A specialized Xie-Beni measure for clustering with adaptive distance

Shidi Deng[1,2], Benoit Albert[2], Violaine Antoine[2], and Jonas Koko[2]

[1] Technische Universität München, Germany
shidi.deng@tum.de
[2] LIMOS, Université Clermont Auvergne, France
{benoit.albert,violaine.antoine,jonas.koko}@uca.fr

Abstract. To certify good data partitioning, it is necessary to use an evaluation measure. This measure must take into account the specificity of the modeled partition. For centroid-based fuzzy partitioning, different measures exist. However, none of them takes into account the adaptive distance that some clustering models use. In our study, we extend the Xie-Beni measure, using both the Mahalanobis distance and the Wasserstein distance. The numerical results show the relevance of our new index.

Keywords: Clustering · Internal measure · Mahalanobis distance · Xie-Beni index.

1 Introduction

Clustering is an unsupervised learning method that does not require prior class labels to implement observational learning. Clustering is employed to group collections of physical or abstract objects into multiple classes of similar objects. There are various clustering algorithms such as partition-based clustering, hierarchical clustering, density-based clustering, grid-based clustering, and model-based clustering. These clustering algorithms can also be split following the type of partition generated: a hard partition or a soft partition. A hard partition assigns with total certainty an object to a cluster, whereas a soft partition allows to produce doubt regarding the class membership of an object. Among soft partitions, the probabilistic partition is the most famous one.

Various clustering methods can be applied for a data analysis. Thus, it is important to choose among the algorithms the partition that best fits the data. For this, validity indexes have been proposed. Such indexes attempt to measure the correspondence between a partition and the underlying structure of the data.

The validity indexes can be divided into internal and external indexes. An external index, such as the Normalized Mutual Information (NMI) or the Adjusted Rand Index (ARI) [14], allows to compare two partitions. It is generally used to measure the accuracy of a clustering partition by comparing it with the partition derived from the ground truth. Inversely, an internal index seeks to describe the intrinsic structure of the data without any prior information. It employs the notion of compactness within clusters and/or the notion of separability between clusters. The compactness quantifies how much the members of each cluster are close to each other. The separability, on the other hand, measures the distance between the different clusters. Cluster validity research is a difficult task and lacks a strict theoretical background [2].

In the case of a fuzzy partition-based clustering algorithm, such as Fuzzy C-Means (FCM) [4], there exists some specific and well-known internal indexes: the *Partition Coefficient* PC, the *Partition Entropy* PE [5], and the *Fuzzy Hyper Volume index* FHV [10] are the indexes that measure only compactness. The *Fuzzy Silhouette* FS [9], the *Xie-Beni* XB [21], and the *Partition Coefficient And Exponential Separation* $PCAES$ [20] are measures combining compactness and separability.

However, with the exception of *Fuzzy Hyper Volume index* [10], they are all based on the Euclidean distance. If a clustering algorithm uses Mahalanobis distances, as it is the case for FCM-GK [12] and its extensions [1], these indexes will not take this information into account and it can lead to incorrect quantification of the compactness and separability of the partition. Plus, although the *Fuzzy Hyper Volume index* [10] handles Mahalanobis distances, it only measures the compactness of the partition. It is therefore necessary to describe a new measure adapted to the compactness and separability for clustering algorithms using Mahalanobis distances.

This study aims to propose an extension of the Xie-Beni index to deal with partitions obtained with Mahalanobis distances. The paper is organized as follows: Section 2 details the necessary knowledge to introduce the Xie-Beni index in Section 3 and its extension in Section 4. Numerical experiments are presented in Section 5 and a conclusion and perspectives are given in the last section.

2 Background

2.1 The fuzzy c-means algorithm

Let $\boldsymbol{X} = (\boldsymbol{x}_1 \dots \boldsymbol{x}_n)$ be a data set with n objects $\boldsymbol{x}_i \in \mathbb{R}^p$ and p be the number of attributes describing the objects. The objective is to obtain a partition that groups objects into c clusters $2 \leq c < n$. A fuzzy partition $\boldsymbol{U} = (u_{ij})$ is a matrix of membership degrees $(n \times c)$ such that $u_{ij} \in [0, 1]$ is the probability that the object $\boldsymbol{x}_i$ belongs to the cluster j. The FCM clustering algorithm and its variants are centroid-based methods, i.e. each cluster is identified by its centroid $\boldsymbol{\mathcal{V}} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_c\}$, $\boldsymbol{v}_j \in \mathbb{R}^p$. The notion of similarity between an object and a group is then the calculation of the distance $\boldsymbol{d}_{ij}^2$ between the object i and the center of gravity j:

- Euclidean distance in the FCM model [4, 7]

$$\boldsymbol{d}_{ij}^2 = (\boldsymbol{x}_i - \boldsymbol{v}_j)^\top (\boldsymbol{x}_i - \boldsymbol{v}_j). \tag{2.1}$$

- Mahalanobis distance in the FCM-GK model [12]

$$\boldsymbol{d}_{ij}^2 = (\boldsymbol{x}_i - \boldsymbol{v}_j)^\top \boldsymbol{S}_j (\boldsymbol{x}_i - \boldsymbol{v}_j), \tag{2.2}$$

In FCM-GK, there exists a specific Mahalanobis distance for each cluster. These Mahalanobis distances are characterized by symmetric positive definite matrices $\boldsymbol{\mathcal{S}} = \{\boldsymbol{S}_1, \ldots, \boldsymbol{S}_c\}$ also referred to as variance covariance matrices. Remark that if the variance covariance matrix $\boldsymbol{S}_j$ defined for the cluster j corresponds to the identity, it represents a Euclidean distance.

In FCM-GK, the unknown variables $(\boldsymbol{U}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{S}})$ are determined by optimizing the following problem:

$$\min_{(\boldsymbol{U}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{S}})} J(\boldsymbol{U}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{S}}) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^m \boldsymbol{d}_{ij}^2, \tag{2.3}$$

with the constraints

$$u_{ij} \geq 0, \quad \forall i, j \in [1, n] \times [1, c] \tag{2.4}$$

$$\sum_{j=1}^{c} u_{ij} = 1, \quad \forall i \in [1, n] \tag{2.5}$$

$$\sum_{i=1}^{n} u_{ij} > 0, \quad \forall j \in [1, c] \tag{2.6}$$

$$\det(\boldsymbol{S}_j) = \rho_j, \quad \forall j \in [1, c] \tag{2.7}$$

The volume constraint (2.7) has been added in order to avoid trivial minimization where all $\boldsymbol{S}_j$ matrices are set to zero.

The method used to resolve this constrained problem is the alternating optimization method (AO) [4, 7, 12]. The resulting minimization steps are described in Algorithm 1. The FCM algorithm is similar except that the co-variance matrices of the set $\boldsymbol{\mathcal{S}}$ are not updated and remain identity matrices.

2.2 The Wasserstein distance

Originating from work on the optimal transport problem, this distance models the difficulty of changing one amount of earth to another, hence its other name Earth Mover's Distance (EMD) [15, 19]. Mathematically, it is defined as the measure of the difference between two probability distributions. Let $g_1 = \mathcal{N}_1(\mu_1, \Sigma_1)$ and $g_2 = \mathcal{N}_2(\mu_2, \Sigma_2)$ be two multivariate Gaussians distribution. The 2-Wasserstein distance between the two Gaussians is:

$$W_2(g_1, g_2)^2 = \parallel \mu_1 - \mu_2 \parallel_2^2 + tr\left(\Sigma_1 + \Sigma_2 - 2\sqrt{\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2}}\right), \tag{2.8}$$

where $\parallel . \parallel_2$ is the Euclidian norm, and $tr(.)$ the trace function. In computer science, this distance is widely used for image comparison, especially in content-based image search [18] and pattern recognition [3].

Algorithm 1 FCM-GK

Intput : $\boldsymbol{X}$ the data set, c the number of cluster
$err = 0, k = 0,$
$\boldsymbol{U}^0$ random initialization.
while $err > 10^{-3}$ **do**
  $k \leftarrow k + 1$
  compute $\boldsymbol{\mathcal{V}}^k$: $\boldsymbol{v}_j^{k+1} = \frac{\sum_{i=1}^n u_{ij}^{k+1} \boldsymbol{x}_i}{\sum_{i=1}^n u_{ij}^{k+1}}$,
  compute $\boldsymbol{\mathcal{S}}^k$:

$$\boldsymbol{\Sigma}_j = \sum_{i=1}^{n} u_{ij}^{k+1} (\boldsymbol{x}_i - \boldsymbol{v}_j^{k+1})(\boldsymbol{x}_i - \boldsymbol{v}_j^{k+1})^\top$$

$$\boldsymbol{S}_j^{k+1} = \det(\boldsymbol{\Sigma}_j)^{\frac{1}{p}} (\boldsymbol{\Sigma}_j)^{-1}$$

  compute $\boldsymbol{U}^k$: $u_{ij}^{k+1} = \left[\sum_{\ell=1}^c \frac{(\boldsymbol{x}_i - \boldsymbol{v}_j^{k+1})^\top \boldsymbol{S}_j^k (\boldsymbol{x}_i - \boldsymbol{v}_j^{k+1})}{(\boldsymbol{x}_i - \boldsymbol{v}_\ell^{k+1})^\top \boldsymbol{S}_\ell^k (\boldsymbol{x}_i - \boldsymbol{v}_\ell^{k+1})} \right]^{-1}$
  $err \leftarrow \| \boldsymbol{U}^k - \boldsymbol{U}^{k-1} \|$
end while
Output : $\boldsymbol{U}^k$,$\boldsymbol{\mathcal{V}}^k$,$\boldsymbol{\mathcal{S}}^k$

3 A valitidy measure : the Xie-Beni index

Xie and Beni proposed a validity measure for fuzzy clustering to evaluate the quality of Fuzzy c-Means (FCM) cluster partitions [21]. This measure takes into account both compactness (intra-cluster gaps) and separability (distances between cluster centers) by computing a ratio between the mean quadratic error and the minimum of the squared distances between the centroids. It is widely used to compare two clustering methods [11, 13, 16].

3.1 Compactness

The compactness formulation is an extension of the "Partition Coefficient" [6] which measures the degree of overlap between fuzzy clusters. It is a weighted center-based distance, with the use of a Euclidean distance and the fuzzy partition as weights:

$$compactness = \frac{1}{n} \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^2 (\boldsymbol{x}_i - \boldsymbol{v}_j)^\top (\boldsymbol{x}_i - \boldsymbol{v}_j). \tag{3.1}$$

Remark that this formulation is very close to the FCM cost-to-minimize function (2.3).

3.2 Separability

In the Xie-Beni index, the separability is defined as the minimum Euclidean distance between two centroids:

$$separability = \min_{j,k \in [1,c], j \neq k} \| \boldsymbol{v}_j - \boldsymbol{v}_k \|_2^2 \,. \tag{3.2}$$

3.3 XB index

The Xie-Beni index (noted V_{XB}) is the ratio between compactness and separability. A good partitioning must have high compactness and high separability, so XB is an index to be minimized.

$$(\downarrow)V_{XB} = \frac{compactness}{separability} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^2 (\boldsymbol{x}_i - \boldsymbol{v}_j)^\top (\boldsymbol{x}_i - \boldsymbol{v}_j)}{n \min_{j,k\in[1,c],j\neq k} \parallel \boldsymbol{v}_j - \boldsymbol{v}_k \parallel_2^2}. \tag{3.3}$$

4 Improvement of V_{XB}

The Xie-Beni index is not appropriate for partitions obtained with clustering algorithms using a specific distance for each cluster, as FCM-GK. The two following examples presents the limits of the Xie-Beni measure and the way to extend the formulas to obtain XBMW, a new Xie-Beni index taking in account Mahalanobis distances.

4.1 Improvement of the compactness measure

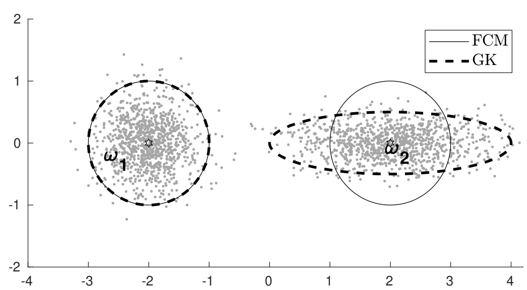


Fig. 1: Data set with two classes. The co-variance matrices obtained by FCM-GK are in dotted lines and in lines for FCM

Let us considerate a 2-dimensional data set with two well-separated classes as shown in the Figure 1. The first class has a spherical structure whereas the second class is characterized by an ellipsoidal shape. The FCM and FCM-GK algorithms have been applied on the data set and the obtained co-variances matrices are presented Figure 1. Note that FCM is represented by identity covariance matrices.

For the first cluster ω_1, both methods detect the same structure. Thus, the compactness is the same. For the second cluster ω_2, the FCM-GK method better detects the real shape of the cluster and should have a better compactness than the FCM algorithm. However, since the compactness measured by the Xie-Beni index uses the Euclidean distance, the values are similar.

Therefore, we propose to modify the Euclidean distance by the Mahalanobis distance :

$$compactness_m = \frac{1}{n} \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^2 (\boldsymbol{x}_i - \boldsymbol{v}_j)^\top \boldsymbol{S}_j (\boldsymbol{x}_i - \boldsymbol{v}_j). \tag{4.1}$$

When $\boldsymbol{S}_j = I$ for all $j \in [1, c]$ then the new compactness measure is similar to the compactness measure of the Xie-Beni index.

4.2 Improvement of the separability measure

In the V_{XB} index, separability is the minimum Euclidean distance between two centroids. Such distance does not take into account the possible difference of importance between attributes that can exists with ellipsoidal shapes. Let us consider an example of three clusters where the second and third cluster have the same centroids but different variance co-variance matrices (cf. Figure 2). The Euclidean distance $d(\omega_1, \omega_2)$ between the cluster $(\omega_1 : \boldsymbol{v}_1, \boldsymbol{S}_1)$ and the cluster $(\omega_2 : \boldsymbol{v}_2, \boldsymbol{S}_2)$ is the same as the Euclidean distance $d(\omega_1, \omega_3)$ between the cluster $(\omega_1 : \boldsymbol{v}_1, \boldsymbol{S}_1)$ and the cluster $(\omega_3 : \boldsymbol{v}_2, \boldsymbol{S}_3)$. It can be noticed in this example that

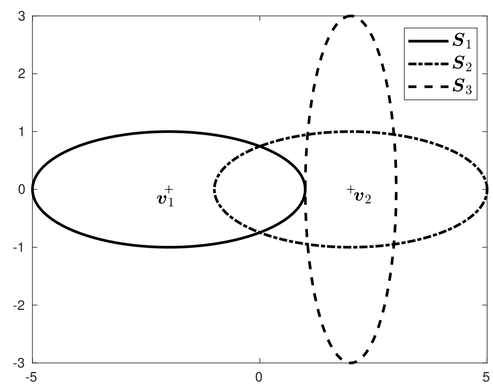


Fig. 2: Figure of two clusters with different shapes

cluster 3 gives much more importance to the attributes carried by the ordinate axis, unlike the two other clusters. We propose to use the Wasserstein distance to measure the difference between two clusters, considering that a cluster can be approximated as a distribution characterized by the mean being the centroid

($\mu = \boldsymbol{v}$) and the variance-covariance matrix being the inverse of the distance matrix ($\Sigma = \boldsymbol{S}^{-1}$). The distance between the two clusters is then:

$$W_2(\omega_j, \omega_k)^2 = \| \boldsymbol{v}_j - \boldsymbol{v}_k \|_2^2 + tr\left(\boldsymbol{S}_j^{-1} + \boldsymbol{S}_k^{-1} - 2\sqrt{\boldsymbol{S}_k^{-1/2}\boldsymbol{S}_j^{-1}\boldsymbol{S}_k^{-1/2}}\right). \quad (4.2)$$

The separability with the Wasserstein distance is

$$separability_w = \min_{j,k\in[1,c],j\neq k} W_2(\omega_j, \omega_k)^2. \quad (4.3)$$

When the distance matrices are all equal as in FCM where $\boldsymbol{S}_j = I, \forall j \in [1, c]$, then the Wasserstein distance is equal to the Euclidean distance.

4.3 XBMW : a new Xie-Beni index

Our new index, referred to as V_{XBMW}, is an extension of V_{XB} using the Mahalanobis distance for the compactness and the Wasserstein distance for the separability:

$$(\downarrow)V_{XBMW} = \frac{compactness_m}{separability_w} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^2(\boldsymbol{x}_i - \boldsymbol{v}_j)^\top \boldsymbol{S}_j(\boldsymbol{x}_i - \boldsymbol{v}_j)}{n \min_{j,k\in[1,c],j\neq k} W_2(\omega_j, \omega_k)^2}. \quad (4.4)$$

5 Numerical experimentation

5.1 Methodology

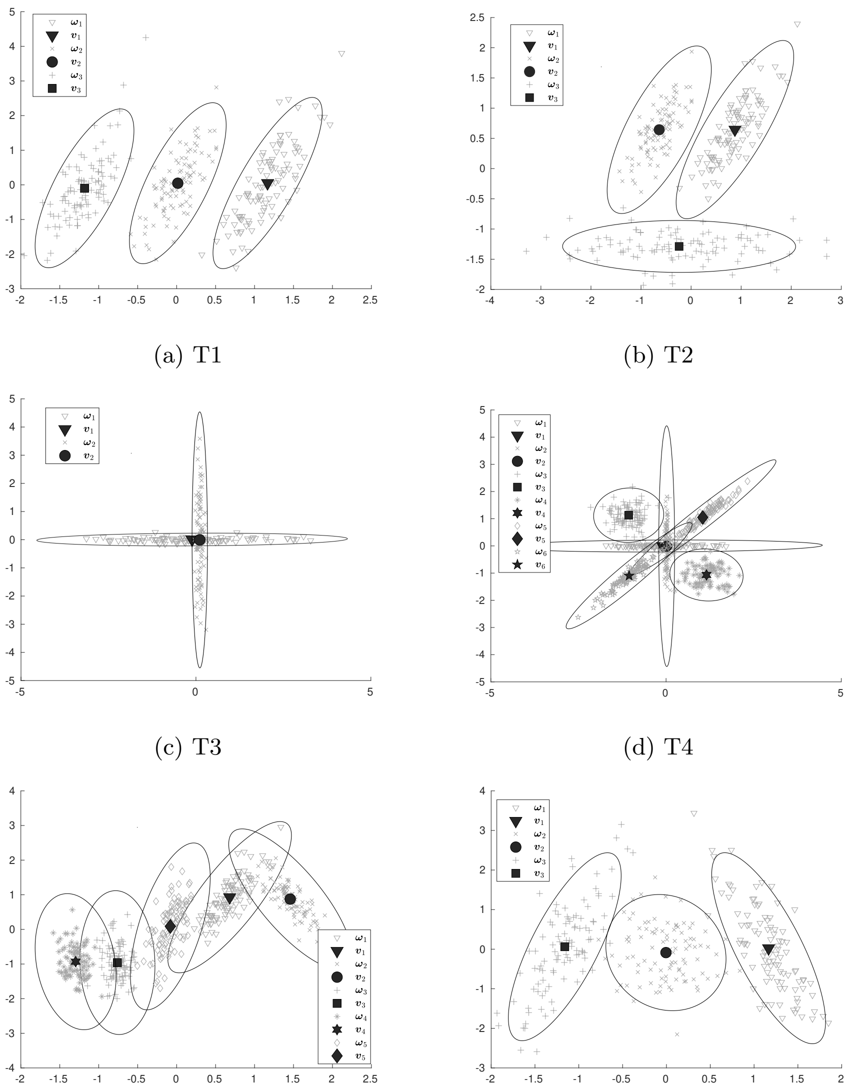
In this section, we evaluate the performance of our index. The idea is to show that there exists a better correlation between an external measure and our internal measure than between the same external measure and the Xie-Benie index. The clustering methods used for the experiments are FCM and FCM-GK. Each algorithm is run 10 times with different centroids initializations and only the partition minimizing the cost function (2.3) is kept.

5.2 Datasets

We used 19 datasets, 6 toys datasets, and 9 from the UCI library [3]: Algerian forest(Af), Drybean(Db), Glass, Iris, classes I, J, and L from Letters(IJL) [8], Seeds, WDBC, Wifi, Wine. We also used two synthetic datasets: Asymetric and Skewed [17]. Table 1 references their characteristics, i.e. the number of classes c, the number of objects n, and the number of attributes p. All datasets are normalized, i.e. centered (mean) and reduced (std) for each attribute.

We also have created six toy datasets using a combination of cluster ω. Each cluster corresponds to a specific Gaussian for which 100 points have been generated. The characteristics of each Gaussian is given in the Table 2: mean value $\boldsymbol{v}$, axis lengths a, b and rotation angle θ. We note $-\omega$, the cluster whose mean is the opposite $-\boldsymbol{v}$. The data set T1 is composed of $(\omega_1, \omega_2, -\omega_1)$, T2 : $(\omega_1, \omega_2, \omega_3)$, T3 : (ω_4, ω_5), T4 : $(\omega_4, \omega_5, \omega_6, -\omega_6, \omega_7, -\omega_7)$,T5 : $(\omega_1, \omega_8, \omega_9, \omega_{10}, \omega_{11})$, and T2 : $(\omega_{12}, \omega_{13}, -\omega_1)$. Figure 3 shows the obtained datasets.

[3] https://archive.ics.uci.edu/ml/datasets.php



(a) T1

(b) T2

(c) T3

(d) T4

(e) T5

(f) T6

Fig. 3: Toys datasets

Table 1: Characteristics of datasets.

	Af	Db	Glass	Iris	IJL	Seeds	WDBC	Wifi	Wine	Asymetric	Skewed
c	2	7	2	3	3	3	2	4	3	5	6
n	243	13611	214	150	2263	210	569	2000	178	1000	1000
p	10	16	9	4	16	7	30	7	13	2	2

Table 2: Characteristics of Gaussians (i.e clusters)

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}	ω_{12}	ω_{13}
$\boldsymbol{v}$	$\begin{pmatrix}\frac{3}{5}\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}\frac{-2}{5}\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$	$\begin{pmatrix}-3\\3\end{pmatrix}$	$\begin{pmatrix}3\\3\end{pmatrix}$	$\begin{pmatrix}1.2\\0\end{pmatrix}$	$\begin{pmatrix}\frac{-1}{2}\\\frac{-1}{3}\end{pmatrix}$	$\begin{pmatrix}\frac{-9}{10}\\\frac{-1}{3}\end{pmatrix}$	$\begin{pmatrix}0\\\frac{-1}{6}\end{pmatrix}$	$\begin{pmatrix}\frac{3}{5}\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\end{pmatrix}$
a	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$	2	2	1	2	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
b	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{10}$	$\frac{1}{10}$	1	$\frac{1}{4}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{18}$
θ	30	30	0	0	90	0	45	-30	0	0	45	-30	0

5.3 External evaluation measure

We used the Ajusted Rand Index[14], which compares two hard partitions. Since FCM and FCM-GK generates fuzzy partitions, these partitions are transformed into hard partitions by assigning to each object the class with the highest membership. Let π_1 and π_2 be two partitions, a be the number of pairs of objects which are in the same group in π_1 and π_2, b be the number of pairs of objects which are in different groups in π_1 and π_2, c be the number of pairs that are in the same group in π_1 but not in π_2 and d be the number of pairs that are in the same group in π_2 but not in π_1. The ARI is then defined as follows:

$$ARI(\pi_1, \pi_2) = \frac{2(ab - cd)}{(a+d)(d+b) + (a+c)(c+b)}$$

If two partitions are identical then the ARI score is one. The better partitioning will have a higher ARI score and a lower index value.

5.4 Results

A better partitioning is a larger ARI and a smaller index. We use a simple matching coefficient (SMC), between the difference in the ARI score for FCM and GK, and the difference in the index. When ARI increases and the index decreases it is a true positive (TP), but if the index increases then it is a false negative (FN). If ARI decreases and the index decreases it is a false positive (FP) but if the index increases it is a true negative (TN).

$$SMC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Table 3: Matching between XB, XBMW and ARI

	TP	TN	FP	FN	SMC
V_{XB}	2	**4**	**1**	10	0.35
V_{XBMW}	**11**	2	3	**1**	**0.76**

As it can be observed Table 3, there exists a better matching for our new index. Details are given in the Tables 4,5, and 6. This is especially the case for the toy sets, which allow us to highlight our index. We observe that for case T1, the limit of the Wasserstein distance is because if the clusters have the same shape then it will be equal to the Euclidean distance. Our new index favors GK (chosen 14 times out of 17) contrary to Xie Beni's index which favors FCM (chosen 14 times out of 6).

Let us also remark that our new index is more sensitive to a high number of attributes, especially for the WDBC dataset.

Table 4: ARI, XB, XBMW for toys datasets

(a) T1

	FCM	GK	
ARI	0.42	1	
V_{XB}	0.18	0.61	FN
V_{XBMW}	0.18	0.21	FN

(b) T2

	FCM	GK	
ARI	0.79	0.97	
V_{XB}	0.18	0.28	FN
V_{XBMW}	0.18	0.13	**TP**

(c) T3

	FCM	GK	
ARI	0.26	0.86	
V_{XB}	0.72	33.3	FN
V_{XBMW}	0.72	0.005	**TP**

(d) T4

	FCM	GK	
ARI	0.61	0.91	
V_{XB}	0.33	13.5	FN
V_{XBMW}	0.33	0.01	**TP**

(e) T5

	FCM	GK	
ARI	0.41	0.93	
V_{XB}	0.33	0.68	FN
V_{XBMW}	0.33	0.31	**TP**

(f) T6

	FCM	GK	
ARI	0.27	0.96	
V_{XB}	0.20	0.53	FN
V_{XBMW}	0.20	0.14	**TP**

Table 5: ARI, XB, XBMW for Synthetic datasets

(a) Asymetric

	FCM	GK	
ARI	0.89	0.96	
V_{XB}	0.09	0.12	FN
V_{XBMW}	0.09	0.06	**TP**

(b) Skewed

	FCM	GK	
ARI	0.65	0.99	
V_{XB}	0.24	0.66	FN
V_{XBMW}	0.24	0.06	**TP**

Table 6: ARI, XB, XBMW for UCI datasets

(a) Algerian forest

	FCM	GK	
ARI	0.34	0.54	
V_{XB}	0.35	0.38	FN
V_{XBMW}	0.35	0.01	**TP**

(b) Dry bean

	FCM	GK	
ARI	0.68	0.70	
V_{XB}	16.55	0.64	**TP**
V_{XBMW}	16.55	6.10^{-6}	**TP**

(c) Glass

	FCM	GK	
ARI	0.55	0.41	
V_{XB}	1.45	0.84	**TN**
V_{XBMW}	1.45	2.10^{-3}	FP

(d) Iris

	FCM	GK	
ARI	0.63	0.74	
V_{XB}	0.22	0.79	FN
V_{XBMW}	0.22	0.16	**TP**

(e) IJL

	FCM	GK	
ARI	0.04	0.26	
V_{XB}	7.06	1.15	FN
V_{XBMW}	7.06	0.10	**TP**

(f) Seed

	FCM	GK	
ARI	0.77	0.72	
V_{XB}	0.21	0.22	**TN**
V_{XBMW}	0.21	0.01	FP

(g) WDBC

	FCM	GK	
ARI	0.68	0.41	
V_{XB}	0.48	2.16	**TN**
V_{XBMW}	0.48	0.02	FP

(h) Wifi

	FCM	GK	
ARI	0.82	0.41	
V_{XB}	0.34	6.10^{4}	**TN**
V_{XBMW}	0.34	1.10^{4}	**TN**

(i) Wine

	FCM	GK	
ARI	0.90	0.33	
V_{XB}	0.47	70.0	**TN**
V_{XBMW}	0.47	4.19	**TN**

6 Conclusion

In this study, the interest was to take into account the adaptability of the metrics to measure the quality of the partitioning methods. Indeed, for the internal criteria, it is important to evaluate the compactness and separability according to the particular distances of each cluster. This is why we have extended the Xie-Beni measure with the Mahalanobis distance for the compactness and the Wasserstein distance for the separability. We compared two methods, one based on Euclidean distance (FCM) and its variant based on adaptive distances (FCM-GK). The results are satisfactory as the index allows us to analyze a good fit with an external measure.

This study is encouraging and offers some perspectives. First of all, it would be interesting to compare two clustering methods that are both using Mahalanobis distances. We can also consider selecting another metric for separability. Finally, we focused our study on the Xie-Beni index, but it could be interesting to adapt other internal validation measures to the Mahalanobis distances, in particular to find an optimal number of clusters.

References

1. Antoine, V., Quost, B., Masson, M.H., Denœux, T.: CECM: Constrained evidential c-means algorithm. Computational Statistics & Data Analysis **56**(4), 894–914 (2012)

2. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I.: An extensive comparative study of cluster validity indices. Pattern recognition **46**(1), 243–256 (2013)
3. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International conference on machine learning. pp. 214–223. PMLR (2017)
4. Bezdek, J.C.: Fuzzy Mathematics in pattern classification. Cornell University (1973)
5. Bezdek, J.C.: Numerical taxonomy with fuzzy sets. Journal of mathematical biology **1**(1), 57–71 (1974)
6. Bezdek, J.C.: Objective function clustering. In: Pattern recognition with fuzzy objective function algorithms, pp. 43–93. Springer (1981)
7. Bezdek, J., Dunn, J.: Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. IEEE Transactions on Computers **100**(8), 835–838 (1975)
8. Bilenko, M., Basu, S., Mooney, R.J.: Integrating constraints and metric learning in semi-supervised clustering. In: Proceedings of the twenty-first international conference on Machine learning. p. 11 (2004)
9. Fukuyama, Y.: A new method of choosing the number of clusters for the fuzzy c-mean method. In: Proc. 5th Fuzzy Syst. Symp., 1989. pp. 247–250 (1989)
10. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence **11**(7), 773–780 (1989)
11. Ghosh, A., Mishra, N.S., Ghosh, S.: Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. Information Sciences **181**(4), 699–715 (2011)
12. Gustafson, D., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: 1978 IEEE conference on decision and control including the 17th symposium on adaptive processes. pp. 761–766. IEEE (1979)
13. Huang, C., Molisch, A.F., Geng, Y.A., He, R., Ai, B., Zhong, Z.: Trajectory-joint clustering algorithm for time-varying channel modeling. IEEE Transactions on Vehicular Technology **69**(1), 1041–1045 (2019)
14. Hubert, L., Arabie, P.: Comparing partitions. Journal of classification **2**(1), 193–218 (1985)
15. Kantorovich, L.V.: Mathematical methods of organizing and planning production. Management science **6**(4), 366–422 (1960)
16. Mitra, S., Pedrycz, W., Barman, B.: Shadowed c-means: Integrating fuzzy and rough clustering. Pattern recognition **43**(4), 1282–1291 (2010)
17. Rezaei, M., Fränti, P.: Can the number of clusters be determined by external indices? IEEE Access **8**, 89239–89257 (2020)
18. Valle, M.E., Francisco, S., Granero, M.A., Velasco-Forero, S.: Measuring the irregularity of vector-valued morphological operators using wasserstein metric. In: Discrete Geometry and Mathematical Morphology: First International Joint Conference, DGMM 2021, Uppsala, Sweden, May 24–27, 2021, Proceedings. pp. 512–524. Springer (2021)
19. Vaserstein, L.N.: Markov processes over denumerable products of spaces, describing large systems of automata. Problemy Peredachi Informatsii **5**(3), 64–72 (1969)
20. Wu, K.L., Yang, M.S.: A cluster validity index for fuzzy clustering. pattern recognition letters **26**(9), 1275–1291 (2005)
21. Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis & Machine Intelligence **13**(08), 841–847 (1991)