

IMPACT OF LEARNERS' QUALITY AND DIVERSITY IN COLLABORATIVE CLUSTERING

Parisa Rastin*, Basarab Matei, Guénaél Cabanes, Nistor Grozavu and Younès Bennani

*LIPN, UMR CNRS 7030, Institut Galile Université Paris 13
99, avenue Jean-Baptiste Clément, 93430 Villetaneuse*

**E-mail: rastin@lipn.univ-paris13.fr*

Submitted: 28th January 2018; Accepted: 3rd July 2018

Abstract

Collaborative Clustering is a data mining task the aim of which is to use several clustering algorithms to analyze different aspects of the same data. The aim of collaborative clustering is to reveal the common underlying structure of data spread across multiple data sites by applying clustering techniques. The idea of collaborative clustering is that each collaborator shares some information about the segmentation (structure) of its local data and improve its own clustering with the information provided by the other learners. This paper analyses the impact of the quality and the diversity of the potential learners to the quality of the collaboration for topological collaborative clustering algorithms based on the learning of a Self-Organizing Map (SOM). Experimental analysis on real data-sets showed that the diversity between learners impact the quality of the collaboration. We also showed that some internal indexes of quality are a good estimator of the increase of quality due to the collaboration.

Keywords: collaborative clustering, topological neural networks, unsupervised learning, diversity, quality

1 Introduction

The current growth in real-time communication networks leads to new categories of problems. One of these new problems is the distribution of information between multiple locations and owners. For example, there are data distributed across different sites (banks, supermarkets, medical organizations, administrations) describing the same people with different information (i.e. with different features) or describing different people with the same features. The analysis of distributed data-sets requires appropriate methods, particularly where the different sites cannot share data directly for privacy reasons. To deal with this kind of data approaches based on the idea of several algorithms working together have been widely studied in the case of su-

pervised learning [1, 2, 3, 4] where they gave birth to the field of *Ensemble Learning*. However, in the field of unsupervised learning (clustering), ensemble approaches still need further investigations. The existing methods can be categorized into two families: cooperative and collaborative clustering.

In cooperative clustering [5, 6], each clustering algorithm produces its result independently. The final clustering is computed in a post-processing step, and the only exchange of information is about when the individual processes are completed, so that post-processing can start. In this case, a set of clustering algorithms is used in parallel on a given data set. The diversity of the local results is obtained through the diversity of the local algorithms, their parameter values, and their initialization. Once all

local computations are completed, a master algorithm takes control and combines the local results to get a hopefully better overall clustering. The resolution of possible conflicts between the local solutions requires an algorithm that is able to compare results that may differ in their format (e.g. different numbers of clusters or different degrees of belief associated with the results) and to find a consensus solution that minimizes the overall violation of the local results. The cooperative framework is closely related to the ensemble methods developed for supervised learning. Indeed, in these approaches, a set of (diverse) classifiers is learned and the classification of new data points is obtained by taking a (potentially weighted) vote of their predictions. Bayesian averaging can be considered as a precursor method. Numerous new ones have been developed, from the error-correcting output coding to Bagging, and Boosting and their application in various domains have become routine with often good results.

By contrast to the cooperative clustering model, the collaborative model does not seek an overall hopefully better clustering of a given data set through the combination of individual solutions. In the collaborative framework, the goal is that each local computation, quite possibly applied to distinct data sets benefits from the work done by the other learners [7, 8]. The learners solve together problems defined and imposed by the central controller, affecting an individual task to each learner. Interactions are recurrent between learners, responsibility is collective, the action of each collaborator is geared to the performance of the group and vice versa. This can be done through the exchange of information about the local data, or the current hypothesized local clustering, or the value of one algorithm's parameters. The validity of the approach rests on the assumption that useful information can be shared among the local tasks.

Collaborative clustering proposes to use the clustering structure of remote data-sets (the learners) to improve the clustering of local data. In that case, if the data of all sites cannot be used in a single analysis, a local clustering is nevertheless possible in each site without breaching, for example, confidentiality rules. The fundamental concept is that each algorithm operates locally on each dataset, then collaborate by exchanging information about

the local data structure [9, 10, 11, 12]. As the actual data are not shared, confidentiality is preserved. The idea is that every learner shares only information on the clustering (i.e. the structure) of its local data and improves its own clustering with information provided by other learners.

Collaborative clustering approaches follow a two-step procedure [7]:

- 1 *Local step*: Each algorithm will individually process the data it has access to and produces a local clustering partition.
- 2 *Collaborative step*: The algorithms share their results and try to confirm or improve their models with the goal of achieving better clustering results.

At the end of the two phases, all the local clustering algorithms have been enriched with remote information. These two steps are sometimes followed by an aggregation step which aims at reaching a consensus with the final results after collaboration. In this work we will not address the aggregation step because it is a field of its own, and that depending on the application it may not always be advisable to aggregate, for instance when the different views, sites or scales have conflicting partitions [13]. We will instead focus on the collaborative step where the algorithms exchange information with a goal of mutual improvement.

We generally distinguish three different types of collaborative learning: horizontal, vertical and hybrid [14, 15, 16]. The vertical collaboration is to collaborate the clustering results obtained from different data sets described by the same variables with different objects. In horizontal clustering, we deal with the same patterns and different feature spaces. The hybrid collaboration is a combination of both horizontal and vertical collaboration.

In [17], the authors showed that the potential learners are not equivalent to the collaboration. Indeed, if the collaborator is not adequately chosen, the resulting partition can be of lesser quality than the local clustering (without collaboration). This variability is still not clearly understood. Within this context, this paper addresses the problem of the choice of the learners, to ensure that collaborative methods between unsupervised algorithms lead to

positive results. To do so, we will focus on a specific family of unsupervised collaborative frameworks based on topological maps: due to their popularity and interesting vector quantization properties, algorithms such as Kohonen's Self-Organizing Maps (SOM) [18, 19] and Bishop's Generative Topographic Maps [20] have been used in unsupervised collaborative learning approaches to transfer information between algorithms [21, 22, 23].

As some learners can decrease the quality of the collaboration, we investigated several indexes to characterize the learners, in order to be able to predict the potential utility of a collaborator for the collaboration. We tested the impact of the Diversity and Internal Quality of the learners to the gain of quality after collaboration. The Diversity indexes measure the similarity between the two partitions before collaboration, whereas the Internal Quality Indexes measure the compactness and the separability of the partition proposed by the collaborator. Diversity and Quality indexes are known to be important for Ensemble Clustering [24, 25], which is a problem related to Collaborative Clustering.

The rest of the paper is organized as follows: The Horizontal and the Vertical Topological Collaborative Clustering algorithms used in this paper are described and discussed in Section 2. Section 3 exposes the impact of the collaborator's choice on the quality of the collaboration. In Section 4, we define the notion of Diversity and we present the different indexes used in the experiments, then the experimental protocol and the obtained results are described. Section 5 focus on the Internal Quality of the learners, we present several indexes and we describe the experiments performed to test their impact on the final quality of the collaboration, before analyzing the obtained results. Finally, the paper ends with a conclusion and some future works.

2 Topological Collaborative Clustering

Depending on the structure of the data-sets, there are two main types of collaboration learning principle: horizontal and vertical collaboration. The vertical collaboration is to collaborate the clustering results obtained from different data-sets described by the same features but having different

objects. In the case of horizontal clustering, all data-sets are described by the same objects but in different feature spaces: the same number of objects but a different number of features. When the data-sets share some objects and some features, it is called a hybrid collaboration.

In this Section, we present two Topological Collaborative Clustering approaches adapted either for a horizontal or a vertical collaboration.

2.1 Horizontal collaboration

In this Section, we are specifically interested in horizontal collaborations. Horizontal collaboration is usually more difficult than vertical collaboration, since in such cases, the data are described in different spaces: each data-set is described by different variables but has the same objects (samples) as other data-sets. In this case, the problem is *how to exchange information between clusters derived out of data-sets having different features?*

In the Horizontal Topological Collaborative Clustering, each data-set is clustered using a Self-Organizing Map (SOM). To simplify the formalism, without loss of generality, we assume that the maps built from various data-sets have been initialized with the same dimensions (number of neurons) and the same structure (topology). In the following we note $[ii]$ the indexes of the different SOM, and i the indexes of neurons in any SOM. There is no special link between the two sets of indexes, e.g. the i -th neuron of the $[ii]$ -th SOM is treated the same way as the j -th neuron of the same SOM. The main idea of the horizontal collaboration principle between different SOM algorithms is that, if an observation from the $[ii]$ -th data-set is projected on the j -th neuron in the $[ii]$ -th map, then that same observation in the $[jj]$ -th data-set will be projected on the same j -th neuron of the $[jj]$ -th map or one of its neighboring neurons. In other words, *corresponding neurons in different maps should capture the same observations*. Therefore, an additional term reflecting the principle of collaboration is added to the classical SOM objective function. A new collaboration step is also added to estimate the importance of the collaboration, during the collaborative learning process. To compute the relevance of the collaboration, two parameters are introduced: the first one is to adapt the distant clustering information and the second is for weighting the collabora-

tive clustering link (the map which receive information about the distant map).

An analysis of the objective function of the SOM algorithm shows there exists some similarities to the K -Means algorithm. For that reason, it is very natural to adopt the same collaborative scheme for SOM maps. That is, introducing an appropriate modification of SOM cost function via the addition of a collaborative term to convey the topological information. This approach was adopted by Grozavu and Benanni [22, 26] where they highlighted the capacity of improving the overall clustering performance of the map. Formally, the following new objective function is composed of two terms

$$R_H^{[ii]}(W) = R_Q^{[ii]}(W) + R_{Col-H}^{[ii]}(W). \quad (1)$$

The first term in the right-hand side of equation (1) corresponds to the classical SOM objective function and writes

$$R_Q^{[ii]}(W) = \sum_{jj=1}^P \sum_{i=1}^N \sum_{j=1}^{|w|} K_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2. \quad (2)$$

The second term in the right-hand side of the equation (1) corresponds to the collaborative procedure

$$R_{Col-H}^{[ii]}(W) = \sum_{jj=1}^P \sum_{i=1}^N \sum_{j=1}^{|w|} L_{ij} \|x_i^{[ii]} - w_j^{[ii]}\|^2. \quad (3)$$

In equations (2) and (3) the parameter P represents the number of datasets (or the classifications), N - represents the number of observations and $|w|$ is the number of prototype vectors from the $[ii] - th$ SOM map (the number of neurons). $\chi(x_i)$ is the assignment function which allows finding the Best Matching Unit (BMU), it selects the neuron with the closest prototype from the data x_i using the Euclidean distance and it's defined as follows

$$\chi(x_i) = \operatorname{argmin} (\|x_i - w_j\|^2). \quad (4)$$

The factor $\sigma(i, j)$ represents the distance between two neurons i and j from the map, and it is defined as the length of the shortest path linking cells i and j on the SOM map.

The factor $K_{\sigma(i, j)}^{[cc]}$ is the neighborhood function on the $SOM^{[cc]}$ map between two cells i and j . This

function depends on the distance between two neurons and is defined as follows

$$K_{\sigma(i, j)}^{[cc]} = \exp \left(-\frac{\sigma^2(i, j)}{T^2} \right), \quad (5)$$

where T is the temperature which allows controlling the neighborhood influence of a cell on the map, it decreases with the T parameter.

The nature of the neighborhood function $K_{\sigma(i, j)}^{[cc]}$ is identical for all the maps, but its value varies from one map to another: it depends on the closest prototype to the observation that is not necessarily the same for all the SOM maps.

In addition, L_{ij} is defined as

$$L_{ij} = \left(K_{\sigma(j, \chi(x_i))}^{[ii]} - K_{\sigma(j, \chi(x_i))}^{[jj]} \right)^2. \quad (6)$$

This factor emulates the best matching unit distance between two different maps. It gives a numerical estimation of the difference of representation of the two maps, by comparing the neighborhood distances of a data point i and the neuron j in both maps. If this distance is similar in the two SOMs for each data point, it means that the BMUs of data points are the same (or at least are close neighbors) in both maps. The factor gives a fine method to avoid the issue of having to be in the same feature space, because it relies solely on the topology of the two maps and the built of a neighborhood distance. However, as it is defined in Equation (5), the neighborhood function K_{σ} relies on a distance metric $\sigma(i, j)$ which depends on the feature space of the data. Therefore, for any comparison to be possible between the neighborhood, we first need to normalize the neighborhood functions across the different feature spaces.

Algorithm 1 Horizontal Collaboration

for $t = 1$ to N_{iter} **do**

1. Local step:

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

 Find the prototypes minimizing the classical SOM objective function

$$w^* = \operatorname{argmin}_w \left[\sum_{i=1}^N \sum_{j=1}^{|w|} K_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2 \right]$$

end for

2. Collaboration step:

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

 Update the prototypes of the $[ii]$ map by using (8) that minimize the objective function $R_H^{[ii]}$ of the horizontal collaboration.

end for

end for

As described in Algorithm 1, to compute the collaborated prototypes matrix, a gradient optimization is used as follows

$$w^{*[ii]} = \operatorname{argmin}_W [R_H^{[ii]}(W)]. \quad (7)$$

The obtained update rule is given by

$$w_{jk}^{*[ii]}(t+1) = w_{jk}^{*[ii]}(t) + \frac{\sum_{i=1}^N K_{\sigma(j, \chi(x_i))}^{[ii]} x_{ik}^{[ii]} + \sum_{\substack{jj=1 \\ jj \neq ii}}^P \sum_{i=1}^N L_{ij} x_{ik}^{[ii]}}{\sum_{i=1}^N K_{\sigma(j, \chi(x_i))}^{[ii]} + \sum_{\substack{jj=1 \\ jj \neq ii}}^P \sum_{i=1}^N L_{ij}}, \quad (8)$$

where L_{ij} is defined by (6). Indeed, during the collaboration with a SOM map, the algorithm takes into account the prototypes of the map and its topology through the neighborhood function.

2.2 Vertical collaboration

In the vertical case, contrary to horizontal collaboration, all the data-sets have the same features (same description space) but have different observations. In this case, the observations of these data-sets are of equal size, and the size of prototype vectors for all the SOM maps will be the same. The basic idea of collaboration, in this case, is the following: a neuron j of ii -th SOM map and the same neuron j of the jj -th map should be very similar using the Euclidean distance. In other words, *corresponding neurons in the different maps should represent groups of similar observations*. That's why we changed the objective function of the classical SOM algorithm to introduce this constraint on the various maps during the collaboration step. Formally, we have achieved the following new objective function

$$R_V^{[ii]}(W) = R_Q^{[ii]}(W) + R_{Col-V}^{[ii]}(W). \quad (9)$$

Here $R_Q^{[ii]}(W)$ is still defined by (2), while $R_{Col-V}^{[ii]}(W)$ writes

$$R_{Col-V}^{[ii]}(W) = \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^{N^{[ii]} |w|} \sum_{j=1}^{|w|} L_{ij} \|w_j^{[ii]} - w_j^{[jj]}\|^2 \quad (10)$$

with L_{ij} defined by (6), where again P represents the number of datasets, N - the number of observations

of the ii -th dataset, $|w|$ is the number of prototype vectors from the ii SOM map and which is equal for all the maps. Note the difference between (10) and (3) in the factors $\|w_j^{[ii]} - w_j^{[jj]}\|^2$ and $\|x_i^{[ii]} - w_j^{[ii]}\|^2$ respectively. Indeed, contrarily to the horizontal collaboration, in the case of vertical collaboration the data are in the same feature space and the value of the collaboration confidence parameters can be computed using the similarity between two collaboration maps (the prototypes similarity) as in Algorithm 2. As for the horizontal collaboration, in the

Algorithm 2 Vertical Collaboration

1. Local step:

for $t = 1$ to N_{iter} **do**

For each $DB[ii]$, $ii = 1$ to P :

Find the prototypes minimizing the classical SOM objective function

$$w^* = \operatorname{argmin}_w \left[\sum_{i=1}^N \sum_{j=1}^{|w|} K_{\sigma(j, \chi(x_i))}^{[ii]} \|x_i^{[ii]} - w_j^{[ii]}\|^2 \right]$$

2. Collaboration step:

For the vertical collaboration of the $[ii]$ -th map with the map $[jj]$:

Collaboration phase:

Update the prototypes of the $[ii]$ map minimizing the objective function of the vertical collaboration $R_V^{[ii]}(W)$ by using the expression (12).

end for

case of vertical collaborative learning, the gradient optimization procedure is used to compute the prototypes matrix

$$w^{*[ii]} = \operatorname{argmin}_W [R_V^{[ii]}(W)] \quad (11)$$

The obtained update rule is given by

$$w_{jk}^{*[ii]}(t+1) = w_{jk}^{*[ii]}(t) + \frac{\sum_{i=1}^N K_{\sigma(j, \chi(x_i))}^{[ii]} x_{ik}^{[ii]} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^{N^{[ii]} |w|} L_{ij} w_{ik}^{[jj]}}{\sum_{i=1}^N K_{\sigma(j, \chi(x_i))}^{[ii]} + \sum_{jj=1, jj \neq ii}^P \sum_{i=1}^N L_{ij}} \quad (12)$$

where L_{ij} is defined by (6). The learning algorithm in this case arises under the scheme given by Algorithm 2.

3 Impact of the learners on the collaboration's quality

In [10], the authors show that not all of the potential learners are suitable to collaborate in the horizontal collaboration (Table 1). Here we present similar results in the vertical case (Table 2).

Table 1. Experimental results of the horizontal collaborative approach on different data sets

Dataset	Map	Purity
Waveform	SOM_1	86.54
	SOM_2	39.5
	$SOM_{1 \rightarrow 2}$	73.24
	$SOM_{2 \rightarrow 1}$	58.34
Isolet	SOM_1	80.79
	SOM_2	93.27
	$SOM_{1 \rightarrow 2}$	81.46
	$SOM_{2 \rightarrow 1}$	92.87
Wdbc	SOM_1	94.02
	SOM_2	96.49
	$SOM_{1 \rightarrow 2}$	95.23
	$SOM_{2 \rightarrow 1}$	96.57
SpamBase	SOM_1	80.57
	SOM_2	84.95
	$SOM_{1 \rightarrow 2}$	82.84
	$SOM_{2 \rightarrow 1}$	83.79

These tables present the purity [27] of the clustering before and after collaboration. The purity is the average proportion of the majority label in each cluster; "true" labels of the data must be known in order to compute the purity. A local clustering is noted M_x and a collaboration between x and y is noted $M_{x \rightarrow y}$ if x uses information from y and $M_{y \rightarrow x}$ if y uses information from x . It is quite clear that $M_{x \rightarrow y}$ is beneficial for x when y have a higher purity than x . However, if the collaborator has a lower purity, the purity of the collaboration usually decreases. This shows that it is important to choose carefully the best learners among the potential candidates. The purity is computed based on already known labels of the data (it is an External Quality Index) and cannot usually be computed as the labels are rarely known in the real case clustering problem. We, therefore, need to find other criteria to estimate the quality of the collaborator for the collaboration.

The objective of this work is to test several criteria in order to help the choice of the most rele-

vant learners. The purpose is to find the best collaborator to collaborate with and improve the local quality. Intuitively, a relevant collaborator should produce a good model of its own data, i.e. the proposed clustering can be trusted. In addition, the collaborator's model should not be too different from the local model to be able to obtain a reasonable agreement between the structure of the local dataset and the model proposed by the learners. To test these ideas, we tested two types of indexes: The Diversity between the local clustering model and the model proposed by the learners (Section 4) and the individual internal Quality of each learner (Section 5).

Table 2. Experimental results of the vertical collaborative approach on different data sets

Dataset	Map	Purity
Waveform	SOM_1	69.71
	SOM_2	69.87
	$SOM_{1 \rightarrow 2}$	74.57
	$SOM_{2 \rightarrow 1}$	70.71
Isolet	SOM_1	98.85
	SOM_2	98.46
	$SOM_{1 \rightarrow 2}$	79.54
	$SOM_{2 \rightarrow 1}$	98.30
Wdbc	SOM_1	96.71
	SOM_2	97.87
	$SOM_{1 \rightarrow 2}$	96.99
	$SOM_{2 \rightarrow 1}$	97.49
SpamBase	SOM_1	76.26
	SOM_2	70.43
	$SOM_{1 \rightarrow 2}$	72.28
	$SOM_{2 \rightarrow 1}$	69.78

4 Impact of the Diversity between Learners

The Diversity is the difference between two cluster partitioning (local and collaborator). In ensemble methods, there is a relation between ensemble efficiency and ensemble diversity, and a diversity measure can be useful to choose the combination method [24, 25].

In this paper, we address the question of the use of diversity information for an unsupervised learning task aiming at finding the best collaboration between several clustering during the learn-

ing. In order to test if the diversity between learners is a relevant index to choose the best collaborator, we first tried several diversity indexes to select the most discriminant index between informative and non-informative learners. We also examined the different way of using these diversity index to find and improve the results. In general, a low diversity means that the different data-sets are partitioned in the same way by the clustering algorithms. A high diversity means that the two data-sets are partitioned in a very different way, either because of differences in the two clustering methods used or because of intrinsic difference in the data representation in the two spaces. In this study, any high diversity was due to a difference in the data structure, because we used the same algorithm to partition both data- sets.

4.1 Relevance of the Diversity indexes

We first investigated the relevance of several diversity indexes for the collaboration task. We therefore compared several well-known indexes and their capability at discriminate similar clustering to random clustering.

When the data-sets represent different objects in the same feature space, Diversity can be defined as the average of the minimal Euclidean distance between clusters' prototypes of the two partitions. When the data-sets represent the same objects in different space, a distance between prototypes cannot be defined and Diversity indexes are usually based on the agreement between the two partitions, i.e. each pair of an object should be either in the same cluster in both partitions or in different clusters in both partitions. We focus on this type of indexes in this paper.

A simple diversity measure consists in calculating the complement of a similarity measure S between two partitions P_1 and P_2 : $D(P_1, P_2) = 1 - S(P_1, P_2)$. We tested six Diversity indexes based on several similarity measures that we recall below. To this end, we define a_{11} as the number of object pairs belonging to the same cluster in P_1 and P_2 , a_{10} denotes the number of pairs that belong to the same cluster in P_1 but not in P_2 , and a_{01} denotes the pairs in the same cluster in P_2 but not in P_1 . Finally, a_{00} denotes the number of object pairs in different clusters in P_1 and P_2 . N the total number of objects, n_i the number of objects in cluster i in P_1 , n_j the num-

ber of objects in cluster j in P_2 and n_{ij} the number of object in cluster i in P_1 and j in P_2 . In Adjusted Rand, n_c is the agreement we would expect to arise by chance alone using Rand index. We will consider the following similarity measures: the Rand measure proposed in [28] defined by

$$R = \frac{a_{00} + a_{11}}{a_{00} + a_{01} + a_{10} + a_{11}}, \quad (13)$$

the Adjusted Rand measure proposed in [29] defined by

$$AR = \frac{a_{00} + a_{11} - n_c}{a_{00} + a_{01} + a_{10} + a_{11} - n_c}, \quad (14)$$

the Jaccard measure proposed in [30] defined by

$$J = \frac{a_{11}}{a_{01} + a_{10} + a_{11}}, \quad (15)$$

the Wallace measure proposed in [31] defined by

$$W_{P_1 \rightarrow P_2} = \frac{a_{11}}{a_{11} + a_{10}}, \quad (16)$$

the Adjusted Wallace measure proposed in [32] defined by

$$AW = \frac{W_{P_1 \rightarrow P_2} - \frac{\sum_i |P_2| n_i (n_i - 1)}{N(N-1)}}{1 - \frac{\sum_i |P_2| n_i (n_i - 1)}{N(N-1)}}, \quad (17)$$

the Normalized Mutual Information measure proposed in [33] defined by

$$NMI = \frac{-2 \sum_{ij} n_{ij} \log \frac{n_{ij} N}{n_i n_j}}{\sum_i n_i \log \frac{n_i}{N} + \sum_j n_j \log \frac{n_j}{N}}, \quad (18)$$

the Variation of Information measure proposed in [34] defined by

$$VI = -2 \sum_{ij} \frac{n_{ij}}{N} \log \frac{n_{ij} N}{n_i n_j} \sum_i \frac{n_i}{N} \log \frac{n_i}{N} - \sum_j \frac{n_j}{N} \log \frac{n_j}{N}, \quad (19)$$

To analyze the relevance of the different indexes, we tested how discriminant each index is between informative and non-informative learners. In that order, we used noisy features to manipulate the quantity of relevant information shared by the collaborator. In that case, the quality of the collaboration depends directly on the percentage of noise in the collaborator's data-set. We used that property to test different diversity indexes. Our experiments are based on the "waveform" data-set. This data-set consists of 5000 instances divided into 3 classes. The original base included 40 attributes, 19

are all noise attributes with mean 0 and variance 1. Each class is generated from a combination of 2 of 3 "base" waves.

We constructed ten subsets, each with five attributes. Five of these subsets (db1 to db5) were constructed with informative attributes (Relevant data-sets) and five subsets (db6 to db10) with uninformative (noisy) features (Noisy data-sets). We then computed several diversity measures between two Relevant data-sets, a Noisy and a Relevant, and two Noisy data-sets, based on the comparison between the two SOM representations.

As non-noisy data-sets were constructed to have the same class structure with three defined clusters, they are expected to have a high similarity (low diversity). On the contrary, each Noisy data-set is different from other Noisy and to non-Noisy data-sets. A relevant diversity index should, therefore, be high (close to 1) as soon as a Noisy data-set is involved, and low (close to 0) otherwise. Table 3 presents the values of the Similarity indexes (i.e. 1-Diversity) in several cases of collaboration. As we can see, the ARI (Adjusted Rand Index) and the NMI (Normal Mutual Information) are both very low (close to zero) as soon as a Noisy data-set is involved and have a much higher value when two non-noisy data-sets collaborate. Therefore, they both are a good candidate for a Diversity measure. The other indexes show little difference between collaboration with and without noisy and are probably not suitable to be used as a Diversity index. The difference between ARI, NMI and the other indexes is that the former are normalized in order to have a value close to zeros when the clusters are chosen randomly. This allows a full use of the [0,1] range and a clearer distinction between random and similar models. In the following experiments, we chose to use the Adjusted Rand Index as a Diversity index.

4.2 Effect of the Diversity on the quality of the collaboration

To evaluate the effect of the diversity between learners on the quality of the collaboration, we used several data-sets of different size and complexity [35]. In addition to the "waveform" data-set already described, the experimental protocol is applied on the "Wisconsin Diagnostic Breast Cancer (WDBC)", "Isolet", "Glass", "Wine" and "Spam-Base" data-sets:

- *Wisconsin Diagnostic Breast Cancer (WDBC)*: This data has 569 instances with 32 variables (ID, diagnosis, 30 real-valued input variables). Each data observation is labeled as benign (357) or malignant (212). Variables are computed from a digitized image of a Fine Needle Aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.
- *Isolet*: This data-set was generated as follows. 150 subjects spoke the name of each letter of the alphabet twice. Hence, we have 52 training examples for each speaker. The speakers are grouped into 5 sets of 30 speakers each. The data consists of 1559 instances and 617 variables. All variables are continuous, real-valued variables scaled into the range $[-1.0, 1.0]$.
- *Glass*: This real data-set contains 6 classes for a total of 214 objects, where each class corresponding to a type of glass. An object is described by 10 attributes.
- *Wine*: These data are the results of a chemical analysis of 178 wines (samples) grown in the same region in Italy but derived from 3 different categories. The attributes describe the quantities of 13 constituents found in each of the three types of wines.
- *Spam Base*: The SpamBase data set is composed for 4601 observations described by 57 variables. Every variable described an e-mail and its category: spam or not-spam. Most of the attributes indicate whether a particular word or character was frequently occurring in the e-mail. The run-length attributes (55-57) measure the length of sequences of consecutive capital letters.

The criteria we chose to estimate the quality of the collaboration is the gain of purity after collaboration (i.e. the difference between the purity of the local clustering before and after collaboration). The purity of a neuron is the percentage of data belonging to the majority class. Assuming that the data labels set $L = \{l_1, l_2, \dots, l_{|L|}\}$ and the prototypes set $C = \{c_1, c_2, \dots, c_{|C|}\}$ are known, the formula that expresses the purity of a map is

$$purity = \sum_{k=1}^{|C|} \frac{|c_k|}{N} \times \frac{\max_{i=1}^{|L|} |c_{ik}|}{|c_k|}, \quad (20)$$

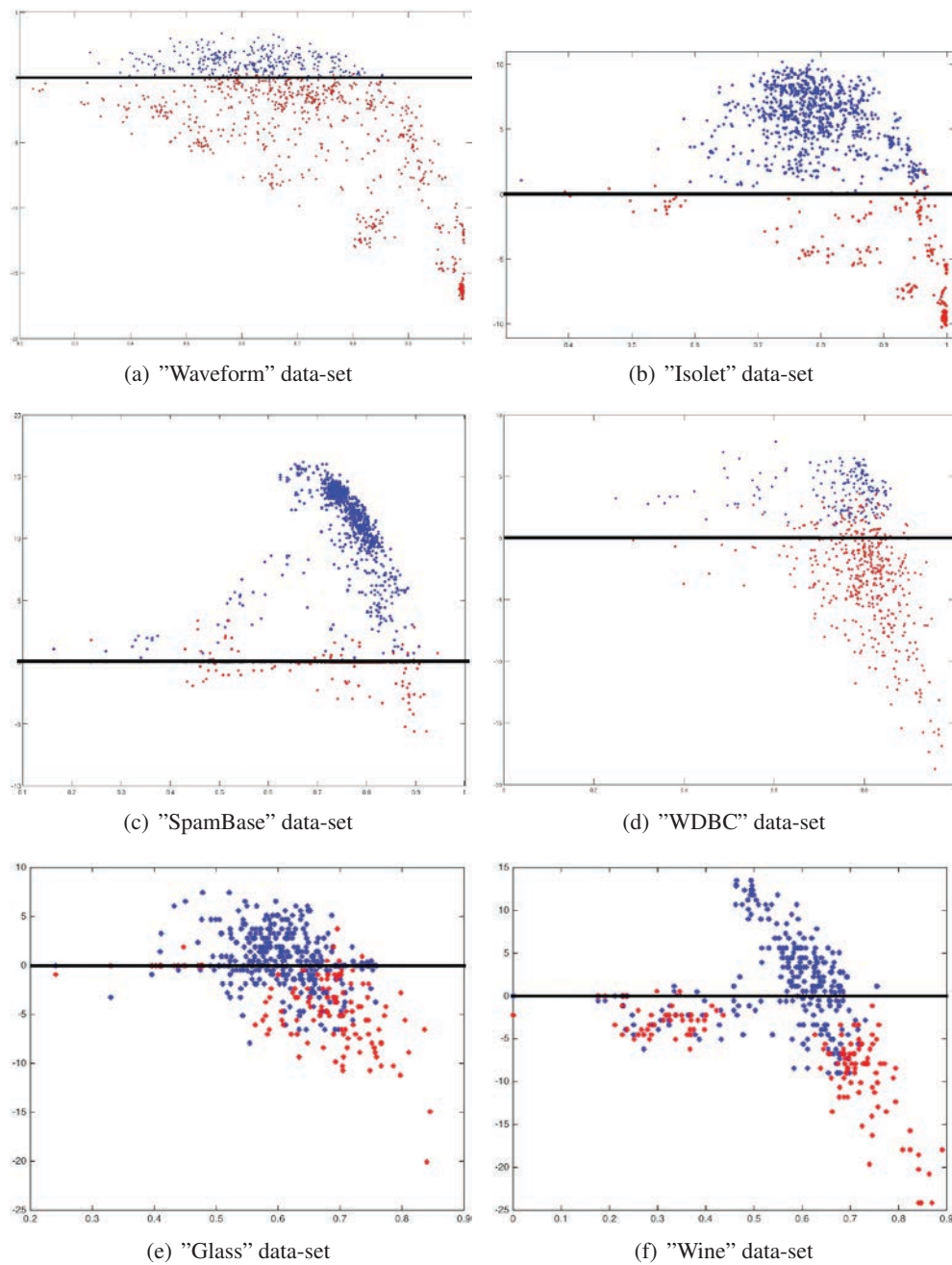


Figure 1. Gain of Purity (ordinate) in function of the Diversity between learners (abscissa) for a horizontal collaboration between a high-quality learner receiving information from a collaborator of random quality.

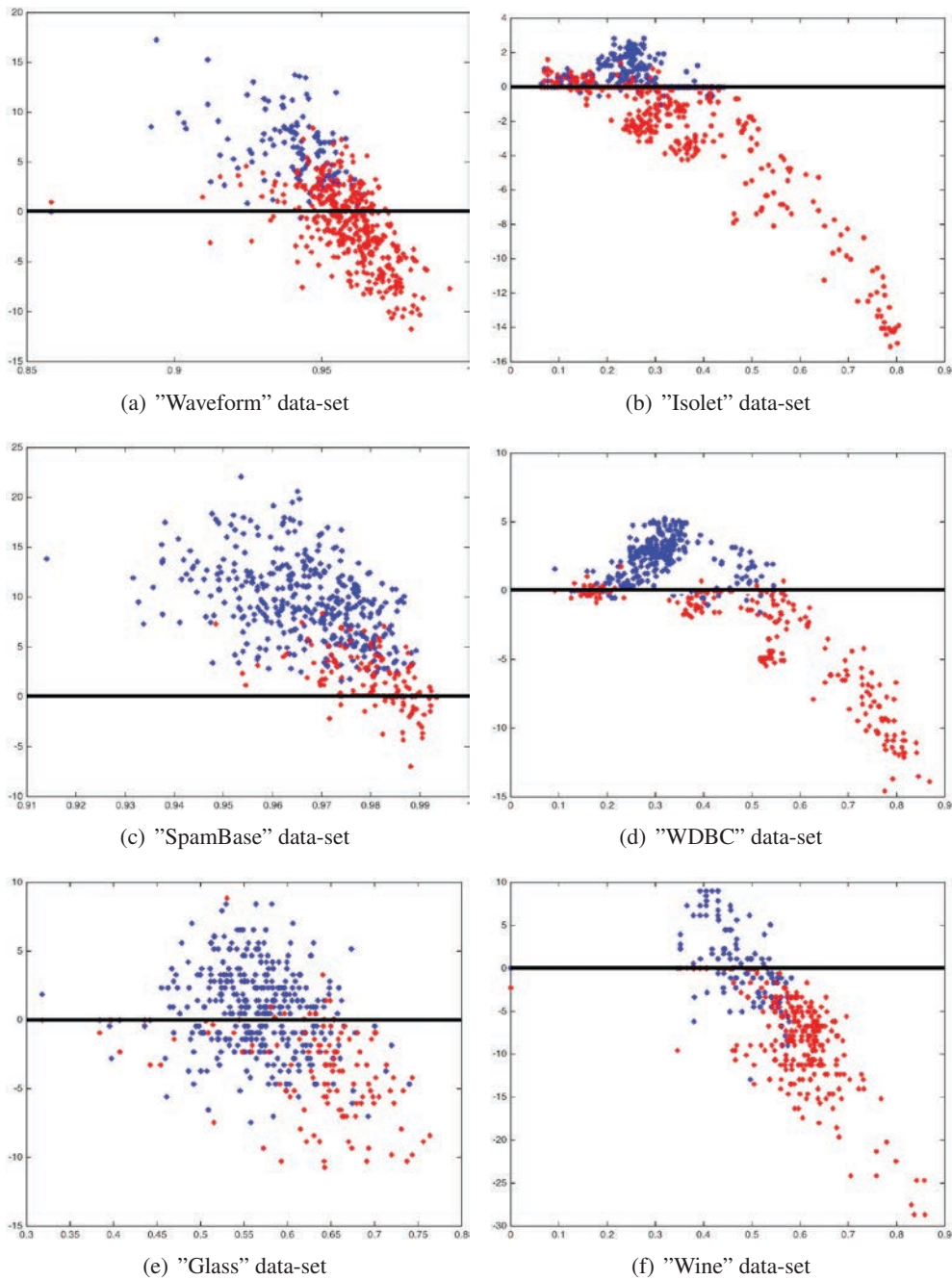


Figure 2. Gain of Purity (ordinate) in function of the Diversity between learners (abscissa) for a vertical collaboration between a high-quality learner receiving information from a collaborator of random quality.

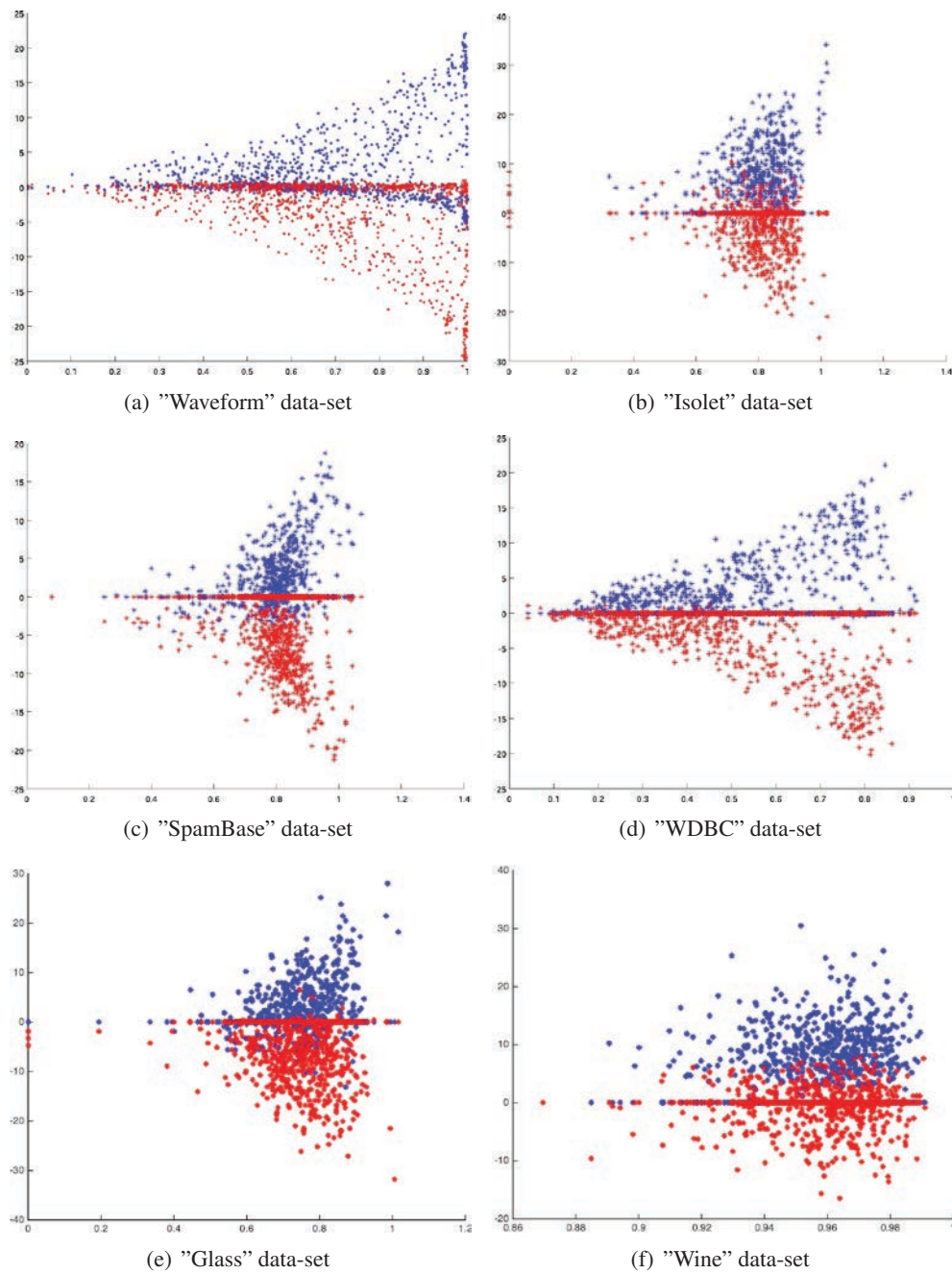


Figure 3. Gain of Purity (ordinate) in function of the Diversity between learners (abscissa) for randomly chosen pairs of learners.

Table 3. Similarity measure on the waveform subsets

Subset Similarity index	Relevant data-sets		Relevant vs Noisy data-sets		Noisy data-sets	
	db2/db3	db3/db4	db2/db8	db4/db9	db7/db8	db9/db10
Rand	0.6707	0.7042	0.5539	0.555	0.5430	0.5553
Adjusted Rand	0.2625	0.3356	0.00008	0.0002	0.0000	0.0000
Jaccard	0.3429	0.3869	0.2017	0.2008	0.2000	0.2003
Wallace's coefficient	0.5079	0.5578	0.3332	0.3342	0.3300	0.3334
Adjusted Wallace	0.5135	0.5581	0.3383	0.3347	0.3500	0.3411
Normal Mutual Information	0.2620	0.3072	0.0002	0.0006	0.0003	0.0004
Variation of Information	2.334	2.1918	3.1577	3.1631	3.1680	3.1664

where $|c_k|$ is the total number of data associated with the neuron c_k , and $|c_{ik}|$ is the number of data of class l_i which are associated with the neuron c_k .

For each data-set we generated 1000 pairs of learners by generating subsets of 4 random features. We applied algorithm ?? and algorithm ?? on each pair (The size of all maps was fixed to 10×10), then we computed the gain of purity of each collaboration, i.e. for each pair. One member of the pair (noted "local" collaborator) was chosen to have a high purity (over 0.8) before collaboration, the other (noted "remote" collaborator) being chosen randomly. The gain in purity (before and after collaboration) was computed only for the "local" collaborator. Figures 1 and 2 show the results for each data-sets, each dot represent a collaboration, either horizontal (Figure 1) or vertical (Figure 2). Blue points represent collaborations where the local collaborator has a lower purity than the remote collaborator, red points represent the opposite case. It is clear from this Figure that the local collaborator should receive information from a remote collaborator with a higher purity to increase its own purity through the collaboration. In most real case in clustering problems we don't have the true labels of the data and it is no possible to compute the Purity index, however, we can see here that the Diversity between Learners can provide some information about the quality of the Collaboration. Indeed, when the diversity is low, the two learners propose very similar results and the collaboration is not informative, leading to a gain close to 0. However, as the local collaborator has a high purity if the remote collaborator proposes a very different solution (high diversity), this solution is probably incorrect and the collaboration will decrease the purity of the local collaborator. An intermediate diversity is therefore

optimal.

If we think about the goal of a collaboration, these results make sense: the goal of a collaboration is to have the SOM algorithms helping each other. Therefore, when several algorithms find solutions that are similar, it is most likely that they have actually found a structure in the data. As such, collaborating with algorithms that have solutions similar to the local partitioning is a convenient way to avoid the risk of negative collaboration. There are actually two good reasons not to collaborate with an algorithm the results of which are too different from the local partition:

- 1 This collaborator may be in a feature space where the clusters to be found are completely different, even for the same objects.
- 2 If this collaborator has a solution that is dissimilar to these of all other algorithms, maybe it is just a bad solution.

This is valid only if the local collaborator has a high quality. In the general case, the relationship between Diversity and gain of purity is different. We tested 1000 collaborations between learners chosen randomly (see Figure 3 for some examples in the horizontal framework). In the general case, it is clear that the Diversity is directly linked to the variability of the quality of the collaboration. A low Diversity between learners will lead to a useless collaboration. On the contrary, a high Diversity can potentially greatly improve the result after collaboration, when the local collaborator proposes an incorrect solution and receive a very good solution from the remote collaborator, or greatly decrease the quality of the clustering after collaboration (in the opposite case), or anything between these two

extremes. Actually, the most important information here seems to be the quality of the collaborator. As the Purity is an external index which cannot be computed most of the time, we investigated the impact of several Internal Quality indexes to predict the gain in purity after collaboration.

5 Impact of the Internal Quality of the Learners

In this Section, we analyze the link between the proportion of relevant information in a data-set and the Internal Quality of a clustering on these data. Then we show how Internal Quality indexes and gain of purity after collaboration are related to each other.

5.1 Relationship between quantity of information and Internal Quality

We used the percentage of noisy features as an indicator of the proportion of relevant information in a data-set. Noisy data-set are less informative than noise-free data-set, because noisy features contain random values. Noise is not the only cause to explain the lack of relevant information in a collaborator: unrelated description of the objects between the two learners can be another one (in that case the two partitions are very different). However, noise is the easiest to manipulate and only depend on the internal property of the data-set.

We investigated the Correlation between the percentage of noise in the data and the internal indexes, in order to find the best index to predict the percentage of relevant information (i.e. here the percentage of noise). In this experiment, we studied six different internal indexes: Calinski-Harabasz, Davies-Bouldin, Krzanowski-Lai, and Silhouette, as well as two SOM-specific indexes: Topological error and Quantization error. We recall the definitions of these internal quality indexes.

In the following k is the number of clusters and n is the number of data x of dimension d . W is the sum of the within-cluster variances and B is the sum of the between-cluster variances. \bar{x}_m is the center of gravity of cluster m . In Silhouette, $a(i)$ is the average distance between x_i and the observations belonging to the same cluster as x_i and $b(i)$ is the lowest average distance between x_i and the observations

in each other clusters.

The Calinski-Harabasz index is introduced in [36]

$$CH(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)}. \quad (21)$$

The Davies-Bouldin index, introduced in [37], is defined as

$$DB = \frac{1}{k} \sum_{l \neq m} \max \frac{\text{avg}_{x_i \in l} \|x_i - \bar{x}_l\| + \text{avg}_i \|x_i - \bar{x}_m\|}{\|\bar{x}_l - \bar{x}_m\|}, \quad (22)$$

where avg is the average function.

The Krzanowski-Lai index [38] writes

$$KL = \left| \frac{W(k-1)(k-1)^{2/d} - W(k)k^{2/d}}{W(k)k^{2/d} - W(k+1)(k+1)^{2/d}} \right|. \quad (23)$$

Finally, the Silhouette index [39] is given by

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}, \quad (24)$$

the two internal indexes are the Topological error index Te [40], which is defined as the proportion of data with the two closest neurons not connected, and the quantization error index Qe [41]: the average distance between each data point and its closest prototype.

The experimental data-set is the waveform data-set. We generated 1000 subsets from the waveform data-set, by choosing randomly 10 features with replacement among the 40 features of waveform. As waveform contains 20 informative and 20 noisy features, the proportion of noisy features varies between 0 and 100% in the subsets. A SOM (size 10x10) was applied to each of the subsets, and the quality indexes were computed to evaluate the SOM quality. Finally, a correlation analysis between the percentage of noise and the quality indexes were performed to find the best index to predict the noise.

Figure 4 shows the correlation between each index and the percentage of noise. There is a clear correlation with at least three indexes: Davies-Bouldin, Calinski-Harabasz and Silhouette. Silhouette index is the best candidate to predict the proportion of relevant information due to the presence of noise in the data.

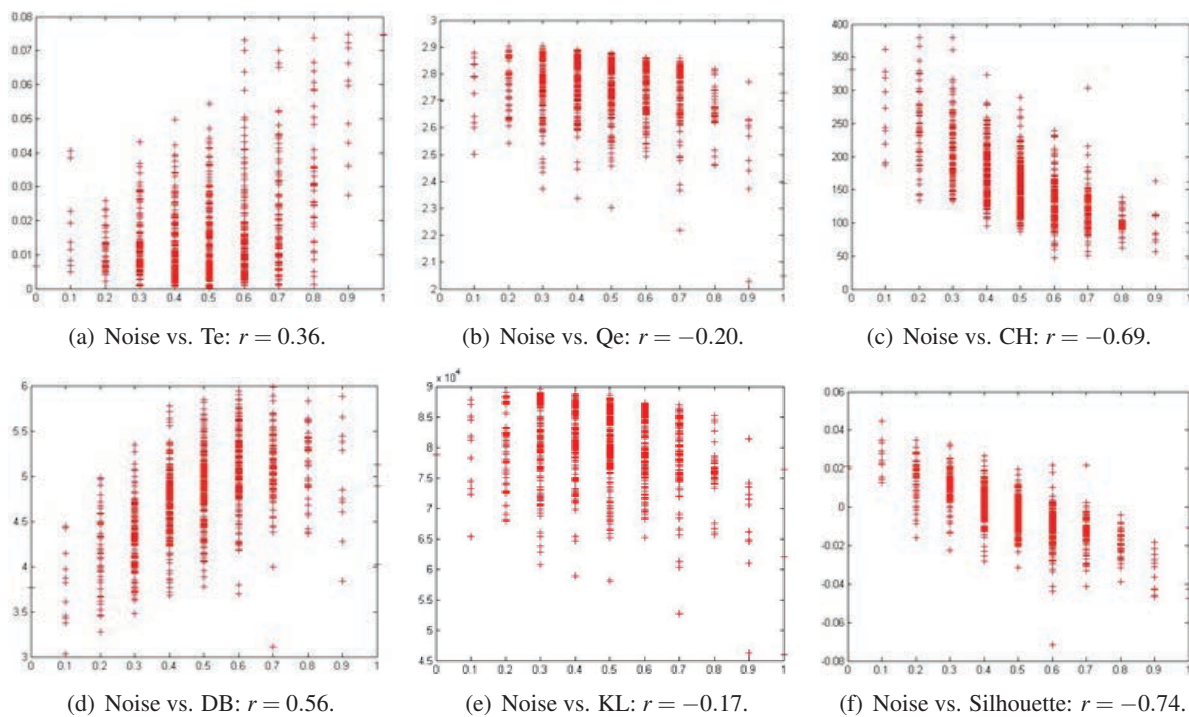


Figure 4. Correlation between the percentage of noisy features and the quality of the clustering for several quality indexes. r is the Pearson’s correlation, all $p < 0.001$ (t test).

5.2 Correlation between collaborator’s Quality and the gain in Purity

Then we conducted in 1000 collaborations between random learners to calculate the Pearson correlation r and the statistical significance p of this correlation (t -test) between the Internal Quality Index of the remote collaborator and the gain in a purity of the local collaborator after collaboration (Table 4).

Table 4. Pearson’s Correlation between the gain in purity and the quality of the collaborator for several quality indexes.

Indices	r	p
Quantification error	-0.3395	0.0087
Topological error	-0.4497	0.0000
Silhouette	0.3915	0.0001
Davis-Bouldin	0.3936	0.0001
Calinski-Harabasz	0.3000	0.1593
Krzanowski-Lai	-0.3193	0.0405

The gain of purity is clearly significantly correlated to most internal Quality index. The best correlations are with Silhouette, Davies-Bouldin and the Topological error. The topological error is the

best predictor of the gain in Purity, but this index is specific to SOM-based methods. It seems that Silhouette index is at the same time a good predictor of the proportion of noise in the data and the potential gain in Purity from a collaboration.

These results again make sense. Collaborative Clustering is about trusting the information received from the learners in order to update its own partition based on local data. If a collaborator obtains a partition with a clear structure (i.e. a good internal quality), the information it shares is worth using. On the contrary, if the structure is unclear, it is probably better to use only information from the local data.

6 Conclusion

In this paper, we studied the impact of the learners’ Diversity and Quality on the output of the collaboration. The results show that the Diversity between learners can be an important information for predicting the gain in purity due to the collaboration. However, the Diversity must be completed with an estimation of the Internal Quality of the learners.

If the Quality of the local clustering is low, any collaborator with higher Quality will improve the quality of the clustering. However, if the Quality of the local clustering is high, the optimal learners should have both a high Quality and an intermediate Diversity (far from 0 and 1).

Indeed, two good clustering solutions must be different enough to add new information to the collaboration. They should not be too different, though, as a collaboration between two very different well-structured partitions will result in an average solution of lower quality.

We showed that the Adjusted Rand index and the Normalized Mutual Information index are good candidates to estimate the diversity between Learners and that Silhouette index is a good estimator of the quantity of noise in the data and the final Quality of the collaboration.

Our future work will be to propose a weighted version of the Topological Collaborative Clustering, allowing an automatic weight to each collaborator based of its importance according to its intrinsic quality and diversity.

References

- [1] R. E. Schapire, The strength of weak learnability, *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, Jul. 1990. [Online]. Available: <http://dx.doi.org/10.1023/A:1022648800760>
- [2] D. H. Wolpert, Stacked generalization, *Neural Networks*, vol. 5, pp. 241–259, 1992
- [3] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, On combining classifiers, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998. [Online]. Available: <http://dx.doi.org/10.1109/34.667881>
- [4] P. Bachman, O. Alsharif, and D. Precup, Learning with pseudo-ensembles, in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3365–3373
- [5] A. Strehl and J. Ghosh, Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions, *Journal on Machine Learning Research (JMLR)*, vol. 3, pp. 583–617, Dec. 2002
- [6] J. da Silva and M. Klusch, Inference on distributed data clustering, in *Machine Learning and Data Mining in Pattern Recognition*, ser. Lecture Notes in Computer Science, P. Perner and A. Imiya, Eds. Springer Berlin Heidelberg, 2005, vol. 3587, pp. 610–619. [Online]. Available: http://dx.doi.org/10.1007/11510888_60
- [7] W. Pedrycz, Collaborative fuzzy clustering, *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1675–1686, 2002
- [8] N. Grozavu, M. Ghassany, and Y. Bennani, Learning confidence exchange in collaborative clustering, in *IJCNN*, 2011, pp. 872–879
- [9] W. Pedrycz and K. Hirota, A consensus-driven fuzzy clustering, *Pattern Recogn. Lett.*, vol. 29, no. 9, pp. 1333–1343, 2008
- [10] N. Grozavu, G. Cabanes, and Y. Bennani, Diversity analysis in collaborative clustering, in *IEEE World Congress on Computational Intelligence*, 2014
- [11] B. Depaire, R. Falcón, K. Vanhoof, and G. Wets, Pso driven collaborative clustering: A clustering algorithm for ubiquitous environments, *Intell. Data Anal.*, vol. 15, no. 1, pp. 49–68, Jan. 2011. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1937721.1937725>
- [12] M. Ghassany, N. Grozavu, and Y. Bennani, Collaborative clustering using prototype-based techniques, *International Journal of Computational Intelligence and Applications*, vol. 11, no. 03, p. 1250017, 2012
- [13] S. Zhang, C. Zhang, and X. Wu, Knowledge Discovery in Multiple Databases, ser. Advanced Information and Knowledge Processing. Springer, 2004. [Online]. Available: <http://dx.doi.org/10.1007/978-0-85729-388-6>
- [14] W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, *Pattern Recogn. Lett.*, vol. 26, no. 15, pp. 2439–2449, 2005
- [15] N. Grozavu and Y. Bennani, Topological collaborative clustering, *Australian Journal of Intelligent Information Processing Systems*, vol. 12, no. 3, 2010
- [16] M. Ghassany, N. Grozavu, and Y. Bennani, Collaborative clustering using prototype-based techniques, *International Journal of Computational Intelligence and Applications*, vol. 11, no. 3, 2012
- [17] N. Grozavu and Y. Bennani, Topological Collaborative Clustering, in *LNCS Springer of ICONIP'10 : 17th International Conference on Neural Information Processing*, 2010
- [18] T. Kohonen, Self-organized formation of topologically correct feature maps, *Biol. Cyb.*, vol. 43, pp. 59–69, 1982
- [19] Analysis of a simple self-organizing process, *Biol. Cyb.*, vol. 44, pp. 135–140, 1982

- [20] C. M. Bishop and C. K. I. Williams, GTM: The generative topographic mapping, *Neural Computation*, vol. 10, pp. 215–234, 1998
- [21] N. Grozavu, Y. Bennani, and M. Lebbah, From variable weighting to cluster characterization in topographic unsupervised learning, in *Proc. of IJCNN09, International Joint Conference on Neural Network*, 2009
- [22] N. Grozavu and Y. Bennani, Topological collaborative clustering, *Australian Journal of Intelligent Information Processing Systems*, vol. 12, no. 2, 2010
- [23] J. Sublime, N. Grozavu, G. Cabanes, Y. Bennani, and A. Cornuéjols, From horizontal to vertical collaborative clustering using generative topographic maps, *International Journal of Hybrid Intelligent Systems*, vol. 12, no. 4, 2016
- [24] L. I. Kuncheva and C. J. Whitaker, Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy, *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, May 2003
- [25] F. Gullo, A. Tagarelli, and S. Greco, Diversity-Based Weighting Schemes for Clustering Ensembles, in *SDM*, 2009, pp. 437–448
- [26] N. Grozavu, M. Ghassany, and Y. Bennani, Learning confidence exchange in collaborative clustering, in *Neural Networks (IJCNN)*, The 2011 International Joint Conference on IEEE, 2011, pp. 872–879
- [27] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988
- [28] W. M. Rand, Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, Dec. 1971
- [29] L. Hubert and P. Arabie, Comparing Partitions, *Journal of the Classification*, vol. 2, pp. 193–218, 1985
- [30] P. Jaccard, The distribution of the flora in the alpine zone, *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912
- [31] D. L. Wallace, A Method for Comparing Two Hierarchical Clusterings: Comment, *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 569–576, 1983. [Online]. Available: <http://www.jstor.org/stable/2288118>
- [32] F. Pinto, J. Carrico, M. Ramirez, and J. Almeida, Ranked Adjusted Rand: integrating distance and partition information in a measure of clustering agreement, *BMC Bioinformatics*, vol. 8, no. 1, p. 44, 2007. [Online]. Available: <http://www.biomedcentral.com/1471-2105/8/44>
- [33] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition (Morgan Kaufmann Series in Data Management Systems). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005
- [34] M. Meila, Comparing clusterings - an information based distance, *Journal of Multivariate Analysis*, vol. 98, pp. 873–895, 2007
- [35] A. Frank and A. Asuncion, *UCI machine learning repository*, 2010. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [36] T. Calinski and J. Harabasz, Dendrite method for cluster analysis, *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974
- [37] D. L. Davies and D. W. Bouldin, A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, Feb. 1979
- [38] W. J. Krzanowski and Y. T. Lai, A criterion for determining the number of groups in a data set using sum-of-squares clustering, *Biometrics*, vol. 44, no. 1, pp. 23–34, 1988. [Online]. Available: <http://www.jstor.org/stable/2531893>
- [39] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, vol. 20, no. 0, pp. 53 – 65, 1987. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0377042787901257>
- [40] K. Kivluoto, Topology Preservation in Self-Organizing Maps, *International Conference on Neural Networks*, pp. 294–299, 1996
- [41] T. Kohonen, *Self-Organizing Maps*. Berlin: Springer-Verlag, 2001



Parisa Rastin is an academic researcher of the University of Paris 13, member of the Computer Science Laboratory of Paris Nord (LIPN - UMR 7030) in the Artificial Learning and Applications (A3) team. She received a Msc degrees in Data Science and Machine Learning in 2015 and a Ph.D. in Computer Science in 2018, at the University of Paris 13. Her research interests lie in automatic and adaptive learning, change detection in data streams, unsupervised learning and complex data structures.



Basarab Matei is a member of the Computer Science Laboratory of Paris Nord (LIPN) in the LIPN UMR 7030 Artificial Learning and Applications (A3) team since September 2015 and Professor assistant (since 2004) at the Galilee Institute. He holds a Ph.D. in Applied Mathematics from Paris 6 University and a Msc degree in Applied Statistics and Optimization from the University of Bucharest. His research interests lie in the theoretical foundations as well as in the applications of artificial learning and data science, particularly for representation and optimization problems, as well as for unsupervised neural networks.



Guénaél Cabanes is an academic researcher of the University of Paris 13, member of the Computer Science Laboratory of Paris Nord (LIPN - UMR 7030) in the Artificial Learning and Applications (A3) team. He received a Ph.D. in Computer Science at the University of Paris 13, France, in 2010. He also received a Msc in Computer Model of Reasoning and Knowledges in 2007. His primary research interests are in data mining, unsupervised learning, artificial neural networks and complex structures. He has participated in several national and international research projects since the beginning of his career.



Nistor Grazavu received his Master of Computer Science degree from Marseille II University in 2006 in Fundamental Informatics. He completed his Ph.D. in Computer Science (Unsupervised Learning) in 2009 in the Computer Science Laboratory of Paris 13 University. He is currently an Associate Professor in Computer Science at the Paris 13 University. His research is with the Machine Learning and Application Team from the LIPN Laboratory. His research interests include Unsupervised Learning, Transfer Learning, Dimensionality reduction, Collaborative Learning and Matrix Factorization. He is also a member of IEEE, INNS, INNS AML group.



Younès Bennani received B.S. degree in Mathematics and Computer Science from Rouen University. Subsequently, he received the M.Sc. and the Ph.D. degree in Computer Science from The University of Paris Sud, Orsay, respectively, and the "Habilitation à Diriger des Recherches" (accreditation to lead research) degree in Computer Science from the Paris 13 University. Dr. Younès Bennani joined the Computer Science Laboratory of Paris-Nord (LIPN-CNRS) at Paris 13 University as Assistant Professor. In 2001, he was appointed to a Full Professor of computer science in the Paris 13 University. Prof. Younès Bennani's areas of expertise are unsupervised learning, deep learning, transfer learning, collaborative clustering, cluster analysis, dimensionality reduction, features selection, features construction, data visualization, and large-scale data mining. He has published 2 books and approximately 250 papers in refereed conferences proceedings or journals or as contributions in books. He is a senior member of the IEEE. and the Vice-President for Digital Transformation at the Paris 13 University - Sorbonne Paris Cité.