

Application of Topology Preserving Ensembles for Sensory Assessment in the Food Industry

Bruno Baruque¹, Emilio Corchado¹, Jordi Rovira², and Javier González²

¹ Department of Civil Engineering, University of Burgos, Spain
bbaruque@ubu.es, escorchado@ubu.es

² Department of Biotechnology and Food Science, University of Burgos, Spain
jrovira@ubu.es, javigonzaalez77@hotmail.com

Abstract. Weighted Voting Superposition (WeVoS) is a novel summarization algorithm that may be applied to the results of an ensemble of topology preserving maps in order to identify the lowest topographical error in a map and thereby, to calculate the best possible visualization of the internal structure of its datasets. It is applied in this research to the food industry field that is studying the olfactory properties of Spanish dry-cured ham. The datasets used for the analysis are taken from the readings of an electronic nose, a device that can be used to recognize the sensory smellprints of Spanish dry-cured ham samples. They are then automatically analyzed using the previously mentioned techniques, in order to detect those batches with an anomalous smell (acidity, rancidity and different type of taints).. The Weighted Voting Superposition of ensembles of Self-Organising Maps (SOMs) is used here for visualization purposes, and is compared with the simple version of the SOM. The results clearly demonstrate how the WeVoS-SOM outperforms the simple SOM method.

1 Introduction

Topology preserving maps [1, 2] are often used for data visualization and inspection tasks. This interesting feature can assist human operators in classification tasks, such as the one presented in this study relating to the olfactory properties of Spanish dry-cured ham. Other features are pattern recognition and automated classification, inherent to many of the unsupervised learning techniques, which are especially relevant in the present application. These models are given enhanced stability in this study through the use of Weighted Voting Superposition (WeVoS), a novel ensemble summarization algorithm.

A combination of an electronic device for the analysis of volatile compounds (hereafter the electronic or “e-nose”) and a novel ensemble summarization algorithm for topology preserving mapping algorithms is used to study a wide variety of samples of “Serrano” Hams, in order to test whether this procedure is able to discriminate, in an easy and reliable way, between hams with different olfactory characteristics.

Consumer trust is a very important factor, when a product is being introduced into a new market or consolidated in an existing one. Dry-cured ham is a widely consumed traditional product in Spain that has also found a market outside Spain and is increasingly exported abroad. “Serrano Ham” is a salted ham that has been cured for over 210 days and is presented to the consumer on and off-the-bone. In these types of

products, rancid and acidic odours may be produced in storage; most of which may increase significantly because of proteolysis and lipid oxidation [3]. It is important to find quick and easy, low-cost techniques that apply simply parameters to evaluate the quality of these products prior to their sale and consumption by the consumer.

Several devices have appeared recently with the aim of enabling analytical techniques in the food industry to support the subjective decisions of professional testers. One disadvantage of these alternative tests is that whatever humans interpret as tastes and smells, machines will interpret as inevitably complex, numeric measurements. Thus, the aim of this multidisciplinary research is to devise an artificial intelligence system capable of interpreting the analyses made by an e-nose and presenting the results in an easily understandable way to human experts.

The rest of this paper is organized as follows: Section 2 and Section 3 present the AI models used in this research. Section 4 outlines the data gathering and pre-processing of the information in the samples, while Section 5 describes the experiments and results and finally, Section 6 presents the conclusions and future lines of research.

2 Topology Preserving Models

Topology preserving mapping comprises a family of techniques with a common target: to produce a low-dimensional representation of the training samples that preserves the topological properties of the input space. From among the various techniques, the best known is the Self-Organizing Map (SOM) algorithm [1, 2]. SOM aims to provide a low-dimensional representation of multi-dimensional datasets while preserving the topological properties of the input space. The SOM algorithm is based on competitive unsupervised learning; an adaptive process in which the neurons in a neural network gradually become sensitive to different input categories, which are sets of samples in a specific domain of the input space [4].

The update of neighbourhood neurons in SOM is expressed as:

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

where w_v is the winning neuron, α is the learning rate of the algorithm, and $\eta(v,k,t)$ the neighbourhood function, in which v represents the position of the winning neuron in the lattice and k the positions of the neurons in the neighbourhood of this one, and x , the network input.

This model can be adapted for classification of new samples using a semi-supervised procedure [5].

3 Ensembles of Topology Preserving Maps

The idea behind the novel fusion algorithm, WeVoS, is to obtain a final map keeping one of the most important features of these types of algorithms: its topological ordering. WeVoS is an improved version of an algorithm presented in several previous works: [6, 7] and in this study is applied for the first time to the SOM in the field of the food industry.

It is based on the calculation of the “quality of adaptation” of a homologous unit of different maps, in order to obtain the best characteristics of the vector in each of the units that make up the final map. This calculation is performed as follows:

$$V(p) = \frac{|x_p|}{\sum |x_p|} \cdot \frac{q_p}{\sum q_p} \quad (2)$$

In this study, two slightly different versions of the WeVoS meta-algorithm are compared: WeVoS (pos) and WeVoS (map). They differ in the way that they consider map units as “homologous” to the summary map units that they calculate. The first version, WeVoS (pos), considers the units that have been assigned to the same position in different maps as homologous. The second, WeVoS (map), considers other units in the neighbourhood of that unit in the same map as homologous.

The general WeVoS meta-algorithm is described in detail in *Algorithm 1*.

Algorithm 1. Weighted Voting Superposition (WeVoS)

- 1: train several networks by using the bagging (re-sampling with replacement) meta-algorithm
 - 2: **for each map** (m) **in the ensemble**
 - 3: **for each unit position** (p) **of the map**
 - 4: calculate the quality measure/error chosen for the current unit
 - 5: **end**
 - 6: **end**
 - 7: calculate an accumulated quality/error total for each homologous set of units $Q(p)$ in all maps
 - 8: calculate an accumulated total of the number of data entries recognized by an homologous set of units in all maps $D(p)$
 - 9: **for each unit position** (p)
 - 10: initialize the fused map (fus) by calculating the centroid (w') of the units of all maps in that position (p)
 - 11: **end**
 - 12: **for each map** (m) **in the ensemble**
 - 13: **for each unit position** (p) **of the map**
 - 14: calculate the vote weight of the neuron (p) in the map (m) by using **Eq. 2**
 - 15: feed the weight vector of the neuron (p), as if it were a network input, into the fused map (fus), using the weight of the vote calculated in **Eq 2** as the learning rate and the index of that same neuron (p) as the index of the BMU.
 - The unit of the composing ensemble (w_p) is thereby approximated to the unit of the final map (w') according to its weighting system.
 - 16: **end**
 - 17: **end**
-

4 A Food Industry Case Study

4.1 Preliminary Analysis of the Ham Samples

Several Spanish hams of different qualities and origins were used in this research. The data sets consisted of measurements taken from seven types of Spanish dry-cured ham from among the various brands available on the Spanish market. The samples also

included some that were tainted and/or that had a rancid/acidic taste. The tainted samples were randomly taken from among all the different quality types and origins of hams. The commercial brands of the hams in the samples were not taken into account in this study.

In this case the e-nose was used to measure the odour of the ham samples. The data collected was presented to the ensemble summarization algorithm of topology preserving maps, WeVoS, in order to achieve a simple and reliable device for testing and analysing the olfactory properties of the hams.

4.2 E-Nose Odour Recognition

The odour recognition process may be summarized as follows:

1. The sample is heated for a given time to generate volatile compounds in the head-space of the vial containing the sample.
2. The gas phase is transferred to a detection device which reacts to the presence of molecules.
3. The differences in sensor reactions are recorded using statistical calculation techniques to classify the odours. The readings taken by each sensor are separated and stored in a simple database for further study.

In this study the analyses are performed using an E-Nose α FOX 4000 (Alpha M.O.S., Toulouse, France) with a sensor array of 18 metal oxide sensors. The e-nose takes readings every 0.5 seconds, and has an acquisition time of 120 seconds and an acquisition delay of 600 seconds. Only the highest reading from each sensor is stored in the database for further analysis.

5 Empirical Evaluation

After having obtained the readings for each sample of cut ham taken from the 18-sensor array in the electronic nose, they are stored in a database along with the corresponding results of the sensory evaluation by the professional testers. These results are normally more detailed, but were restricted in this initial study to three possible values: “unspoilt”, “rancid/acid” and “tainted”. Thus, our final dataset consisted of readings taken from a total of 154 samples of ham, the readings on each ham being composed of 18 different variables measured over three possible categories.

Regarding the visual inspection of the maps obtained in Figs. 1(a) to 1(d), the projection of the dataset over its two first Principal Components (obtained by a conventional PCA analysis [8]) is shown alongside three maps obtained by training over the same dataset. If attention is paid to Fig. 1(a), it may be observed that the dataset is clearly ordered. Most of the unspoilt samples are situated in a compact group on the right of the image (triangles), while tainted samples are represented in 2 main groups to the left of the image (circles). The rancid/acidic samples (squares), although they might have been considered unspoilt, were on the point of spoiling. In Fig. 1(a) they are clearly shown as separate from the group of the normal samples and lie within the group of definitively tainted samples.

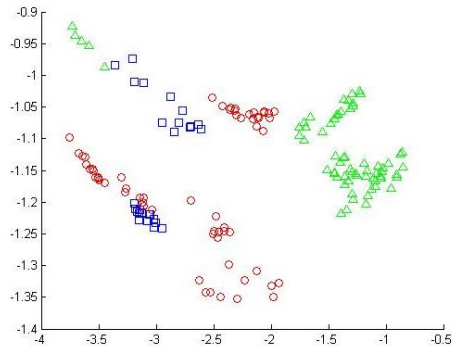


Fig. 1(a). Ham dataset projection over the first 2 Principal Components

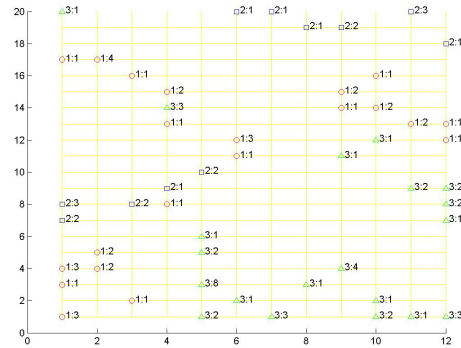


Fig. 1(b). Map obtained by training a single SOM over the ham dataset

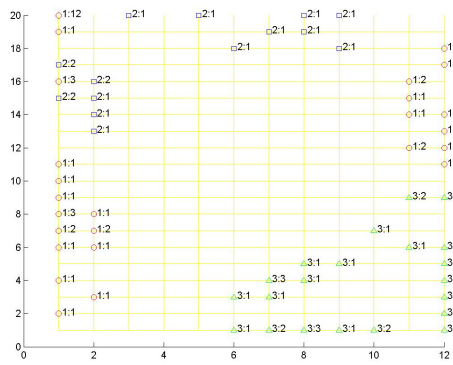


Fig. 1(c). Map obtained by calculating a WeVoS-SOM (pos) from an ensemble of SOMs trained over the ham dataset

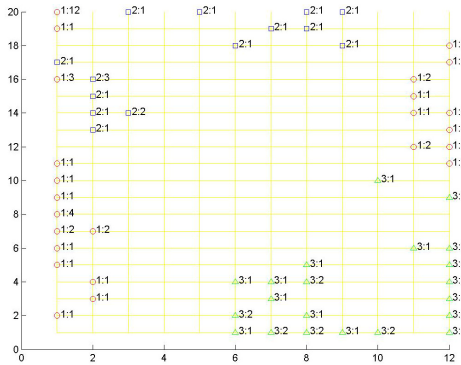


Fig. 1(d). Map obtained by calculating a WeVoS-SOM (map) from an ensemble of SOMs trained over the ham dataset

This same organization can be observed in the maps representing the dataset, although it is much less clearly represented in Fig. 1(b), which depicts the map obtained by a single SOM. The samples are scattered across this map, and there are even some unspoilt samples among the spoilt ones. On the contrary, in Figs. 1(c) and 1(d), the data appears more ordered, with all the samples in the unspoilt group to the bottom right-hand corner, clearly separated from the rest. Those maps even represent the gap separating the two groups of tainted samples, which emerges due to the different origin of the samples: one group is composed of samples originally of high quality hams that became spoilt, while the other is composed of samples of standard quality hams that also became spoilt. This situation is less evident if the only map observed is the one obtained from a single model (Fig. 1(b)).

The next step in the study was the training of single maps and ensembles of a different number of maps over the same subset of samples, in order to compare their characteristics. This was done using a standard 5-fold cross-validation technique in

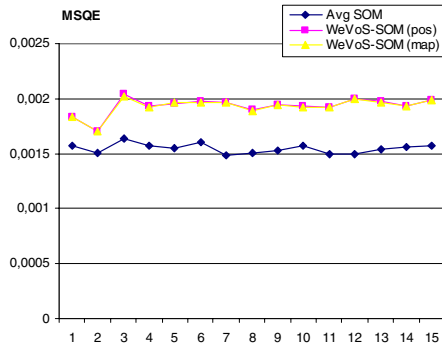


Fig. 2 (a). Mean Square Quantization Error

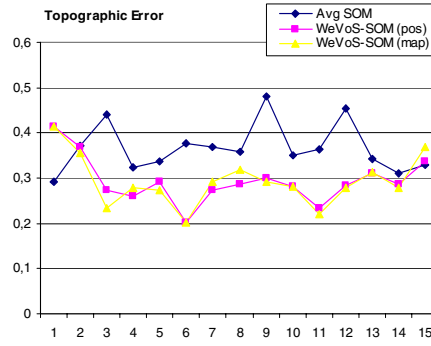


Fig. 2 (b). Topographic Error

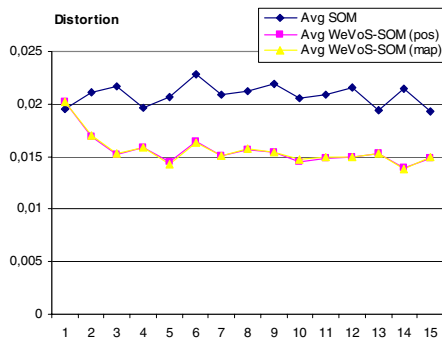


Fig. 2 (c). Distortion

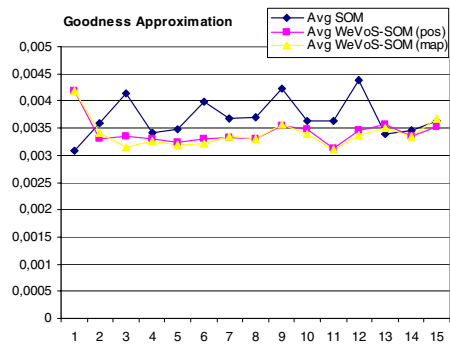


Fig. 2 (d). Goodness of Approximation

Fig. 2. Error readings for each of the models compared in the study (single map, WeVoS (pos) and WeVoS (map)). They were all obtained from the same basic SOM model. The x-axis represents the number of maps used in the ensemble and the y-axis the value of the measure.

order to be able to use the whole dataset for the tests. Each measure obtained represented the average of the measures obtained by each of the maps trained with 4-folds and tested over the other remaining fold.

Fig. 2 shows several measurements obtained from three models compared in the study, all of which are error measurements in different areas of representation of the dataset. The definitions of the Mean Square Quantization Error (MSQE), Topographic Error and Distortion are found in [9], while the Goodness of Approximation is described in [10]. As may be observed in the last three measurements (Fig. 2(b) to Fig. 2(d)), not only do the ensemble summarization methods (WeVoS) obtain a lower error, but they are also more stable, and do not depend on any specific execution of the training. As expected, the only exception to this is the MSQE, because it is a measurement of how far or how close the samples are from the unit that represents them, whereas the WeVoS meta-algorithm improves the visual representation of the dataset; the remaining measurements denote the accuracy of the representation in relation to the topological organization of the map.

6 Conclusions and Future Work

It has been shown that the summarization algorithm presented in this article is capable of providing a better visualization than the simple version of the SOM model.

In this case, it has been successfully applied to the readings taken from an E-nose in an assessment of the olfactory properties of different ham samples. The results suggest that this combination of techniques may easily be adapted to assist professional food testers in their work of classifying food samples or may even replace them in cases where simple explanations of taste are required.

Future work will include a more thorough study and classification of the samples in order to provide the professional food tester with maps that include more detailed information on the quality of the samples provided. Another line of work consists in adapting the algorithm to other extensions of the SOM to improve visualization, such as the Visualization Induced SOM (ViSOM).

Acknowledgements

This research has been partially funded through project BU006A08 of the JCyL.

References

- [1] Kohonen, T.: *Self-Organizing Maps*. Springer, Berlin (1995)
- [2] Kohonen, T.: The self-organizing map. *Neurocomputing* 21, 1–6 (1998)
- [3] Monahan, R.L., Brunton, N.P., Cronin, D.A., Durcan, R.: Determination of hexanal in cooked turkey using solid phase microextraction (SPME)/GC. In: *44th International Congress of Meat Science and Technology (ICoMST)*, vol. 1, pp. 586–587 (1998)
- [4] Kohonen, T., Lehtio, P., Rovamo, J., Hyvarinen, J., Bry, K., Vainio, L.: A principle of neural associative memory. *Neuroscience* 2, 1065–1076 (1977)
- [5] Kraaijveld, M.A., Mao, J., Jain, A.K.: A nonlinear projection method based on Kohonen's topology preserving maps. *IEEE Transactions, Neural Networks* 6, 548–559 (1995)
- [6] Corchado, E., Baruque, B., Yin, H.: Boosting Unsupervised Competitive Learning Ensembles. In: *International Conference of Neural Network (ICANN 2007)*, pp. 339–348 (2007)
- [7] Baruque, B., Corchado, E., Yin, H.: Quality of Adaptation of Fusion ViSOM. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) *IDEAL 2007*. LNCS, vol. 4881, pp. 728–738. Springer, Heidelberg (2007)
- [8] Hotelling, H.: Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Education Psychology* 24, 417–444 (1933)
- [9] Pözlbauer, G.: Survey and Comparison of Quality Measures for Self-Organizing Maps. In: *Fifth Workshop on Data Analysis (WDA 2004)*, pp. 67–82 (2004)
- [10] 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)