

UNSUPERVISED HORIZONTAL COLLABORATION BASED IN SOM

Nadia Mesghouni¹ Khaled Ghedira² and Moncef Temani³

¹University of Tunis Laboratory LI3 ISG Tunis, Tunis, Tunisie
nadia.mesghouni@yahoo.fr

²University of Tunis Laboratory LI3 ISG Tunis, Tunis, Tunisie
Khaled.ghedira@isg.rnu.tn

³University of Tunis Laboratory LI3 ISG Tunis, Tunis, Tunisie
Moncef.temanni@fst.rnu.tn

ABSTRACT

In this paper we present a new approach of collaborative classification allowing protecting the confidentiality of the data by using the self organizing map of Kohonen.

Having a collection of databases distributed on several different sites, so, the problem consists in clustering each of these bases by considering the data and the classifications of the others base co-workers, without omitting however to respect the constraint of confidentiality which forbids the sharing of data between the various centers. To do it, our approach is subdivided into two phases: a local phase and a collaborative phase.

The local phase would mean applying the classic algorithm of Kohonen, locally and independently on each of the databases, what will end in the obtaining of a map (SELF-ORGANIZING MAP) for each of these bases. The phase of collaboration would mean making each of the databases collaborate with all the map SOM partners in the other bases obtained during the local phase.

So, as result we obtain on each of the sites a map close to the SOM which we would have obtained if we had disregarded the constraint of confidentiality, namely make databases collaborate they same. In the stemming from both phases, all the maps will be enriched. The article presents the formalism of the approach as well as its validation.

The proposed approach was validated on several databases and the experimental results showed very promising performances.

KEYWORDS

Self organizing Map, unsupervised learning, collaborative classification, confidentiality of data

1. INTRODUCTION

In an industrial context of increasing competition, companies are constantly called up on to work together (to collaborate) to face up to this strategic issue and to be able to provide the level of required service for their customers. To benefit from this collaboration, the tasks of Data Mining (Clustering, mining and knowledge management ...) should consider all the datasets associated with these collaborating companies although they are distributed on several different sites.

Obviously, for confidentiality reasons (ex. medical or bank data), sharing data between collaborating companies is not allowed. So, centralizing their data by combining them into one dataset and then performing the task of Data Mining is not appropriate.

In this paper, we are interested in the problem of clustering and specifically in collaborative clustering preserving data confidentiality and using self organizing maps of Kohonen.[1]

The rest of this article is organized as follows: we'll present our vertical collaboration approach in section 2, after introducing the problem of collaborative clustering. In Section 3, we'll present different results. Finally, we completed by the conclusion.

2. Collaborative Clustering

Collaboration between the companies has intensified in recent years and has become one of the usual corporate strategies. Often, large local companies outsource the manufacture of certain parts or the provision of certain services to small businesses. Moreover, there are small companies in the same activity sector or even in the same industry forming between them strategic organizations which favours flexible specialization and collective efficiency.

In both cases, to better take advantage of collaboration and clustering, the collaborating Companies must have a global clustering result considering all their data. Obviously, for reasons of confidentiality, data sharing between these companies is forbidden, which prevents centralizing their data by combining them into one database and perform clustering on the latter. Thus, in literature [2,3,4] some studies have been proposed which allow collaboration of several distributed datasets over several different sites while preserving the confidentiality of these data. According to the structure of collaborated datasets, there are three main types of collaboration: horizontal, vertical and hybrid collaboration. In this work, we are particularly interested in horizontal topological collaborative clustering inspired by the work of fuzzy collaborative c-means [5] and introduce the notion of self organizing map of Kohonen [6].

We will introduce our approach in the dimensionality reduction, but imagine the case when we have a very large data set and all the features a pertinent and no reduction can be made. In this case, to accelerate the data mining of these data sets we can separate the data in some subsets (using a separate collaborative function) and then we can use the collaborative clustering to cluster this data.

Collaborative clustering was first investigated by Pedrycz [7, 8, 9, 10, 11], using a fuzzy k-means algorithm. The fundamental concept of collaboration is: “the clustering algorithms operate locally (namely, on individual data sets) but collaborate by exchanging information about their findings” Pedrycz.

2.1. Topological unsupervised horizontal collaborative clustering

In the case of horizontal clustering, all data sets describe the same observations. So, all these collaborative datasets have the same number of observations but a different number of variables. What we would like is that after the collaboration, if an observation of the i^{th} data set is projected onto the j^{th} neuron of the ii^{th} SOM map, then the same observation of the jj^{th} data set is projected on the same neuron j in the jj^{th} map or on one of the neighboring neurons.

In other words, neurons that correspond to different maps should capture the same observations. That’s why we have added a term to the objective function of the classical SOM learning algorithm in order to approximate the neurons tow hich an individual belongs on all the maps. We weighted this function by the collaboration parameter α which is set by the expert depending on the confidence of the collaborated map.

Formally, we have achieved the following new objective function:

$$R_H^{[ii]}(\mathcal{X}, w) = R_{SOM}^{[ii]}(\mathcal{X}, w) + R_{Col_H}^{[ii]}(\mathcal{X}, w) \quad (1)$$

with

$$R_{SOM}^{[ii]}(\mathcal{X}, w) = \sum_{i=1}^N \sum_{j=1}^{|w|} K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2$$

$$R_{Col_H}^{[ii]}(\mathcal{X}, w) = \sum_{jj=1, jj \neq ii}^P \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \left(K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} - K_{\delta(j, \mathcal{X}(x_i))}^{[jj]} \right)^2 \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2$$

Where P is the number of data sets, N –the number of observation son the dataset and ii is the same for all bases, |w| is the number of prototype vectors of the map ii. This function is minimized for each dataset during the collaboration step.

Figure1 shows the schema of horizontal collaboration between multiple maps from several datasets.

The minimization of the horizontal collaborative clustering

$$R_{SOM}^{[ii]}(\mathcal{X}, w) = \sum_{i=1}^N \sum_{j=1}^{|w|} K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2 + \sum_{jj=1, jj \neq ii}^P \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{|w|} \left(K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} - K_{\delta(j, \mathcal{X}(x_i))}^{[jj]} \right)^2 \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2$$

So, we have $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ et $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})$ and the $R^{HCol}_{SOM}[ii]$ function can be rewritten as following:

$$R_{SOM}^{[ii]}(\mathcal{X}, w) = \sum_{i=1}^N \sum_{j=1}^{|w|} K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} \sum_{k=1}^n [ii] \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2 + \sum_{jj=1, jj \neq ii}^P \alpha_{[ii]}^{[jj]} \sum_{i=1}^N \sum_{j=1}^{\bar{w}} \left(K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} - K_{\delta(j, \mathcal{X}(x_i))}^{[jj]} \right)^2 \sum_{k=1}^{n[ii]} \left\| x_{ik} [ii] - w_{jk} [ii] \right\|^2$$

(1)

The necessary condition for finding the minimum of the :

$$\begin{aligned}
 & R_{som}^{HCol} [ii] s \nabla w [ii] R_{SOM}^{HCol} [ii] = 0 \\
 & \frac{\partial R_{SOM}^{HCol}}{\partial w_{jk} [ii]} = 0 \Rightarrow -2 \sum_{i=1}^N K_{j, \mathcal{X}(xi)} (x_{ik} - w_{jk} [ii]) - 2 \\
 & \sum_{jj=1, jj \neq ii}^P \sum_{i=1, N} \alpha [ii, jj] (K_{j, \mathcal{X}(xi)} [ii] - K_{j, \mathcal{X}(xi)} [jj])^2 \\
 & (x_{ik} [ii] - w_{jk} [ii]) = 0 \dots (1) \\
 & (1) \Rightarrow \sum_{i=1}^N K_{j, \mathcal{X}(xi)} (x_{ik} - w_{jk} [ii]) + \\
 & + \sum_{jj=1, jj \neq ii}^P \sum_{i=1, N} \alpha [ii, jj] (K_{j, \mathcal{X}(xi)} [ii] - K_{j, \mathcal{X}(xi)} [jj])^2 \\
 & (x_{ik} [ii] - w_{jk} [ii]) = 0 \\
 & (1) \Rightarrow \sum_{i=1}^N k_{s, \mathcal{X}(xi)} [ii] - \sum_{i=1}^N k_{s, \mathcal{X}(xi)} w_{sk} [ii] + \\
 & \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2 \\
 & x_{ik} [ii] - \sum_{i=1}^N [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2 w_{jk} [ii] \\
 & (1) \Rightarrow w_{jk} [ii] \left(\sum_{i=1}^N k_{j, \mathcal{X}(xi)} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2 \right) = \\
 & \sum_{i=1}^n k_{j, \mathcal{X}(xi)} x_{ik} [ii] + \\
 & + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2 x_{ik} [ii] \\
 & (1) \Rightarrow w_{jk} [ii] = \\
 & \frac{\sum_{i=1}^n k_{j, \mathcal{X}(xi)} x_{ik} [ii] + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2 x_{ik} [ii]}{\sum_{i=1}^N k_{j, \mathcal{X}(xi)} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha [ii, jj] (k_{j, \mathcal{X}(xi)} [ii] - k_{j, \mathcal{X}(xi)} [jj])^2}
 \end{aligned}$$

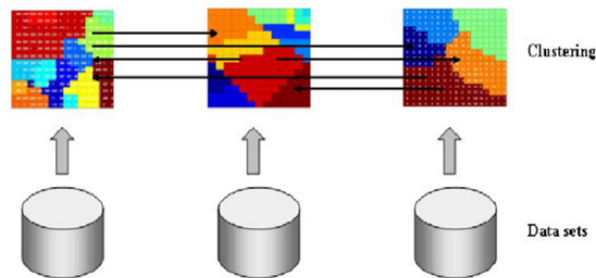


Figure 1: Horizontal collaborative SOM

The topological horizontal collaboration clustering algorithm is summarized in the procedure 20. In this procedure the two steps: local and collaborative are indicated. We note, that in case of collaborating of existed maps (clustering results) the algorithms cape the local phase because the map existed already and will start directly the collaboration step.

Algorithm1: Topological-horizontal collaboration Algorithm

Initialization:

Initialize all the map prototypes W randomly. Set the matrix of collaboration $\alpha[ii, jj]$

1. Local step:

for each DB, $X[ii], ii= 1$ to P do

Minimize the objective function of the classical SOM:

$$R_{SOM}^{[ii]}(\mathcal{X}, w) = \sum_{i=1}^N \sum_{j=1}^{|w|} K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} \left\| x_i^{[ii]} - w_j^{[ii]} \right\|^2 \quad (2)$$

For $t= 1$ to T_{max} (T_{max} indicates the number of iterations) **do**

1. Learning step:

At each iteration t choose an input $x(t)$ in general, randomly, and present it to the map.

2. Competition step:

Choose the best matching unit (BMU) i^* by computing $\|x_i - w_j(t)\|^2$

3. Updating step:

Update the winner neuron i^* and its neighbors: $w_j(t+1) = w_j + \varepsilon(t) K_{j, \mathcal{X}(x_i)} (x_i - w_j(t))$

Decreasing the size of the bmus neighborhood area and the learning coefficient $\varepsilon(t) = 0$.

end for

end for

2. Collaborative step:

for each DBX $[ii], ii= 1$ to P **do**

Minimize the horizontal collaborative learning objective function (1)

Compute prototypes of the new ii -th map:

for $s = 1$ to $|w|$ **do**

for $t = 1$ to $|x|$ **do**

$$w_{jk}^{*ii} = \frac{\sum_{i=1}^{N^{[ii]}} K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} x_{ik}^{[ii]} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^{N^{[ii]}} \alpha_{[ii]}^{[jj]} \left(K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} - K_{\delta(j, \mathcal{X}(x_i))}^{[jj]} \right)^2 w_{ik}^{[jj]}}{\sum_{i=1}^N K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N \alpha_{[ii]}^{[jj]} \left(K_{\delta(j, \mathcal{X}(x_i))}^{[ii]} - K_{\delta(j, \mathcal{X}(x_i))}^{[jj]} \right)^2} \quad (3)$$

end for

end for

end for

3. EXPERIMENTAL RESULTS

To validate our approach, we computed the quantization error (distortion) on several maps of different size. To show the importance of our approach to the clustering problem, we calculated the purity index of each map.[12 , 13].

The purity of a neuron is the percentage of data belonging to the majority class. Assuming knowledge of all classes of data $L = (l_1, l_2, \dots, l_{|L|})$ and all the neurons $C = (c_1, c_2, \dots, c_{|C|})$, the term which expresses the purity of a card is defined as follows:

$$purity = \sum_{k=1}^{|C|} \frac{|C_k|}{k} \max_{i=1}^{|L|} \frac{|C_{k,i}|}{|C_k|} \quad (4)$$

Where $|C_k|$ represents the total number of data associated with the neuron C_k , $|C_{k,i}|$ represents the number of data class l_i which are associated with neuron c_k and n the total number of data.

The purity of the map is equal to the average purity of neurons. A good SOM map should have a greater purity.

3.1 Results on the waveform dataset

We use the waveform dataset [14] in order to show the improvement of the collaborative topological clustering approaches because this dataset has twenty noisy features added to the data, and allows us to provide a visualization of each result. We note that the analysis of these results must be made in a color mode. The computed validation indices are shown in tables 1.

Waveform data set: This data set consists of 5000 instances divided into 3 classes.

The original base included 40 variables, 19 are all noise attributes with mean 0 and variance 1. Each class is generated from a combination of 2 of 3 "base" waves.

3.2 Improvement of the horizontal approach

In order to have two datasets with the same observations but described by different variables, we divided the Waveform dataset size 5000 x 40 into four databases: the first and the second part of the dataset (5000x10; 5000x10) which corresponds to all the relevant variables [1,2,...,20]; the second and third part (5000 x10; 5000 x10) containing the noise of the base - the set of variables [21,22,...,40]. We use these datasets to show the entire process that will enable collaboration of all datasets in a horizontal manner.

The collaboration matrix α was set initially a $t \alpha = [11, 11]$. We selected maps of dimension 10x10, and then we execute the local step of the collaboration clustering algorithm on the first dataset which is to learn a SOM map for all the observations of this dataset. Another map has been used to learn the observations of the second dataset during the local learning step. Figure 2(a) and Figure 2(b) present respectively the prototype vectors obtained on the first and second map after the local learning step. The two axes X and Y represent respectively the indices of variables and prototypes. We then construct another map for the 1st dataset by collaborating with the second map (SOM2) and obtained a map presented in the figure 2(c). The same collaboration was done for the 2nd dataset in collaboration with the first map (SOM1) and we obtained the map21 which is shown in the figure 2(d).

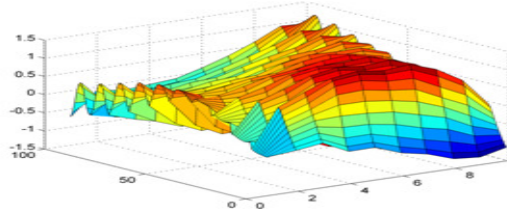
If we compare the first local map (SOM1) which has a purity index equal to 75.71 and the one obtained after the collaboration (SOM12), we can notice that the second local map has influenced the latter. It is thus that, due to the influence of the high important variables from the

second local map (SOM2) having 79.61 purity index, the variables [1– 4] of the collaborated SOM12 map has more importance (red color) compared to the SOM1 map obtained from the same dataset. The collaborated SOM12 map purity increase to 76.21 compared to the first local map thanks to the SOM2 map with which the map collaborated.

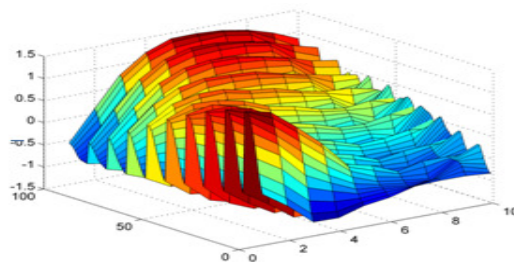
And similarly, the first our variables from the collaborated SOM21 map has less importance Compared to the local SOM2 map due to the collaboration with the first map (SOM1) where The variables [1–4]has less importance compared to the SOM2local map, and respectively the accuracy index decrease to 78.72 as a result of the collaboration with a map which has a smaller accuracy. The same analysis can be done for the learning quantization errors of these maps. The respective purity indexes and quantization errors can be find in the table1 from this chapter.

Table 1 summarizes the evaluation criteria associated with these maps. The local step of the collaboration approach was done also for the third and fourth datasets which contains noise variables. The figure 4shows the both local maps SOM3 and SOM4 issued from these dataset which shows the noise features for all the prototypes. Due to the features noise in these datasets, the purity indices a very smaller compared to the first two maps, and are equals to 47.19 for SOM3 map and to 51.26 for SOM4 map (table1) Now, we will collaborate the local SOM1 map (figure 2(a)) with local the SOM3 map (figure 3(a)) in order to see the

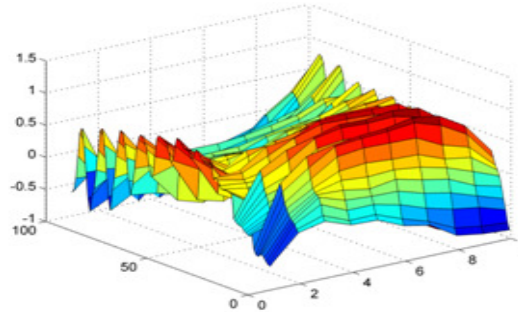
influence of the noise features on the important features and vice versa. On the SOM13 map (figure 4(a)) obtained after the collaboration with the SOM3 map, we can see that the high important variables [6–10] reduced the importance due to the collaboration with the noise variables compared to the local SOM1map and the purity index decrease considerably to 62.47(from75.71). Contrarily, the collaborated SOM31map



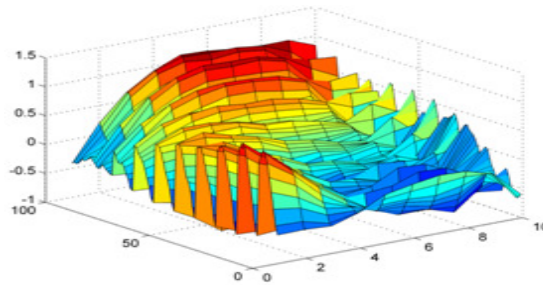
(a) The map of the 1st dataset before collaboration : SOM1



(b) The map of the 2nd dataset before collaboration: SOM2



(c) The map of the 1st dataset after collaboration with SOM2 : SOM12



(d) The map of the 2nd dataset after collaboration with SOM1 : SOM21

Figure 2: Horizontal collaboration between the datasets 1 and 3

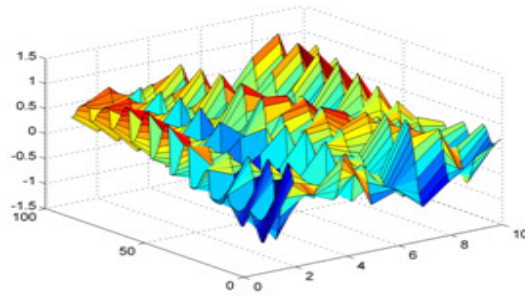
(figure 4 (b)) has increase only a little the importance of the features [1 – 4] compared to the local SOM3 map (the accuracy index increase to 54.63 from 47.19)

We, will made the same collaboration process between the SOM1 local map containing relevant variables and the local SOM4 map containing noise variables. As for the SOM13 map, the noise SOM4 map influenced the SOM1 map but a little less than collaborating with the SOM3 map. The relevant features decrease their importance after the collaboration due to the noise features presented in the local SOM4 map (figure 5). Relevant variables from can't increase the importance of features from the 4th map, to do this we need to increase the confidence parameter α for the 1st map. The purity indices and quantization errors which explain these results are shown in the table 1.

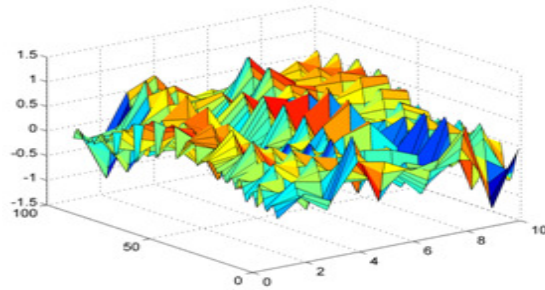
In the (figure 6) we shows the collaboration between two maps SOM14 and SOM21 obtained after the collaboration between 1st and 4th maps; and between the 2nd and 1st maps respectively.

This result shows that in the case of collaborating of two maps containing relevant variables, then the obtained collaborated map will contains high relevant variables which will increase the map accuracy (table 4.1). We remind that red color represent important variables for each map's prototype. We note obtained map SOM14-21.

As, the SOM14-21 map contains high relevant features, we will collaborate this one with the SOM41 map containing noise variables, but increasing the confidence parameter α to 8

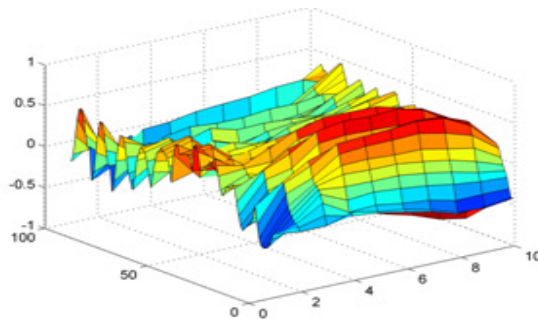


(a) The map of the 3th dataset before collaboration: SOM3

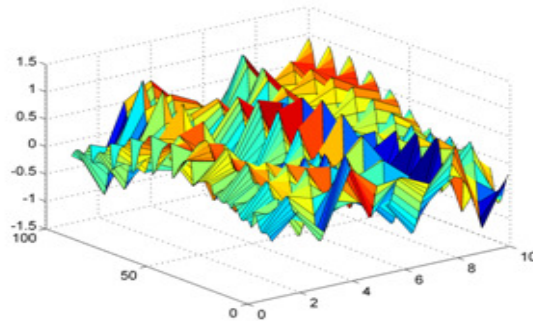


(b) The map of the 4th dataset before collaboration: SOM4

Figure 3: SOM maps for the dataset 3 and 4 **before** collaboration

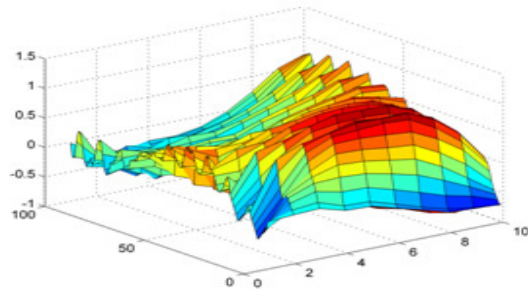


(a) The map of the 1st dataset after collaboration with SOM3 : SOM13

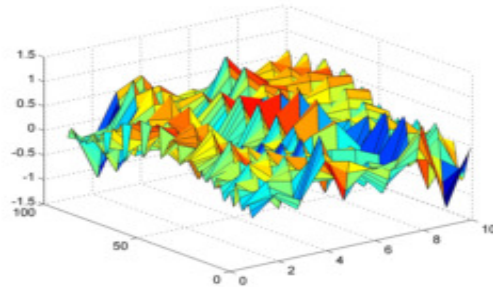


(b) The map of the 3th dataset after collaboration with SOM1 : SOM31

Figure 4: Horizontal collaboration between the **datasets** 1 and 3



(a) The map of the 1st dataset after collaboration with SOM4 : SOM14



(b) The map of the 4th dataset after collaboration with SOM1 : SOM41

Figure 5: Horizontal collaboration for **datasets 1 and 4**

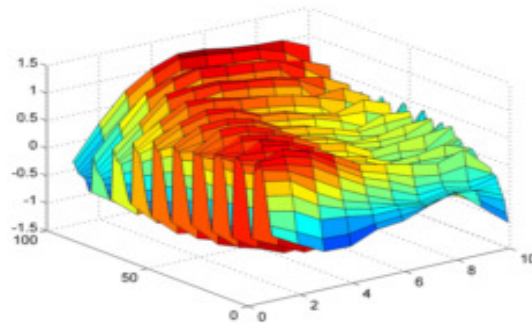


Figure 6: The SOM14 map after collaboration with SOM21 : SOM14-21

For the SOM14-21 map and decreasing α to 0.2 for the SOM41 map in order to weight the learning to take more into account the SOM14 map. As we can see on the (figure 7), the collaborated obtained map changed its structure by increasing the importance of features [4–9], that means that the important features from the first map high influence don't be noise features, and increase the purity index to 66.84, and decreasing the quantization error to 2.01.

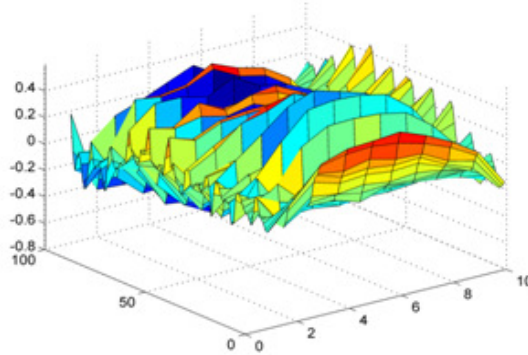


Figure 7: The SOM14-21->41 map after collaboration of SOM14-21 with SOM41: SOM14-21->41

Table 1: Experimental validation of the horizontal collaboration

Measures	Purity	Quantization error
SOM1	75.71	1.98
SOM2	79.61	1.87
SOM3	47.19	2.64
SOM4	51.26	2.41
SOM12	76.21	1.93
SOM21	78.72	1.81
SOM13	62.47	2.14
SOM31	54.63	2.27
SOM14	64.17	2.05
SOM41	56.18	2.35
SOM41- 21->41	66.84	2.01

From Table 1, we can notice that then collaborating a map with a higher purity index with another which has a smaller purity, the accuracy of the first one will decrease and vice versa. The same situation can be analyzed for the quantization error, where it will increase while collaborating a map with another which has a bigger quantization error. This is the main Enigma of the collaborating technique: to know when the collaborating process will increase clustering validation indices; but in the unsupervised learning, computing these indices are not usually possible, so, the main goal is to make collaborating the maps.

4. CONCLUSION

In this study, we have proposed approach for clustering of multiple databases derived from several datasets using Kohonen's self-organizing maps: horizontal collaborative SOM.

The horizontal collaborative SOM approach is adapted for horizontal collaboration datasets that describe the same observations but with different variables.

Thanks to the principle of this approach, data confidentiality is preserved. During the collaboration phase, each data set is collaborated with all the maps obtained during the local phase.

Thus, each site uses its dataset and the information from other SOM maps, there by learning a new map that is similar to the map that would be obtained if we had centralized Datasets and then clustering it. Thus, the SOM maps obtained after the collaboration step are similar. This

approach has been validated on multiple datasets and experimental results have shown very promising performance.

REFERENCES

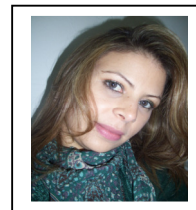
- [1] T.Kohonen. Self-organizing maps. Springer berlin, 2001.
- [2] Costa da silva j. & klusch m. (2006). Inference in distributed data clustering, eng. Appl. Artificial intelligence 19, p. 363- 69
- [3] Pedrycz w. (2002) collaborative fuzzy clustering, pattern recognition lett. 23, p. 675-686.
- [4] Strehl a., ghosh j. (2002). Cluster ensembles: a knowledge reuse framework for combining multiple partitions. Journal of machine learning research (jmlr), 3, p. 583 – 617
- [5] Pedrycz w. & rai p. (2008). Collaborative clustering with the use of fuzzy c-means and its quantification, fuzzy sets and systems, doi: 10.1016/j.fss.2007.12.030.
- [6] Kohonen t. (1989). Self-organization and associative memory, heidelberg: springer-verlag, berlin, 3rd edition.
- [7] Witold pedrycz. Collaborative fuzzy clustering. Pattern recognition letters,23(14):1675–1686, 2002.
Witold pedrycz. Fuzzy clustering with a knowledge –based guidance. Patternre cogn. Lett., 25(4):469–480, 2004.
Witold pedrycz. Interpretation of clusters in the framework of shadowed sets. Pattern recogn.lett., 26(15):2439–2449, 2005.
- [8] Sushmita mitra, haider banka, and witold pedrycz. Collaborative rough clustering. In premi,pages768–773,2005.
- [9] Witold pedrycz and kaoru hirota. A consensus-driven fuzzy clustering. Pattern recogn.lett., 29(9):1333–1343, 2008.
- [10] Vesanto j. Et alhoniemi e. (2000). Clustering of the self organizing map, iee transactions on neural networks 11 (3), p.586-600
- [11] Jain a., murt m., flynn p. (1999). Data clustering: a review, acm comput. Surveys 31, p. 264-323
- [12] Asuncion a. & newman d.j. (2007). Uci machine learning repository <http://www.ics.uci.edu/~mllearn/mlrepository.html> Irvine, ca: university of california, school of information and computer science.

Authors

Nadia MESGHOUNI, *Has a Master's degree in knowledge and extraction of data mining 2007FSEGJ Tunisie and INSA lyon 2, registered in thesis in University Paris13 laboratory LIPN and University of Tunis Laboratory LI3ex-SOIE, axes of research the unsupervised learning, collaborative learning and self organizing Map(Kohonen).*

Moncef TEMANNI, *is a doctor INSA lyon2, occupies the function of director ISI Tunis, these axes of research are the systems multi-agents, the problems of satisfaction of constraints and the security of the information .*

Khaled GHEDIRA *is Engineer ENSEEIHT, Engineer of specialization ENSIMAG, Doctor ENSAE Toulouse, habilitated ENSI of Tunis, formerly DR to the Institute of Computing and Artificial intelligence of*



Neuchâtel in Switzerland(Swiss) and consultant at British Telecom, present a director(manager) of the National School of the Sciences of the computing of Tunis(ENSI)and a president of the ATTIA.