HAIS'10

5th International Conference on HYBRID ARTIFICIAL INTELLIGENCE SYSTEMS

A Bio-Inspired Fusion Method for Data Visualization

Bruno Baruque[1]

Emilio Corchado[2]

1 GICAP – Grupo de Inteligencia Computacional Aplicada
University of Burgos, Spain.
bbaruque@ubu.es

2 BISITE –
University of Salamanca, Spain
escorchado@usal.es

Outline

- Introduction
- Competitive learning algorithms
  - The Self-Organizing Map
  - The Visualization Induced SOM (ViSOM)
  - Quality Measures
- Ensembles and combinations
  - Fusion by Euclidean Distance
  - Weighted Voting Superposition (WeVoS)
- Experiments and Results
- Conclusions

Introduction

- <u>Objective: to obtain a clearer and more truthful representation of the structure of a multi-dimensional data set</u>
- This is very useful in several types of AI systems such as Case Based Reasoning Systems

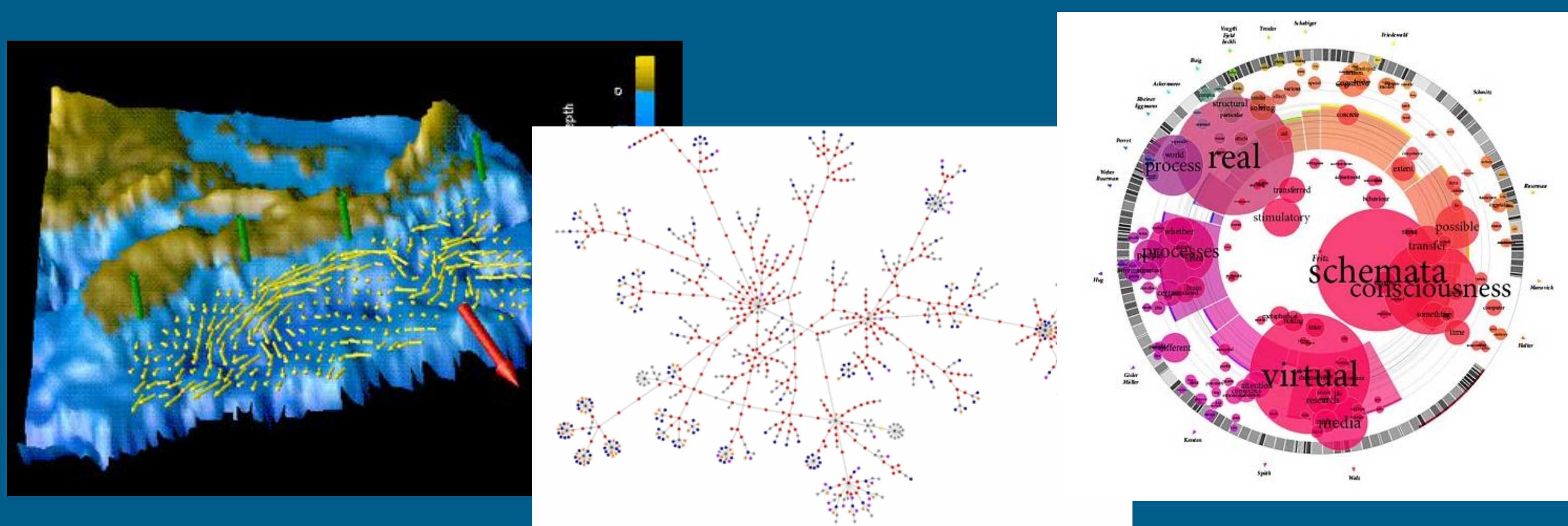


Fig. 1. Different ways of performing data visualization

Introduction

- Intended to use ANN
- Preferably Non-Supervised Learning
- Problems:
  - it is very difficult to asses the quality of a single map without comparing
  - the training of a map with the same dataset and parameters can yield quite different results

Outline

- Introduction
- Competitive learning algorithms
  - The Self-Organizing Map
  - The Visualization Induced SOM (ViSOM)
  - Quality Measures
- Ensembles and combinations
  - Fusion by Euclidean Distance
  - Weighted Voting Superposition (WeVoS)
- Experiments and Results
- Conclusions

Competitive Learing Alg.

SOM (Self-Organizing Maps)

- Objective
  - low dimensional representation (2-D)
  - preserve the topological properties of the input space
- Unsupervised Learning
- Neighbouring function makes close neurons activate for close ‘patterns’ in the input space

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

[Kohonen, 84]

Topology Preserving Maps

SOM *(Self-Organizing Maps)*

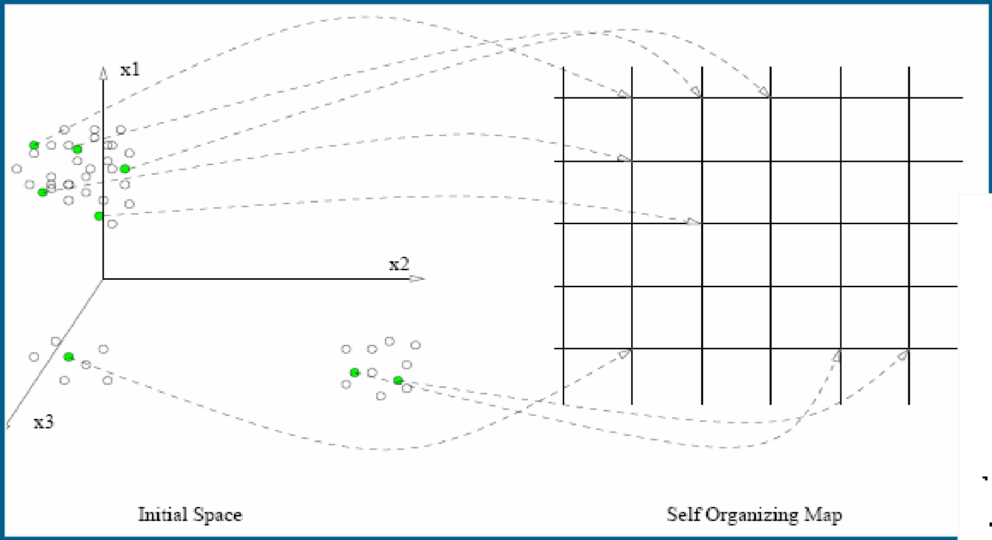
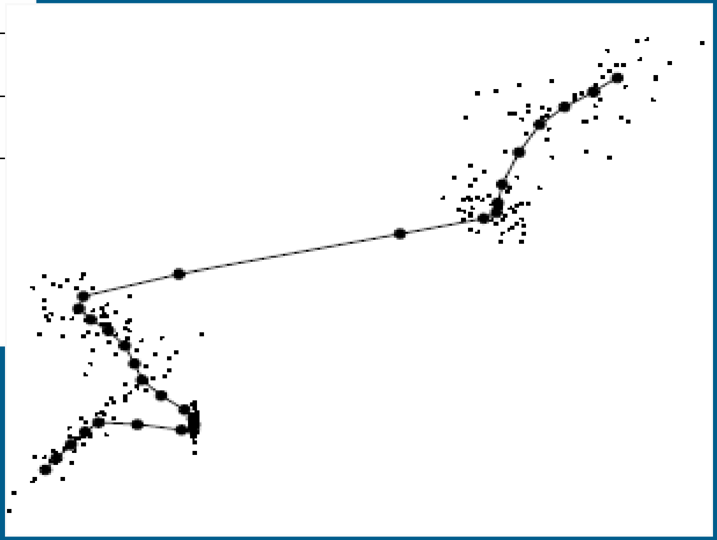


Fig. 3. 3D dataset mapped into a 2D map using a Self-Organizing Map

Fig. 4. 2D dataset mapped into a 1D map using a Self-Organizing Map

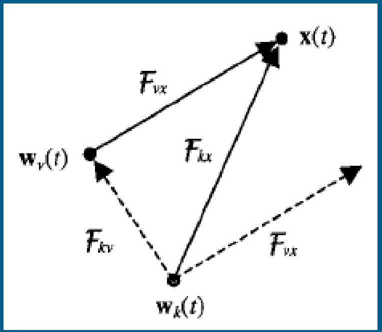


Topology Preserving Maps

ViSOM (Visualization Induced SOM)

- Objective
  - directly preserve the local distance information on the map, along with the topology
- Constrains the lateral contraction forces between neurons so that distances between neurons in the data space are in proportion to those in the input space

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



[Yin, 02]

Topology Preserving Maps

ViSOM (Visualization Induced SOM)

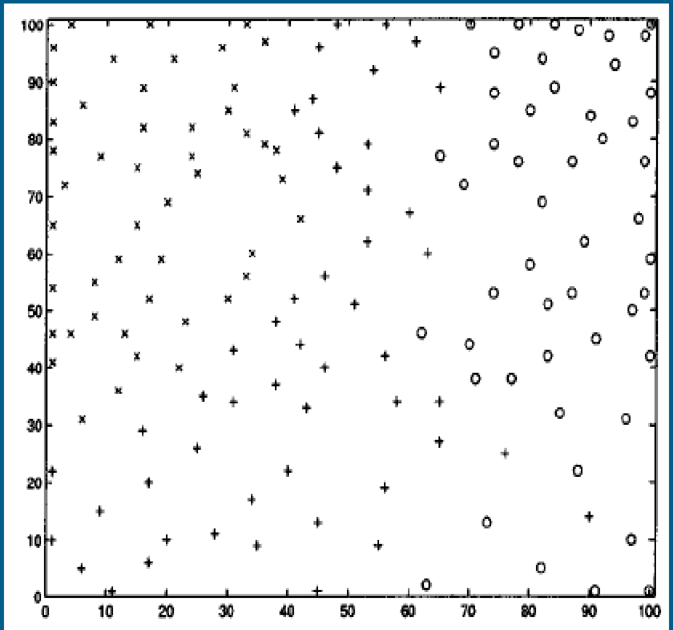


Fig 3. Iris dataset represented by a SOM

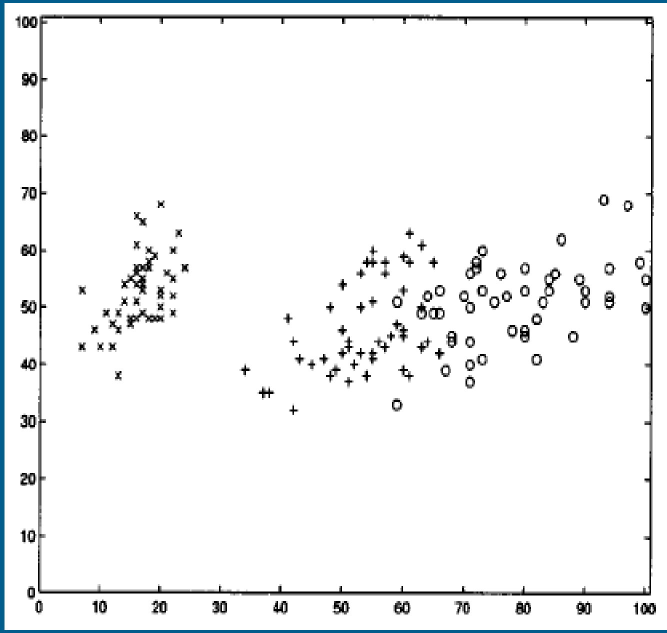


Fig 4. Iris dataset represented by a ViSOM

Topology Preserving Maps

Quality Measures

- Distortion
  - Measures in a more detailed way the topological order of the map

$$E_d = \sum_{x_i \in D} \sum_{w_k \in W} \eta(v_i, k) \left\| x_i - w_k \right\|^2$$

[Lampinen,92]

- Goodness of Approximation
  - measuring both the continuity of the mapping from the dataset to the map grid, and the accuracy of the map in representing the set

$$d(x_i) = \left\| x_i - v_i \right\| + \min \sum_{k=0}^{\left| K_{v'_i} \right| - 1} \left\| w_{Ii(k)} - w_{Ii(k+1)} \right\|$$

[Kaski & Lagus,96]

Outline

- Introduction
- Competitive learning algorithms
  - The Self-Organizing Map
  - The Visualization Induced SOM (ViSOM)
  - Quality Measures
- Ensembles and combinations
  - Fusion by Euclidean Distance
  - Weighted Voting Superposition (WeVoS)
- Experiments and Results
- Conclusions

Ensembles and Combinations

- The sets of patterns misclassified by the different classifiers would not necessarily overlap
- Different classifier designs offer complementary information about the patterns to be classified to improve the performance of the selected classifier
- Ensemble Summarization
  - Combination of answers of different classifiers (i.e. some kind of voting) Known as aggregation. **Not suitable for representation**
  - Combination of classifiers to obtain a final one (that should outperform the individual classifiers)

Ensembles and Combinations

Bagging

- Originally devised to improve classification accuracy
- Not suitable for representation
- Training of networks individually using slightly different datasets
- Datasets obtained using re-sampling with replacement
- Classification results obtained by weighted voting

Ensembles and Combinations

Bagging

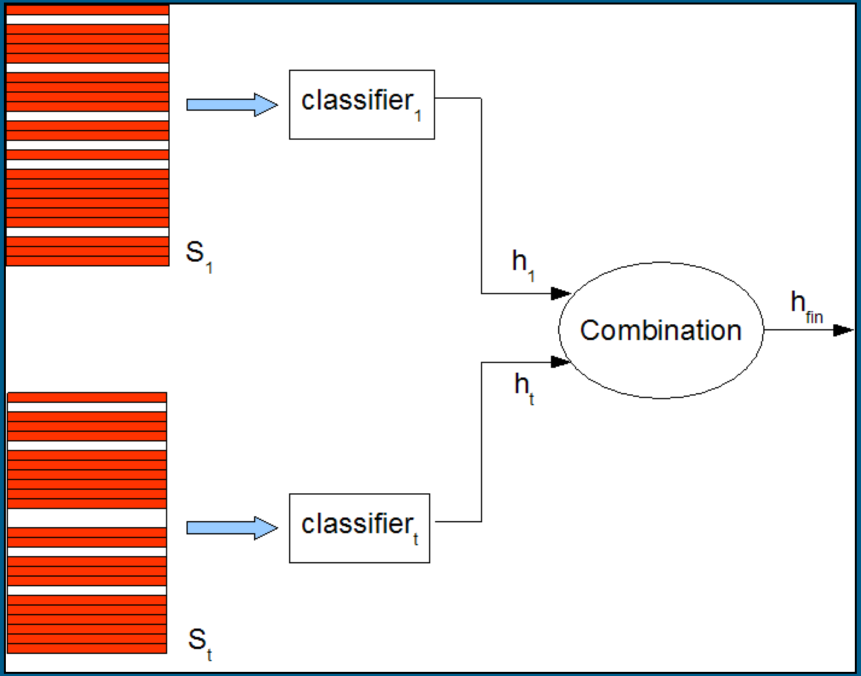


Fig 5. Scheme of the training of an ensemble using the bagging algorithm [Breiman,96]

Ensembles and Combinations

Fusion by Euclidean Distance

- Neurons must be aligned first.
- Based on Euclidean distance on neuron's weights and calculation of centroids.
- The set of data entries recognized for each neuron is updated as the **unions** of the sets corresponding to the clusters of neurons.

Ensembles and Combinations

Fusion by Euclidean Distance

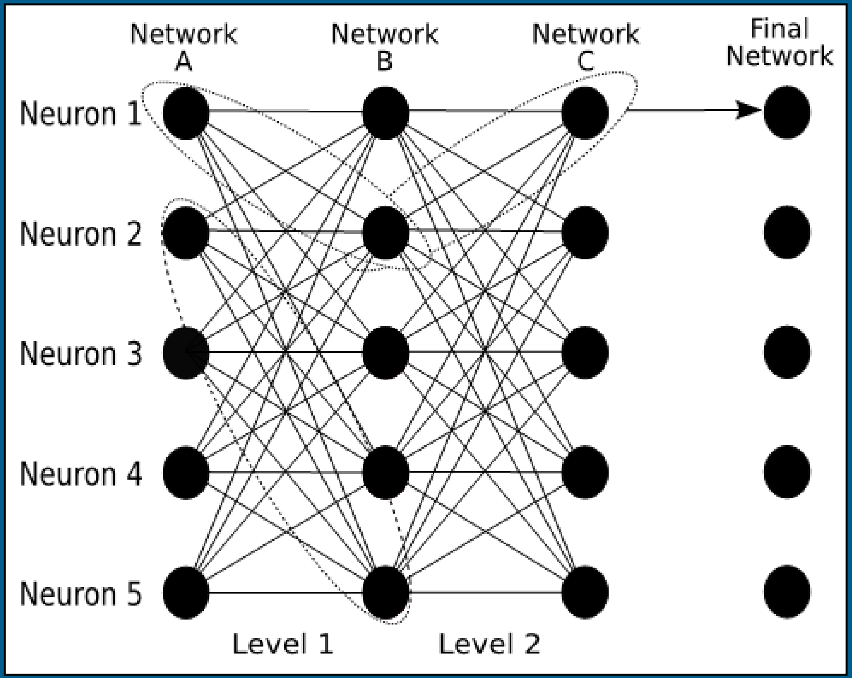


Fig 6. Alignment of three networks and their merging into one final map [Georgakis, 96]

Weighted Voting Superposition

- Objective:
  - to obtain a final map keeping the most important features of the maps composing the ensemble
- based on the calculation of the “quality of adaptation” of homologous units of different maps
- calculates the best adapted vector of characteristics in each of the units that make up the final map

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

Ensembles and Combinations

Weighted Voting Superposition

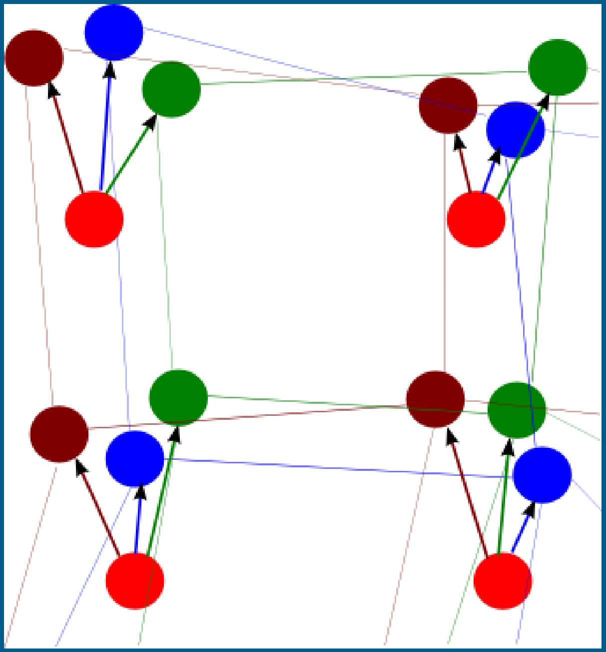


Fig. 5. Diagram showing how units of ensemble maps vote to determine the position of the final units

Outline

- Introduction
- Competitive learning algorithms
  - The Self-Organizing Map
  - The Visualization Induced SOM (ViSOM)
  - Quality Measures
- Ensembles and combinations
  - Fusion by Euclidean Distance
  - Weighted Voting Superposition (WeVoS)
- Experiments and Results
- Conclusions

Experiments and Results

- Initial Study
  - Well-known datasets obtained from the UCI Repository: Iris and Echocardiogram
- Practical case
  - Visualization of ham samples quality

Experiments and Results

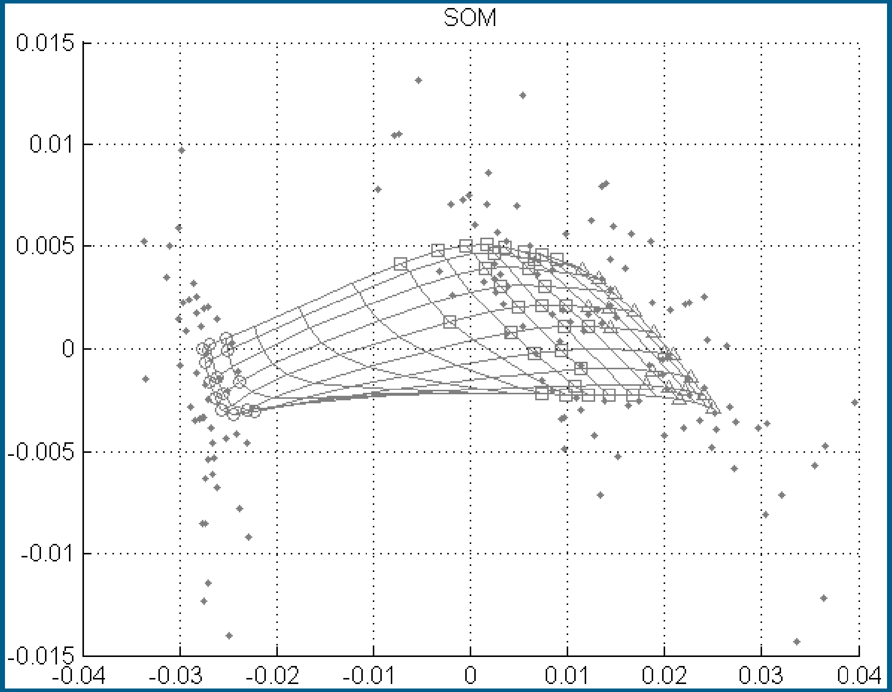


Fig. 7. SOM represented over the Iris Dataset (1st two Principal Components)

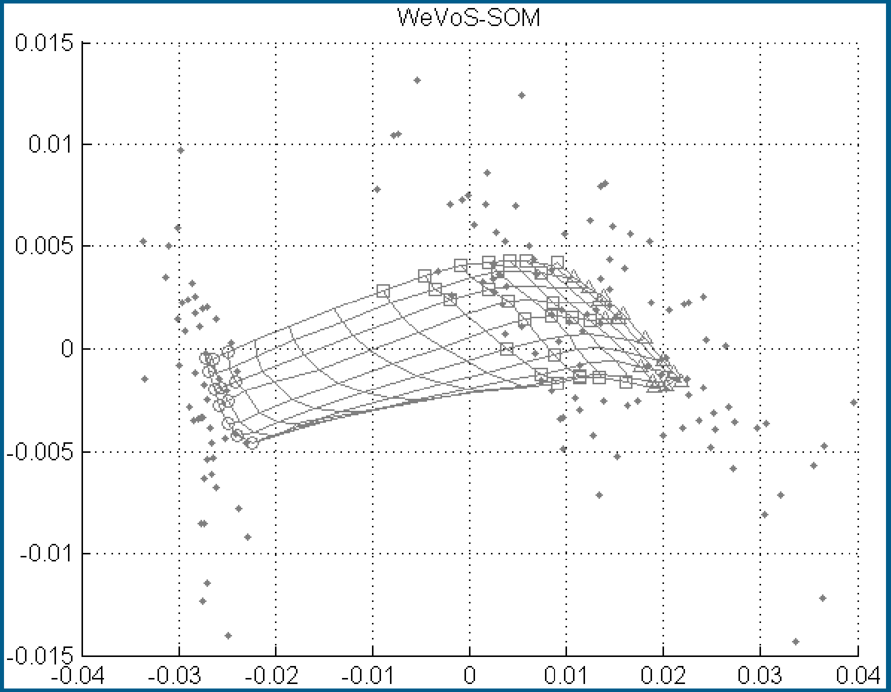


Fig. 10. WeVoS obtained from 5 SOM maps over the Iris Dataset (1st two Principal Components)

Experiments and Results

Introduction | Competitive Learning Algorithms | Ensembles and Combinations | Experiments and Results | Conclusions

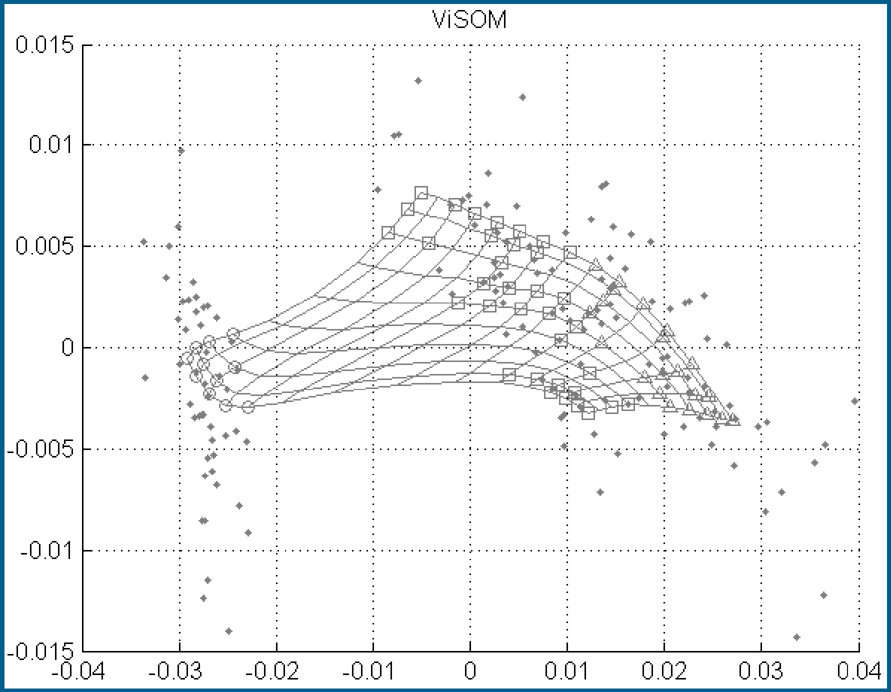


Fig. 9. ViSOM represented over the Iris Dataset (1st two Principal Components)

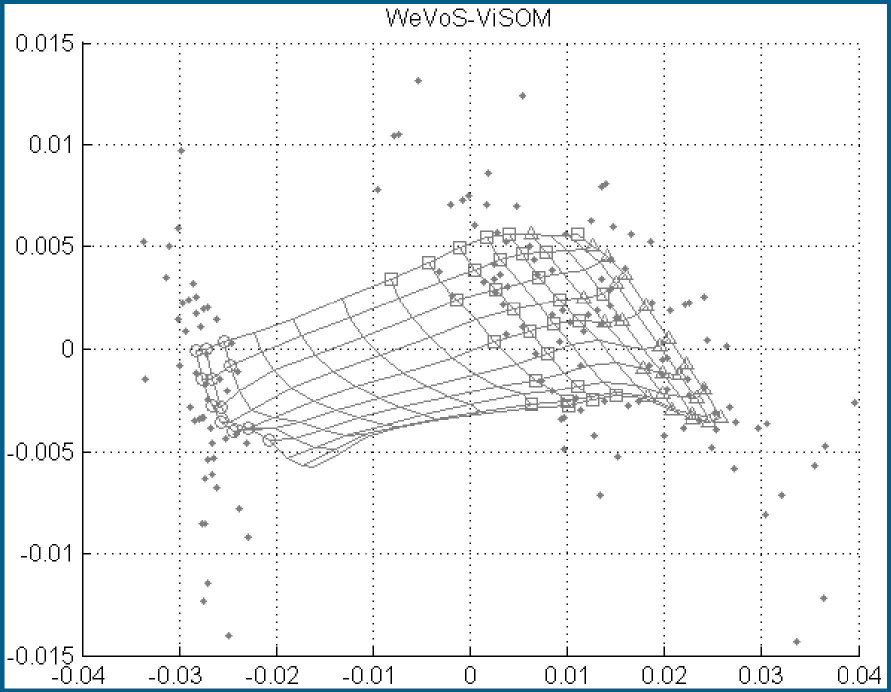


Fig. 10. WeVoS obtained from 5 ViSOM maps over the Iris Dataset (1st two Principal Components)

Experiments and Results

Introduction | Competitive Learning Algorithms | Ensembles and Combinations | **Experiments and Results** | Conclusions

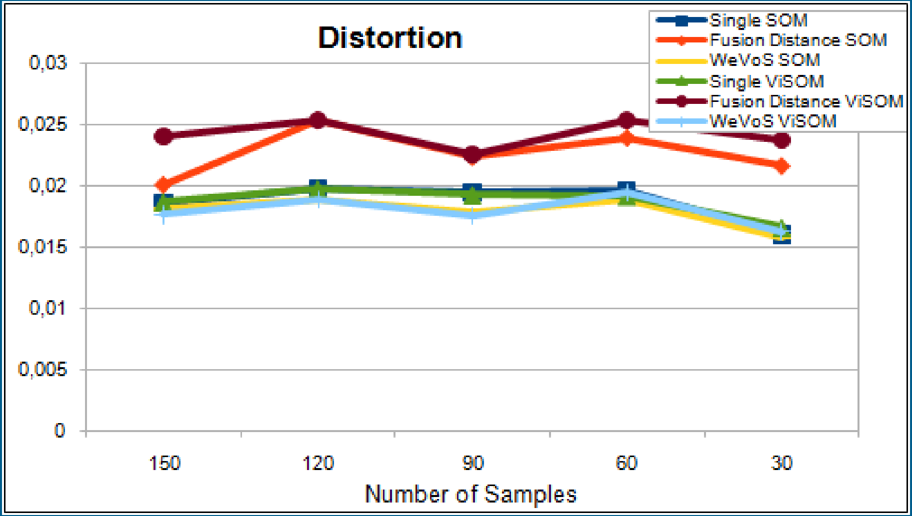


Fig. 11. Distortion for the models compared over the Iris Dataset

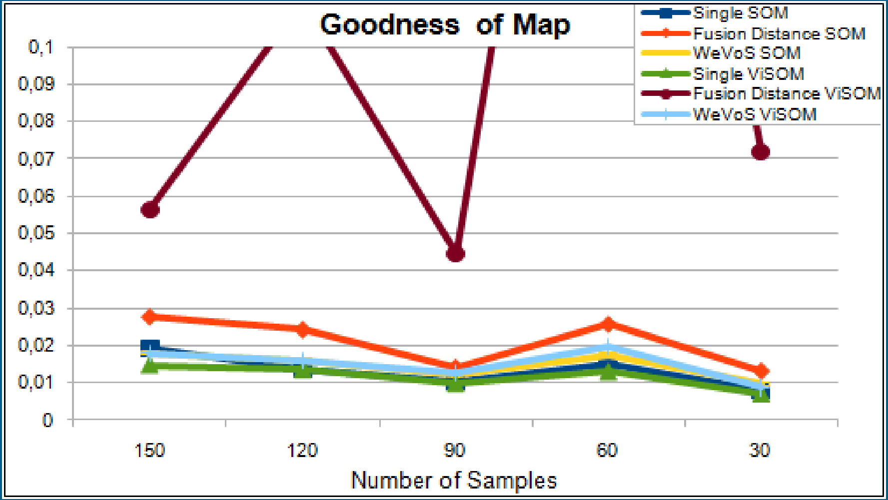


Fig. 12. Goodness of Map for the models compared over the Iris Dataset

Experiments and Results

Introduction | Competitive Learning Algorithms | Ensembles and Combinations | **Experiments and Results** | Conclusions

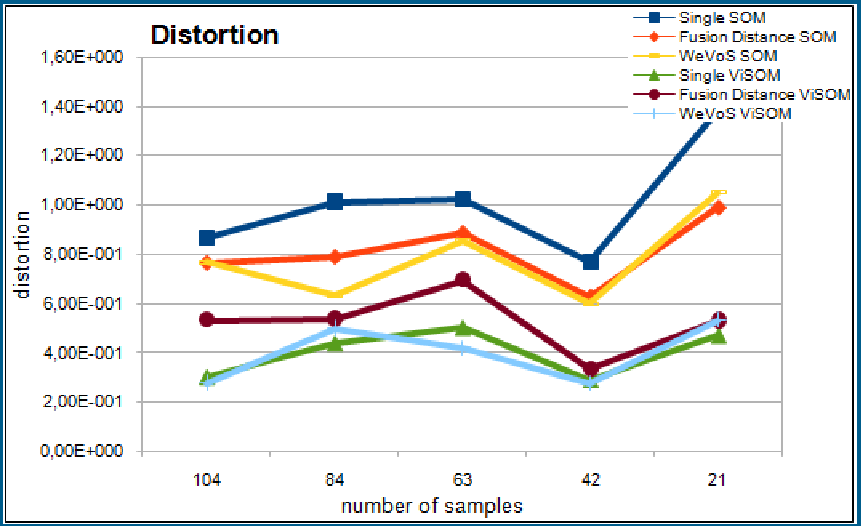


Fig. 13. Distortion for the models compared over the Echo-Cardiogram Dataset

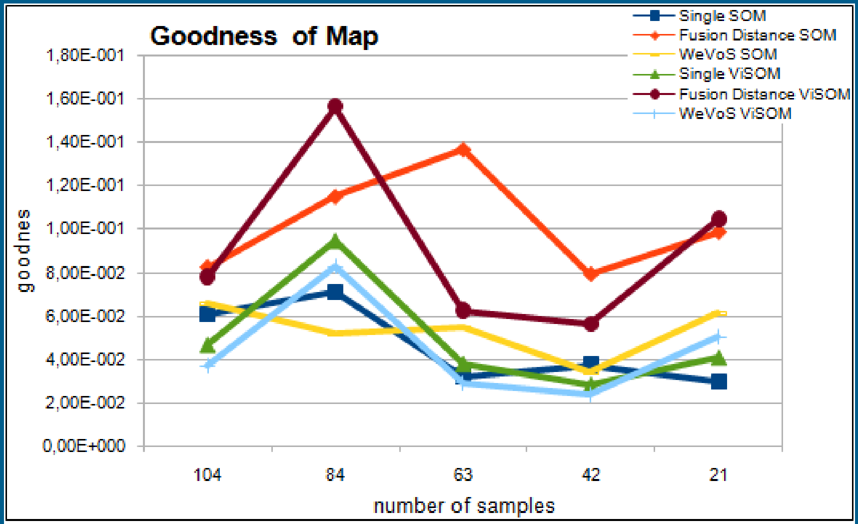


Fig. 14. Goodness of Map for the models compared over the Echo-Cardiogram Dataset

Experiments and Results

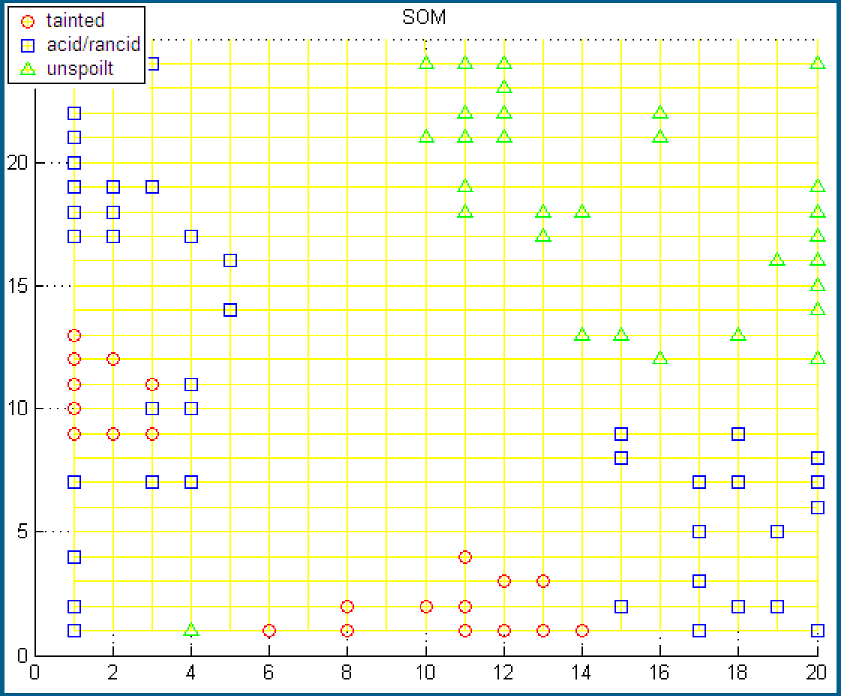


Fig. 15. Map of the Ham dataset created using a single SOM

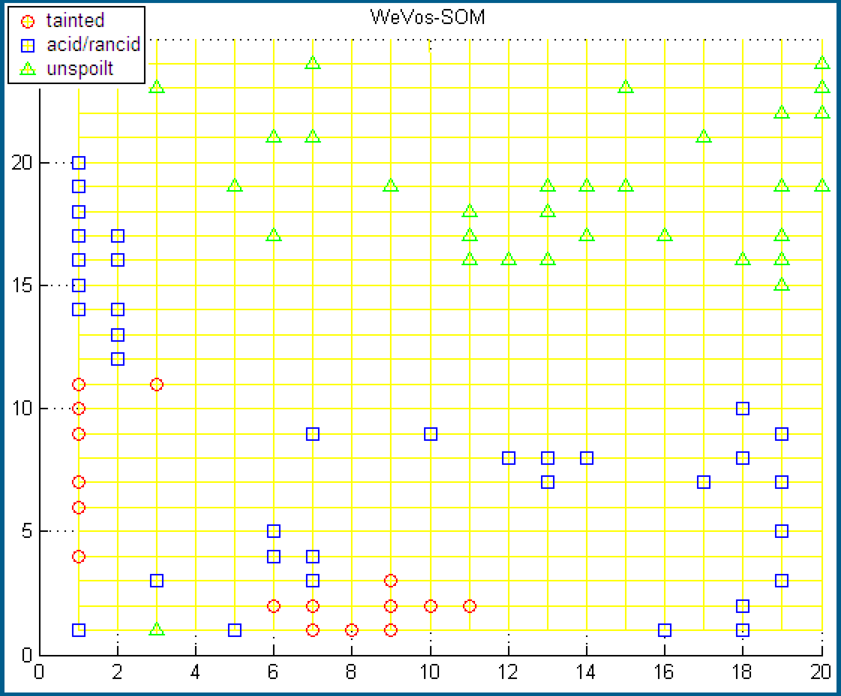


Fig. 16. Map of the Ham dataset created using an ensemble of 5 SOMs and the WeVoS summarizing algorithm

Experiments and Results

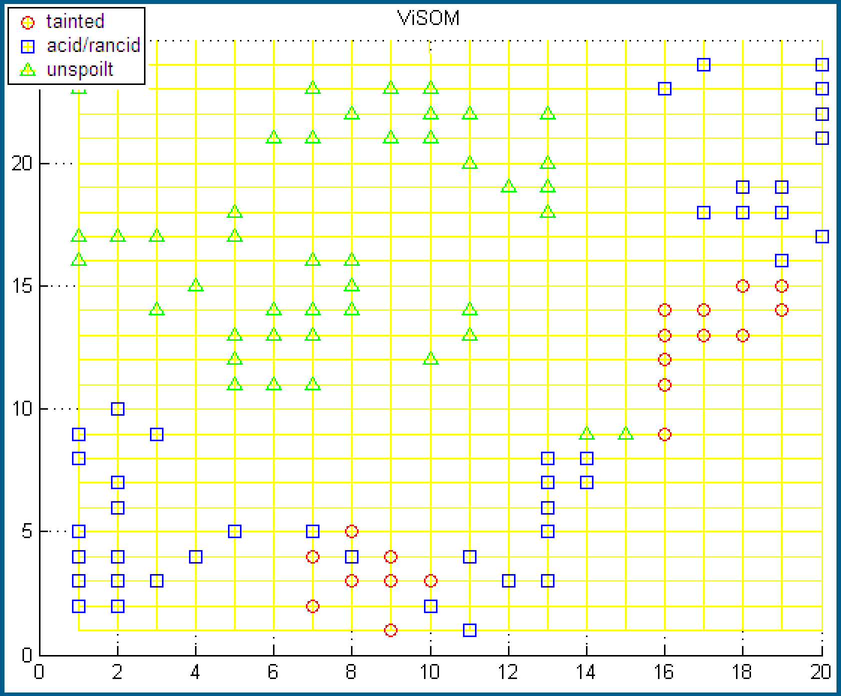


Fig. 17. Map of the Ham dataset created using a single ViSOM

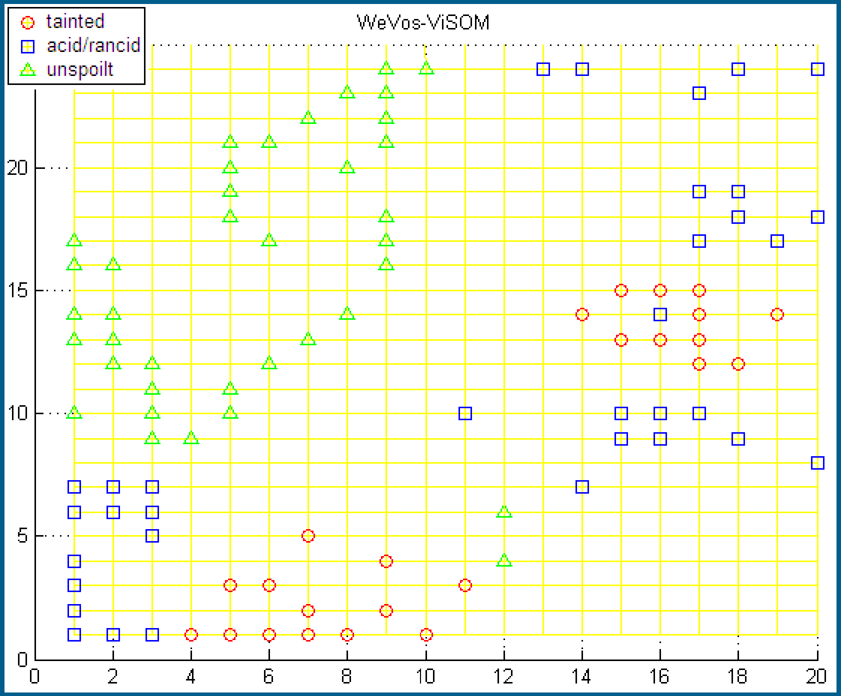


Fig. 18. Map of the Ham dataset created using an ensemble of 5 ViSOMs and the WeVoS summarizing algorithm

Outline

- Introduction
- Competitive learning algorithms
  - The Self-Organizing Map
  - The Visualization Induced SOM (ViSOM)
  - Quality Measures
- Ensembles and combinations
  - Fusion by Euclidean Distance
  - Weighted Voting Superposition (WeVoS)
- Experiments and Results
- Conclusions

Conclusions

- A novel technique of obtaining fusion of unsupervised visualization algorithms has been presented.
- It has been used for the first time with more advanced models than the single SOM (ViSOM) with good results.
  - Lower distortion errors. Slightly increased quantization error
- The model proves its full potential with more difficult to analyse datasets (faulty or insufficient data)

Conclusions

- Useful also for organizing information for more accurate and faster retrieval (i.e. CBR systems)
- Future Work Includes:
  - More complex ensemble training algorithms
  - Comparison on a wider range of datasets
  - Use of other similar competitive learning algorithms

Questions ?

- Thank you for your attention !