A Novel Ensemble of Scale-Invariant Feature Maps

Bruno Baruque and Emilio Corchado

Abstract. A novel method for improving the training of some topology preserving algorithms as the Scale Invariant Feature Map (SIM) and the Maximum Likelihood Hebbian Learning Scale Invariant Map (MAX-SIM) is presented and analyzed in this study. It is called Weighted Voting Superposition (WeVoS), providing two new versions, the WeVoS-SIM and the WeVoS-MAX-SIM. The method is based on the training of an ensemble of networks and the combination of them to obtain a single one, including the best features of each one of the networks in the ensemble. To accomplish this combination, a weighted voting process takes place between the units of the maps in the ensemble in order to determine the characteristics of the units of the resulting map. For comparison purposes these new models are compared with their original models, the SIM and MAX-SIM. The models are tested under the frame of an artificial data set. Three quality measures have been applied for each model in order to present a complete study of their capabilities. The results obtained confirm that the novel models presented in this study based on the application of WeVoS can outperform the classic models in terms of organization of the presented information.

1 Introduction

Among the great variety of tools for multi-dimensional data visualization, several of the most widely used are those belonging to the family of the topology preserving maps [16]. Two interesting models are the Scale Invariant

Bruno Baruque
University of Burgos
bbaruque@ubu.es

Emilio Corchado
University of Burgos
escorchado@ubu.es

M. Kurzynski and M. Wozniak (Eds.): Computer Recognition Sys. 3, AISC 57, pp. 285–293.
springerlink.com

Map (SIM) [8] and the Maximum Likelihood Scale Invariant Map (MAX-SIM) [5]. Both are designed to perform their best with radial data sets due to the fact that both create a mapping where each neuron captures a "pie slice" of the data according to the angular distribution of the input data. The main difference between this mapping and the SOM is that this mapping is scale invariant. When the SOM is trained, it approximates a Voronoi tessellation of the input space [17]. The scale invariant map, however, creates a mapping where each neuron captures a "pie slice" of the data according to the angular distribution of the input data.

The main problem of all the neural network algorithms is that, they are rather unstable [11]. The use of ensembles is one of the most spread techniques for increasing the stability and performance of an analysis model [3,12]. This meta-algorithm consists on training several slightly different models over the same dataset and relying on their combined results, rather than in the results of a single model.

There are many combination algorithms in classification ensembles literature; but few of them, up to the knowledge of the authors, is directly applicable to topology preserving algorithms. Several algorithms for topographic maps summarization have been previously proposed [19, 10, 22], although there are some characteristics of the topology preserving models that have not been taken into account. In this research we present and analyse a new fusion algorithm called Weighted Voting Superposition (WeVoS) applied for the first time to the SIM and MAX-SIM. The study reports the application of these algorithms under one artificial dataset, created accordingly to the main characteristic of the models under study.

2 Topology Preserving Maps

The main target of the family of topology preserving maps [15] is to produce low dimensional representations of high dimensional datasets maintaining the topological features of the input space.

The Scale Invariant Map (SIM) [8] uses a simple network which uses negative feedback of activation and simple Hebbian learning to self-organize. By adding neighbourhood relations to its learning rule, it creates a feature map which has the property of retaining the angular properties of the input data, i.e. vectors of similar directions are classified similarly regardless of their magnitude.

A SIM is also a regular array of nodes arranged on a lattice. Competitive learning and a neighbourhood function are used in a similar way as with the SOM. The input data (x) is fed forward to the outputs yi in the usual way. After selection of a winner, the winner, c, is deemed to be firing (yc=1) and all other outputs are suppressed (yi = 0,). The winner's activation is then fed back through its weights and this is subtracted from the inputs to calculate the error or residual e as shown in Eq. 1:

$$e = x - W_c \cdot y_c,\ (y_c = 1) \tag{1}$$

The Maximum Likelihood Scale Invariant Map (MAX-SIM) [5] is an extension of the Scale Invariant Map (SIM) based on the application of the Maximum Likelihood Hebbian Learning (MLHL) [6,9].

The main difference with the SIM is that the MLHL is used to update the weights of all nodes in the neighbourhood of the winner, once this has been updated as in Eq. 1. This can be expressed as in Eq. 2:

$$\triangle W_i = h_{ci} \cdot \eta \cdot sign\left(e - W_c\right) \left|e - W_c\right|^{p-1}, \forall i \in N_c \tag{2}$$

By giving different values to p, the learning rule is optimal for different probability density functions of the residuals. h_{ci} is the neighbourhood function as in the case of the SOM and N_c is the number of output neurons. Finally η, represents the learning rate.

During the training of the SIM or the MAX-SIM, the weights of the winning node are fed back as inhibition at the inputs, and then in the case of the MAX-SIM, MLH learning is used to update the weights of all nodes in the neighbourhood of the winner as explained above.

2.1 Features to Analyse

Several quality measures have been proposed in literature to study the reliability of the results displayed by topology preserving models in representing the dataset that have been trained with [20, 21]. There is not a global and unified one, but rather a set of complementary ones, as each of them asses a specific characteristic of the performance of the map in different visual representation areas. The three used in this study are briefly presented in the following paragraphs.

Topographic Error [14]. It consists on finding the first two best matching units (BMU) for each entry of the dataset and testing whether the second is in the direct neighbourhood of the first or not.

Distortion [18]. When using a constant radius for the neighbourhood function of the learning phase of a SOM; the algorithm optimizes a particular function. This function can be used to quantify in a more trustful way than the previous one, the overall topology preservation of a map by means of a measure, called distortion measure in this work.

Goodness of Map [13]. This measure combines two different error measures: the square quantization error and the distortion. It takes account of both the distance between the input and the BMU and the distance between the first BMU and the second BMU in the shortest path between both along the grid map units, calculated solely with units that are direct neighbours in the map.

3 Previous Work: Fusion of SOM

Several algorithms for fusion of maps have been tested and reviewed recently by the authors of this work [2, 4]. In the present study, two of them will be employed. The first one, called Fusion based on Voronoi Polygons Similarity [22], is characterized by determining the units of different maps that are suitable to be fused by comparing the input space covered by each unit [1]. That is, comparing what are called the Voronoi polygons of each unit. This summary is very good at recognizing and adapting its structure in the input space of the dataset, but is not really able to represent that same dataset in a 2-D map; thus being of no use for dimensional reduction and visualization tasks.

The second one, called Fusion based on Euclidean Distance [10] uses the classic Euclidean distance between units to determine their suitability to be fused, instead. In this case the model is able to represent the dataset as a 2-D map, but the way it computes the neurons to fuses is an approximate one, so is prone to minor errors in the topology preservation of the map.

4 Weighted Voting Superposition (WeVoS)

The novel algorithm presented in this work tries to overcome the problems outlined for the previous described models. The principal idea is to obtain the final units of the map by a weighted voting among the units in the same position in the different maps, according to a quality measure. This measure can be any found in literature, as long as can be calculated in a unit by unit basis.

The voting process used is the one described in Eq. 3:

$$V_{p,m} = \frac{\sum b_{p,m}}{\sum_{i=1}^{M} b_{p,i}} \cdot \frac{q_{p,m}}{\sum_{i=1}^{M} q_{p,i}} \tag{3}$$

where, $V_{p,m}$ is the weight of the vote for the unit included in map m of the ensemble, in its in position p, M is the total number of maps in the ensemble, $b_{p,m}$ is the binary vector used for marking the dataset entries recognized by unit in position p of map m, and $q_{p,m}$ is the value of the desired quality measure for unit in position p of map m. The detailed description of the WeVoS algorithm is included below under the title Algorithm 1.

5 Experiments and Results

An artificial 2-dimensional dataset was created for testing and comparing the different algorithms described in this study. It was generated by using classic Gaussian distributions. The Gaussian distributions were centred along five

Algorithm 1. Weighted Voting Superposition

1: Train several maps by using the bagging (re-sampling with replacement) meta-algorithm
2: Calculate the quality measure/error measure chosen for the unit in each position (p) of each map (m)
3: Calculate an accumulated total of the quality/error for each position $Q(p)$ in all maps
4: Calculate an accumulated total of the number of data entries recognized by a position in all maps $D(p)$
5: initialize the fused map (fus) by calculating the centroid (w') of the units of all maps in that position (p)
6: for each map in the ensemble (m) calculate the vote weight of each of its units (p) by using Eq. 3
7: for each unit (p) in each map (m), feed to the fused map (fus) the weights vector of the unit (p) as if it was an input to the map, using the weight of the vote calculatedin step 6 as the learning rate and the index of that same unit (p) as the index of the BMU

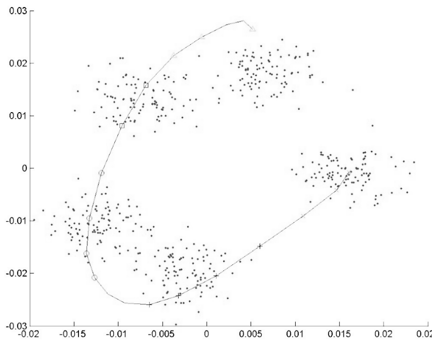
different points of the 2-D data space. The centres were disposed forming a "circular shape".

Tests were run using a classic five-fold cross-validation in order to use the complete dataset for training and testing. The ensembles were trained using one of the simplest meta-algorithm for ensemble training: the bagging meta-algorithm [3].

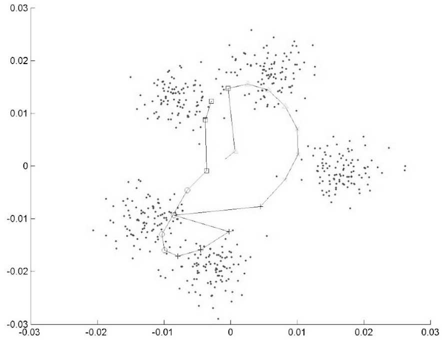
When a 1-D topology preserving map grid has to adapt to a circular shaped dataset (see Fig. 1) the most appropriate shape to use should intuitively be a circular shape, then the application of a SIM and MAX-SIM is very appropriated, as they tend to generate circular shaped grids, which are scale invariant.

Fig. 1(a) to Fig. 1(d) show the results obtained by training a single SIM and an ensemble of 6 different SIMs over the circular shaped dataset. The figures show the results of calculating each of the summarizing algorithms discussed for the ensembles.

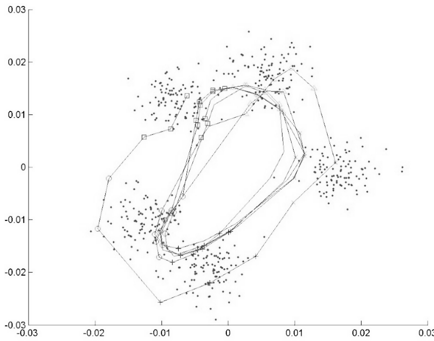
In this case (Fig. 1), all models use a circular shape for their adaptation to the dataset. The single SIM (Fig. 1(a)) adapts to the dataset correctly, but using a slightly too open map. The Fusion by Distance (Fig. 1(b)) suffers from an expected problem: several twists appear on its structure. To find the actual closest unit to another one using the Euclidean distance would be a NP-complete problem. That is why the algorithm works with an approximation to find the units to fuse. For classification problems that does not seem to be an important issue, but when dealing with visual inspection this approach has the problem of not always preserving strictly the topography of the network. Regarding the Fusion by Similarity (Fig.1(c)) it can easily be seen that, although the shape is correct, there are too many unnecessary units. This is due to the fact that, with such a big dataset, is impossible for a 1-D network



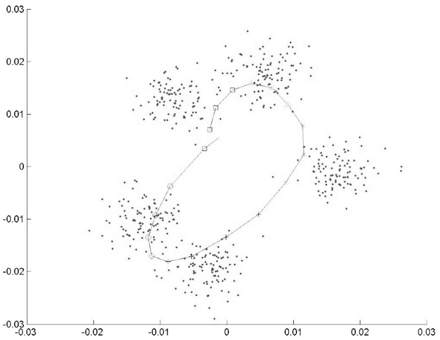
(a) A single SIM



(b) Fusion by Euclidean Distance



(c) Fusion by Voronoi Polygon Similarity



(d) Weighted Voting Superposition

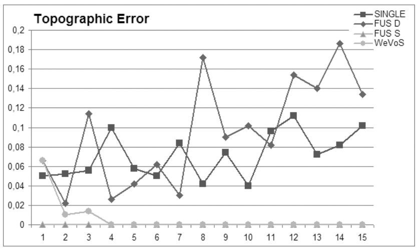
Fig. 1 The single SIM and the three summarizations for the same 6 network ensemble. All trained over the circular dataset employing the SIM learning Algorithm. The resultant 1-D networks are showed embedded in the 2-D dataset

to cover the data space correctly, so very few units are related with Voronoi polygons that overlap because they recognized mainly the same data entries. The WeVoS-SIM (Fig. 1(d)) obtains an oval shape without many twist or dead neurons, providing the best results.

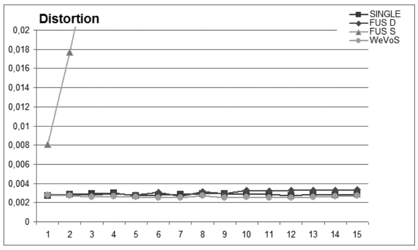
Results yielded by the application of the WeVoS-MAX-SIM are very similar to those obtained by the WeVoS-SIM . This second experiment's corresponding figures are not showed due to the limit of space.

Regarding the different summarization models, it can be observed how single models, adapt to the dataset correctly, but that their results can be improved by the use of the ensemble meta-algorithms. Among these summarization methods, the WeVoS obtains a simple network, without major twists and obtains a result more adapted to the dataset shape than the single model.

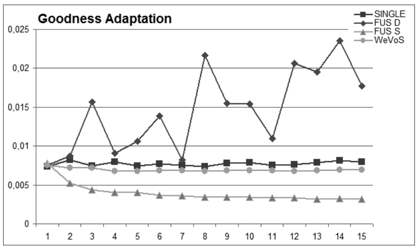
As stated in the introduction, the aim of the novel model presented (WeVoS) is to obtain a truly topology preserving representation of the dataset



(a) Topographic Error



(b) Distortion



(c) Goodness of Adaptation

Fig. 2 Measures for the MAX-SIM model both for the single version and the summarization algorithms obtained for the circular dataset

in a map. Thus, the most important features to evaluate in this case are the neighbouring relationship of the units of the map and the continuity of the map. These features are assessed by topographic error, distortion and to some extent goodness of map.

The following figures show the different three measures described in section 2.1 obtained for the single model and for the three different summarization algorithms described in this work.

What is shown for each measure in each figure is the comparison of quality obtained by each of the four fusion algorithms calculated over the same ensemble of maps. They represent how the measures vary when the number of maps included in the summary increases from 1 to 15. X-axis represents this number of maps, while Y-axis represents the measure obtained. All the three measures are errors, so the closest to 0, the best a result can be considered.

As expected, according to the measures regarding topographic ordering the WeVoS obtains better results than other models both for the SIM and the MAX-SIM. The exception to this is the Fusion by Similarity. Its results are not directly comparable, as the number of units contained in it is different from the rest, thus altering the results. The other three models behave in a more consistent way, being the WeVoS the model obtaining lower error rates. As stated before, the preservation of neighbourhood and topographic ordering is the main concern of the presented meta-algorithm. As a kind of confirmation of these results, again with the exception of the Fusion by

Similarity, the Goodness of adaptation of the map for the WeVoS is the best of all models.

6 Conclusions

An algorithm to summarize an ensemble of topology preserving maps has been presented in this work. This algorithm aims to obtain the best topology preserving summary as possible, in order to be used as a reliable tool in data visualization. Specifically in the case of the present work it has been applied to a model called Scale Invariant map and an extension of it, called Maximum Likelihood – SIM. Tests performed over an artificial dataset prove this algorithm as useful, obtaining better results than other similar ones. Future work includes the application of this algorithm to other topology preserving models and its combination with the use of other ensemble generation algorithms to boost its performance. Also real life dataset will be used in order to assess its usefulness for real problems.

Acknowledgements. This research has been partially supported through project BU006A08 by the Junta de Castilla y Len.

References

1. Barna, G., Chrisley, R., Kohonen, T.: Statistical pattern recognition with neural networks. Neural Networks 1, 1–7 (1988)
2. Baruque, B., Corchado, E., Yin, H.: ViSOM Ensembles for Visualization and Classification. In: Sandoval, F., Prieto, A.G., Cabestany, J., Graña, M. (eds.) IWANN 2007. LNCS, vol. 4507, pp. 235–243. Springer, Heidelberg (2007)
3. Breiman, L.: Bagging Predictors. Machine Learning 24, 123–140 (1996)
4. Corchado, E., Baruque, B., Yin, H.: Boosting Unsupervised Competitive Learning Ensembles. In: de Sá, J.M., Alexandre, L.A., Duch, W., Mandic, D.P. (eds.) ICANN 2007. LNCS, vol. 4668, pp. 339–348. Springer, Heidelberg (2007)
5. Corchado, E., Fyfe, C.: Maximum Likelihood Topology Preserving Algorithms. In: U.K. Workshop on Computational Intelligence (2002)
6. Corchado, E., Fyfe, C.: The Scale Invariant Map and Maximum Likelihood Hebbian Learning. In: International Conference on Knowledge-Based & Intelligent Information & Engineering System (2002)
7. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data Mining and Knowledge Discovery 8, 203–225 (2004)
8. Fyfe, C.: A scale-invariant feature map. Network: Computation in Neural Systems 7, 269–275 (1996)
9. Fyfe, C., Corchado, E.: Maximum likelihood Hebbian rules. In: European Symposium on Artificial Neural Networks (ESANN) (2002)

10. Georgakis, A., Li, H., Gordan, M.: An ensemble of SOM networks for document organization and retrieval. In: International Conference on Adaptive Knowledge Representation and Reasoning (AKRR 2005), pp. 6–141 (2005)
11. Heskes, T.: Balancing between bagging and bumping. In: Advances in Neural Information Processing Systems, vol. 9, pp. 466–472 (1997)
12. Johansson, U., Lofstrom, T., Niklasson, L.: Obtaining accurate neural network ensembles. In: International Conference on Computational Intelligence for Modelling, Control & Automation Jointly with International Conference on Intelligent Agents, Web Technologies & Internet Commerce, Proceedings, vol. 2, pp. 103–108.
13. Kaski, S., Lagus, K.: Comparing Self-Organizing Maps. In: Vorbrüggen, J.C., von Seelen, W., Sendhoff, B. (eds.) ICANN 1996. LNCS, vol. 1112, pp. 809–814. Springer, Heidelberg (1996)
14. Kiviluoto, K.: Topology preservation in self-organizing maps. In: IEEE International Conference on Neural Networks (ICNN 1996), vol. 1, pp. 294–299 (1996)
15. Kohonen, T.: The self-organizing map. Neurocomputing 21, 1–6 (1998)
16. Kohonen, T.: Self-Organizing Maps. Springer, Berlin (1995)
17. Kohonen, T., Lehtio, P., Rovamo, J., Hyvarinen, J., Bry, K., Vainio, L.: A principle of neural associative memory. Neuroscience 2, 1065–1076 (1977)
18. Lampinen, J.: On Clustering Properties of Hierarchical Self-Organizing Maps. Artificial Neural Networks 2, 1219–1222 (1992)
19. Petrakieva, L., Fyfe, C.: Bagging and Bumping Self Organising Maps. Computing and Information Systems Journal, 1352–1404 (2003)
20. Polani, D.: Measures for the Organization of Self-Organizing Maps. In: Self-organizing Neural Networks: Recent Advances and Applications. Studies in Fuzziness and Soft Computing, pp. 13–44 (2003)
21. Polzlbauer, G.: Survey and Comparison of Quality Measures for Self-Organizing Maps. In: Fifth Workshop on Data Analysis (WDA 2004), pp. 67–82. Elfa Academic Press, London (2004)
22. Saavedra, C., Salas, R., Moreno, S., Allende, H.: Fusion of Self Organizing Maps. In: Sandoval, F., Prieto, A.G., Cabestany, J., Graña, M. (eds.) IWANN 2007. LNCS, vol. 4507, pp. 227–234. Springer, Heidelberg (2007)
23. Vesanto, J., Sulkava, M., Hollen, J.: On the Decomposition of the Self-Organizing Map Distortion Measure. In: Proceedings of the Workshop on Self-Organizing Maps (WSOM 2003), pp. 11–16 (2003)
24. Voronoi, G.: Nouvelles applications des parametres continus a la theorie des formes quadratiques. Math. 133, 97–178 (1907)