ICNN'96)*, vol. 1, 1996, pp. 294–299.
- [19] J. Vesanto, M. Sulkava, J. Hollmen, On the decomposition of the self-organizing map distortion measure, in: *Proceedings of the Workshop on Self-Organizing Maps WSOM'03*, 2003, pp. 11–16.
- [20] J. Vesanto, Data mining techniques based on the self-organizing map, Ph.D. Thesis, Helsinki University of Technology, Espoo, Finland, 1997.
- [21] S. Kaski, K. Lagus, Comparing self-organizing maps, in: C.v.d. Malsburg, W.v. Seelen, J.C. Vorbruggen, B. Sendhoff (Eds.), *Lecture Notes in Computer Science*, Bochum, Springer, Germany, Berlin, 1996, pp. 809–814.
- [22] L. Breiman, Bagging predictors, *Machine Learning* 24 (2) (1996) 123–140.
- [23] D. Ruta, B. Gabrys, Classifier selection for majority voting, *Information Fusion* 6 (1) (2005) 63–81.
- [24] A.P.M. Henriques, A.D.D. Neto, R.F. Amaral, Classification of multispectral images in coral environments using a hybrid of classifier ensembles, *Neurocomputing* 73 (7–9) (2010) 1256–1264.
- [25] M. Ron, R. Gunnar, An Introduction to Boosting and Leveraging, 2003.
- [26] F. Wang, C. Yang, Z. Lin, Y. Li, Y. Yuan, Hybrid sampling on mutual information entropy-based clustering ensembles for optimizations, *Neurocomputing* 73 (7–9) (2010) 1457–1464.
- [27] T.G. Dietterich, Ensemble methods in machine learning, in: *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, Springer-Verlag, London, UK, 2000, pp. 1–15.
- [28] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.
- [29] B. Baruque, E. Corchado, H. Yin, Visom ensembles for visualization and classification, in: *9th International Work-Conference on Artificial Neural Networks (IWANN'07)*, San Sebastian, Spain, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, pp. 235–243.
- [30] B. Baruque, E. Corchado, H. Yin, J. Rovira, J. González, Automated ham quality classification using ensemble unsupervised mapping models, in: B. Apolloni, R.J. Howlett, L. Jain (Eds.), *Lecture Notes in Computer Science*, Vietri Sul Mare, Italy, Springer, Berlin, Heidelberg, 2007, pp. 435–443.
- [31] E. Corchado, B. Baruque, H. Yin, Boosting unsupervised competitive learning ensembles, in: *17th International Conference in Artificial Neural Networks (ICANN 2007)*, Porto, Portugal, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2007, pp. 339–348.
- [32] A. Georgakis, H. Li, Content based image retrieval using a bootstrapped SOM network, in: *Third International Symposium on Neural Network (ISNN'06)*, 2006, pp. 595–601.
- [33] F. Aurenhammer, R. Klein, Voronoi diagrams, in: *Handbook of Computational Geometry*, no. 5, North-Holland, Amsterdam, Netherlands, 2000, pp. 201–290.

- [34] B. Baruque, E. Corchado, A weighted voting summarization of SOM ensembles, *Data Mining and Knowledge Discovery* 21 (2010) 398–426. URL <http://dx.doi.org/10.1007/s10618-009-0160-3>.
- [35] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Education Psychology* 24 (1933) 417–444.
- [36] A. Sharkey, N. Sharkey, Diversity, selection and ensembles of artificial neural nets, in: *Third International Conference on Neural Networks and their Applications*, IUSPIM, 1997, pp. 205–212.



Emilio S. Corchado is an Associate Professor of Computer Science at University of Salamanca (BISITE Research Group and GICAP Research Group). He is chair of the IEEE Systems, Man and Cybernetics—Spanish Chapter; member of the IEEE Spanish Section Board and member of the IEEE R8 PA subcommittee. He is or was a member of some European networks of Excellence as Nature-inspired Smart Information Systems (NiSIS) and COST Action IC0806: Intelligent Monitoring, Control and Security of Critical Infrastructure Systems (IntelliCIS). He is reviewer and evaluator for the European Commission for the 7FP, since 2007.

His research interests include neural networks, with a particular focus on exploratory projection pursuit, maximum likelihood hebbian learning, self-organising maps, multiple classifier systems and Hybrid Artificial Systems and its applications to industrial and environmental problems. He has published over 170 peer-reviewed articles in a range of topics from modelling of industrial processes, knowledge management and risk analysis, intrusion detection systems, food industry, artificial vision, and so on.

He is co-editor in Chief of the *International Journal of Computational Intelligence Research (IJ CIR)* and member of the Editorial Boards of the *International Journal of Computational Intelligence and Applications (IJ CIA)* and also of the *International Journal of Reasoning-based Intelligent Systems (IJRIS)*. Dr. Corchado is a guest editor of several special issues of reputed journals as *Information Science (Elsevier)*, *Knowledge and Information Systems (Springer)*, *Neurocomputing (Elsevier)*, *International Journal of Neural Systems (World Scientific)*, *Logic Journal Of the IGPL (Oxford Journals)*, *Neural Networks World (Institute of Computer Science AS CR and Faculty of Transportation Sciences)*, *Expert Systems (Wiley-Blackwell)*. He is general chair for the following conferences: *International Conference on Hybrid Artificial Intelligence Systems (HAIS)*, *International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO)* and *International Conference on Computational Intelligence in Security for Information Systems (CISIS)*.



Bruno Baruque holds an assistant lecturer position at the University of Burgos, Spain. He obtained his Ph.D. degree in Computer Science/Artificial Intelligence (European Mention) in 2009 and his Master Degree in Computer Science in 2004, both at the University of Burgos.

He is an active member of the GICAP (University of Burgos) and BISITE (University of Salamanca) Research Groups, having participated in many of the competitive projects that have been developed by those groups; both in the Computer Science Research and Teaching areas. He has also collaborated with other research groups at the LABRI (Bordeaux, France) or the University of Manchester (Manchester, UK), during research stays in the mentioned centres.

His research interests are focused on the automated learning field, with special emphasis on artificial neural networks. His work has mainly been related with unsupervised learning algorithms used for data visualization and ensemble learning models used to improve the results of those algorithms, having published many peer-reviewed papers, in those topics including theoretical and practical applications (food industry, environmental analysis, etc.) in reputed international journals and conferences.

He collaborates as reviewer of several international journals (*Neurocomputing*, *Pattern Analysis and Applications*, etc.) and as programme committee member in numerous conferences (*International Conference on Hybrid Artificial Intelligence Systems (HAIS)*, *International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO)*, *International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*).