ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Clustering of Self-Organizing Map

Hanane Azzag[1] and Mustapha Lebbah[1,2]

1- LIPN-UMR 7030
Université Paris 13 - CNRS
99, av. J-B Clément - F-93430 Villetaneuse
{hanane.azzag, mustapha.lebbah}@lipn.univ-paris13.fr

2- Université Paris 13, UFR (SMBH) - LIM&BIO Lab
74, rue Marcel Cachin 93017 Bobigny Cedex France

Abstract. In this paper, we present a new similarity measure for a clustering self-organizing map which will be reached using a new approach of hierarchical clustering. (1) The similarity measure is composed from two terms: weighted Ward distance and Euclidean distance weighted by neighbourhood function. (2) An algorithm inspired from artificial ants named AntTree will be used to cluster a self-organizing map. This algorithm has the advantage to provide a hierarchy of referents with a low complexity (near the $n\log(n)$). The SOM clustering including the new measure is validated on several public data bases.

1 Introduction

The data clustering is identified as one of the major problems in data mining. Popularity and different variations linked to the clustering problem [1], have given birth to a several methods of resolution. These methods can both use heuristic or mathematics principles. In this paper we are interested by the clustering methods which use topological maps. These methods have the advantage to propose both visualization tools and unsupervised clustering of different types of data (continuous and binary). The basic model proposed by Kohonen, is solely used for continuous data. Extensions and reformulations of Kohonen model have been proposed in the literature [2, 3]. In learning topological maps process, quality criterion are very difficult to define ; They revolve around the interpretation of different mergers or obtained cluster. In this paper we study the the automatic clustering algorithm of the topological map. We find in literature several methods to cluster self-organizinge map, all of these use hierarchical clustering or K-mean combined with a quality index to find the good partition [4]. Thus, we introduce a new similarity measure dedicated to topological map combined with hierarchical clustering algorithm which will be applied to the referent vectors. In this paper we introduce a new hierarchical method named AntTree defined by [5] which is inspired from real ants and their ability to connect themselves to build complex structures.

2 Topological Clustering

Self-organizing maps are increasingly used as tools for visualization, as they allow projection over small areas that are generally two dimensional. The basic model proposed by Kohonen consists on a discrete set $\mathcal{C}$ of cells called map. This map

has a discrete topology defined by undirected graph, usually it is a regular grid in 2 dimensions. We denote p the number of cells. For each pair of cells (c,r) on the map, the distance $\delta(c,r)$ is defined as the length of the shortest chain linking cells r and c on the grid. For each cell c this distance defines a neighbor cell; in order to control the neighborhood area, we introduce a kernel positive function $\mathcal{K}$ ($\mathcal{K} \geq 0$ and $\lim_{|x|\to\infty} \mathcal{K}(x) = 0$). We define the mutual influence of two cells c and r by $\mathcal{K}(\delta(c,r))$. In practice, as for traditional topological map we use smooth function to control the size of the neighborhood as $\mathcal{K}(\delta(c,r)) = \exp(\frac{-\delta(c,r)}{T})$. Using this kernel function, T becomes a parameter of the model. As in the Kohonen algorithm, we decrease T from an initilal value T_{max} to a final value T_{min}. Let $\mathcal{R}^d$ be the euclidean data space and $\mathcal{A} = \{\mathbf{z}_i; i = 1, \ldots, N\}$ a set of observations, where each observation $\mathbf{z}_i = (z_i^1, z_i^2, ..., z_i^d)$ is a continuous vector in $\mathcal{R}^d$. For each cell c of the grid, we associate a referent vector $\mathbf{w}_c = (w_c^1, w_c^2, ..., w_c^j, ..., w_c^d)$ of dimension d. We denote by $\mathcal{W}$ the set of the referent vectors. The set of parameter $\mathcal{W}$, has to be estimated from $\mathcal{A}$ iteratively by minimizing a cost function defined as follows :

$$\mathcal{J}(\phi, \mathcal{W}) = \sum_{\mathbf{z}_i \in \mathcal{A}} \sum_{r \in C} \mathcal{K}(\delta(\phi(\mathbf{z}_i), r)) ||\mathbf{z}_i - \mathbf{w}_r||^2 \quad (1)$$

where ϕ assign each observation $\mathbf{z}$ to a single cell in the map $\mathcal{C}$. In this expression $||\mathbf{z} - \mathbf{w}_r||^2$ is square of the Euclidean distance. At the end of learning, SOM provide a partition of p subsets. This partition and subsets will be denoted by $\mathcal{P} = \{P_1, \ldots, P_c, \ldots, P_p\}$. Each subset P_c is associated to a referent vector $\mathbf{w}_c \in \mathcal{R}^d$. Often the use of SOM is followed by a clustering step of referents. This step is realized using traditional K-means or hierarchical clustering [6]. The choice of two referents which are the best to be merged is done by using similarity measure between two subsets. Different rule or measure are proposed [6, 7, 8]. Very often these measures don't take into account the topological organization provided by SOM. The most widely considered rule for merging subsets may be the method proposed by Ward [6] which is defined as follows:

$$Dist_W = \left(\frac{n_c n_r}{n_c + n_{c_r}} \right) ||\mathbf{w}_c - \mathbf{w}_r||^2 \quad (2)$$

where n_c and n_r represent respectively the size of subsets P_c and P_r. Let's $\mathcal{P}^{t-1}$ denotes the partition before merging two subsets P_c and P_r associated to the referent c et r; and $\mathcal{P}^t$ the partition obtained after merging subsets P_c and P_r. It's easy to prove that loss of inertia after merging two subsets is equal to the criterion Ward (expression 2). Thus the agglomerative algorithm computes a dissimilarity matrix associated to the new partition and choose the subsets which limit the increase of intra-class. The agglomerative Hierarchical clustering can be described by 5 main steps:

1. Initialize the dissimilarity matrix using a subset provided by SOM
2. Find the two closest subsets P_c and P_r according to Ward criterion.
3. Agglomerate them to form a new cluster.

4. Update the dissimilarity matrix using a new obtained partition.
5. Repeat from step 2

Often the clustering map is combined with extern quality index which allow to choose size of the partition. In order to optimize the process of clustering of topological map we propose two modifications. The first one consists to use a new hierarchical clustering which ignores the step 4 and provide automatically the "best" size of partition. Thus, the new hiearchical clustering deals with only one dissimilarty matrix computed at $t = 0$ ($\mathcal{P}^0$). This algorithm is presented in the section below 2.1. The second proposition consists on modifications of similarity criterion in order to take into account the topological organization provided by SOM algorithm. It is necessary to weight the Ward measure by the loss of inertia with value that measures the topological modification after merging two subsets. We propose to quantify this topological modification by as follows: $\sum_{u \in C} K(\delta((c,r),u)))$ where $\delta((c,r),u) = \min\{\delta(c,u), \delta(r,u)\}$. This quantity allows to quantify the topological modification, but do not allow to take into account the referent proximity on the map. Thus we propose to subtract value that measures this proximity as follows:

$$\begin{aligned} Dist_{Neigh-W} &= \left(\sum_{u \in C} K(\delta((c,r),u)))\right) \frac{n_c n_r}{n_c + n_{c_r}} ||\mathbf{w}_c - \mathbf{w}_r||^2 \\ &- K(\delta(c,r))(n_c + n_r)||\mathbf{w}_c - \mathbf{w}_r||^2 \end{aligned} \quad (3)$$

This measure is composed with two terms. The first term computes the inertia loose after merging P_c and P_r. The second term brings subsets corresponding to two referent neighbors on the map, in order to maintain topological order between subsets. Small proximity between two neighbor c and r infers small $\delta(c,r)$, thus the neighbor function $K(\delta(c,r))$ becomes high. Hence, the second term reduces the first term depending on the neighbor function. It's obvious that for null neighborhood our measure computes only Ward criterion. The criterion we proposed allows to obtain a regularized Ward criterion, and this regularization is obtained with the topological order used. Finally the new measure defines a dissimilarity matrix which take into account the inertia loss and the topological order. In order to cluster map using this new matrix, we use hierarchical clustering defined below which is based on artificial ants.

2.1 Hierarchical clustering [5]

To cluster the map we have used an artificial approach inspired from the self assembly behavior observed on real ants. These insects may become fixed to one another to build live structures with different functions [9]. In this artificial model it is shown that this behavior can be used to build an hierarchical tree-structured clustering of the data according to the similarities between those data. The general principles that rule these behaviors are the followings: Each ant represent one data, ants start from an initial point (called the support). They begin to connect themselves to this support and then progressively to previously connected ants. When an ant is connected, it becomes a part of the

structure and other ants may move over this ant or connect themselves to it. The structure grows over time according to the local actions performed by the ants. Moving ants are influenced by the local shape of the structure and by a visual attractor (for instance, the point to reach). Ants which are in the middle of the structure cannot easily disconnect themselves. We can note that in the obtained tree, each node represent one data and the mother ant is more representative of the similarities than to the (daughter) ants.

AntTree has the advantage to have a low complexity (near the $n\ log(n)$). A detailed study was presented in [5] , it confirms that compared to other methods in (n^2), the time required by AntTree can be up to 1000 times less than those obtained by AHC (Ascendant Hierarchical Clustering) method on several large databases and this for the same error quality. Those times will be further reduced since AntTree will be applied on the referent provided by the topological map. This greatly reduces the clustering complexity of the map. Thus, the tree structured obtained is as the best representative of referent set $\mathcal{W} = \{\mathbf{w}_1, ..., \mathbf{w}_p\}$ (with the new similarity measure (3)). Each node parent of the tree is more representative of their node son. So the algorithm, AntTree-SOM-Neigh-W can be described as follows:

- **Input**: $\mathcal{W} = \{\mathbf{w}_1, ..., \mathbf{w}_p\}$, set of referents which constitute the topological map at the end of learning process,
  - Compute the similarity measure using the new method defined in (3),
  - Building the tree using AntTree algorithm,
- **Output** : Tree structure of referents

The tree obtained provides a clustering of topological map $\mathcal{P} = \{P_1..., P_s\}$ where the value s represent the number of clusters founded (each sub-tree represent one cluster) provided by AntTree. Thus, in the same process we propose to cluster the map without any quality index.

3 Validation

We have evaluated and compared our algorithms on a set of 19 databases (table 1). The databases ART1 to ART6 are artificial and have been generated with gaussian and uniform distributions. The others have been extracted from the machine learning repository [10]. To evaluate the quality of map clustering, we adopt the approach of comparing the results to a "ground truth". We use the clustering accuracy for measuring the clustering results. The index is classification rate, usually named purity measure which can be expressed as the percentage of elements of the assigned class in a cluster. This is a common approach in the general area of data clustering. This procedure is defined by [6] as "validating clustering by extrinsic classification", and has been followed in many other studies. Thus, to adopt this approach we need labeled data sets, where the external (extrinsic) knowledge is the class information provided by labels. We compare our map clustering process using a new measure (SOM-AntTree-Neigh-W) and map clustering using Ward index (SOM-AntTree-W). We compare also the impact of this measure using ascendent hierarchical clustering (SOM-AHC-Neigh-W) compared to Ward index (SOM-AHC-W). Table

Datasets	Cl_R	d	N	Datasets	Cl_R	d	N
Atom	2	3	800	Tetra	4	3	400
Anneaux	2	3	1000	Two diamonds	2	2	800
Dem-icercle	2	2	600	WingNut	2	2	1016
Engytime	2	2	4096	ART1	4	2	400
Glass	7	9	214	ART2	2	2	1000
Hepta	7	3	212	ART4	2	2	200
Lsun	3	2	400	ART5	9	2	900
Pima	2	8	768	ART6	4	8	400
Target	6	2	770				

Table 1: Databases used in the experimentation

Datasets/ %	SOM-AT-W	SOM-AT-N-W	SOM-AHC-W	SOM-AHC-Neigh-W
Atom (2)	85.87 (5)	99.9 (7)	97 (10)	100 (9)
Anneaux (2)	97.8 (6)	81.5 (5)	100 (11)	100(11)
Demi-cercle (2)	58.833 (2)	72.67 (4)	100 (9)	100 (9)
Engytime (2)	74.14 (5)	88.04 (7)	95.73 (15)	93.90 (5)
Glass (7)	38.32 (5)	59.81 (6)	50.47 (5)	54 (6)
Hepta (7)	43.4 (4)	43.4 (4)	100 (7)	100(7)
Lsun (3)	55 (3)	93 (5)	92.5 (6)	99.25 (6)
Pima (2)	67 (5)	72.4 (5)	65 (2)	65.10 (2)
Target (6)	83.25 (5)	94.42 (6)	98.44 (11)	98.44 (9)
Tetra (4)	62.5 (3)	81.75 (5)	97 (5)	98.5 (4)
Twodiamonds (2)	100 (4)	100 (5)	98.5 (10)	96.88 (5)
WingNut (2)	95.67 (3)	87.11 (5)	95.28 (6)	98.82 (4)
ART1 (4)	50.5 (4)	84.75 (4)	81 (7)	81(6)
ART2 (2)	94.9 (4)	97.7 (4)	98 (2)	98(2)
ART4 (2)	100 (3)	100 (5)	100 (5)	100 (3)
ART5 (9)	31.78 (4)	50.33 (6)	75 (9)	75.78 (9)
ART6 (4)	24.25 (2)	78.75 (4)	98 (4)	99.75 (4)

Table 2: Comparison between Hierarchical Clustering of map using Ward criterion and a new measure. The number following the good classification rate (purity rate) indicates the number of subsets provided by map clustering. SOM : Self-Organizing Map; AntTree : Hierarchical clustering based on artificial ant; AHC: Ascendant hierarchical Clustering; SOM-AT-W:SOM-AntTree-W; SOM-AT-N-W:SOM-AntTree-Neigh-W

2 lists the classification accuracy obtained with different methods; the purity rate results were provided. Using a new measure $Dist_{Neigh-W}$, we observe that the results are generally better than the clustering map using traditional ward measure. Thus, we validate the assumption that using the neighborhood result provided by topological map in SOM learning phase improve the traditional map clustering without any neighborhood information. Looking to column associated to AnTree, we observe as example, for *Pima* data set, we improve the purity from 67% to 72.4% with the same number of subsets. For *Hepta* data set we obtain identical results 43.4%. In the case of *Anneaux* data set we observe a decline for the purity, moving from 97.8% to 81.5% with less subsets. Finally, we can see, a clear improvement in the purity when new measure proposed (SOM-AntTree-Neigh-W) is used. Looking to column associated to AHC method, we observe the same improvement of purity between Ward and a new measure. Usually, using a new measure we obtain best result. Hence taking into account the topological order improves significatively the results of the map clustering in both cases. We remind here that clustering map algorithms using the Euclidean

distance as the AHC require to use quality index to define the optimal partition of referents, [4, 8]. For the model based on AnTree (with the new measure 3) no index is necessary to obtain the optimal partition. Usually, we observe that AHC provides more clusters than AnTree.

4 Conclusions and perspectives

In this work we have first developed a new similarity measure dedicated to cluster Self-Organizing Maps. This measure takes into account the existing neighborhood between referents of this map. Secondly, we have introduced an original hierarchical clustering algorithm based on artificial ants; this approach is competitive with standard method such as ascending hierarchical clustering by using much lower complexity and processing times. There are many perspectives to study after this series of results. The first consists on approving the similarity measure which seems to be an important tool in the clustering map process. For clustering method, we would like secondly to introduce some heuristics in order to evaluate their influence on the results, like removing clusters with a small number of instances.

Acknowledgmets

This research was supported by SILLAGES project which is funded by ANR agency.

References

[1] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999.

[2] T. Kohonen. *Self-organizing Maps*. Springer Berlin, 2001.

[3] Mustapha Lebbah, Nicoleta Rogovschi, and Younés Bennani. Besom : Bernoulli on self organizing map. In *International Joint Conferences on Neural Networks. IJCNN 2007, Orlando, Florida, August 12-17*, 2007.

[4] J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on*, 11(3):586–600, May 2000.

[5] Hanane Azzag. *Classification hiérarchique par des fourmis artificielles : applications à la fouille de données et de textes pour le Web*. Thèse de doctorat, Laboratoire d'Informatique, Université de Tours, Decembre 2005.

[6] Anil K. Jain and Richard C. Dubes. *Algorithms for clustering data*. Prentice Hall advanced reference series:Computer Science, 1988.

[7] Méziane Yacoub, Fouad Badran, and Sylvie Thiria. A topological hierarchical clustering: Application to ocean color classification. In *ICANN '01: Proceedings of the International Conference on Artificial Neural Networks*, pages 492–499, London, UK, 2001. Springer-Verlag.

[8] C. Ambroise, G. Séze, F. Badran, and S. Thiria. Hierarchical clustering of self organizing map for cloud classification. *Neurocomputing*, 30:47–52, 1998.

[9] C. Anderson, G. Theraulaz, and J.L. Deneubourg. Self-assemblages in insect societies. *Insectes Sociaux*, 49:99–110, 2002.

[10] C.L. Blake and C.L Merz. Uci repository of machine learning databases. technical report. Technical report, University of California, Department of information and Computer science, Irvine, CA, available at: *ftp://ftp.ics.uci.edu/pub/machine-learning-databases*, 1998.