# Fuzzy Cluster Validity with Generalized Silhouettes

**Mohammad Rawashdeh and Anca Ralescu**
School of Computing Sciences and Informatics
University of Cincinnati
Cincinnati, USA
rawashmy@mail.uc.edu; Anca.Ralescu@uc.edu

**Abstract**

A review of some popular fuzzy cluster validity indices is given. An index that is based on the generalization of silhouettes to fuzzy partitions is compared with the reviewed indices in conjunction with fuzzy *c*-means clustering.

## Introduction

Prevalent in many applications, (Jain, Murty, & Flynn 1999), the problem of clustering involves design decisions about representation (i.e. set of features), similarity measure, criterion and mechanism of a clustering algorithm. The clustering literature is very rich in various schemes that address these ingredients (Jain, Murty, & Flynn 1999; Xu & Wunsch 2005). However, the problem itself is centered on the intuitive and easily stated goal of partitioning a set of objects into groups such that *objects within one group are similar to each other and dissimilar to objects in other groups*, which has become a common description of clustering (Jain, Murty, & Flynn 1999; Berkhin 2002; Kleinberg 2002; Xu & Wunsch 2005). As opposed to classification, only few of the existing clustering algorithms are widely used. Indeed, clustering is less appreciated among practitioners of data analysis due to the lack of class labels. Labels are used in the evaluation of loss functions, formal assessments of the goal. This has encouraged researchers to treat clustering in a semi-supervised manner by incorporating as much information as available such as in *must-link* and *cannot-link* constraints (Blienko, Basu, & Mooney 2004) in order to achieve satisfactory results. Usually, there is an end-goal to clustering of a dataset or an end-use of the final clustering. For example, clustering of documents by topic, clustering of images by common content and clustering of proteins by function have as respective end goal a better understanding of a corpus of documents, or of one or more proteins. This suggests that a better treatment for clustering should be in the context of end-use rather than in an application-independent mathematical manner (Guyon, von Luxburg, & Williamson 2009). Accordingly, the unknown desired clustering is the only ground truth assumed about the problem. The properties of the similarity measure sufficient to cluster well, that is, to achieve low error with respect to the ground-truth clustering, are given in (Balcan, Blum, & Vempala 2008). The features, the measure and the algorithm all should be chosen in the context of the end-use. For example, it would be unwise to employ a measure that pairs two images because they show the same person while a clustering by facial expression is desired. This applies to the set of features as well; the features should accommodate for the different possible expressions. In the absence of end-use, clustering becomes an exploratory approach to data analysis, looking for the right ingredients to get the best structure.

The *c*-means, alternatively *k*-means (MacQueen 1967), is one popular clustering algorithm that partitions a set of data points $X = \{x_j | j = 1, .., n\}$ into disjoint subsets $U = \{u_i | i = 1, \dots, c\}$. The exclusive cluster assignment characterizes hard clustering and hence it is also referred by hard *c*-means (HCM). Fuzzy *c*-means (FCM) family of algorithms imposes relaxed constraints on cluster assignment by allowing nonexclusive but partial memberships, thereby, modeling cluster overlapping. The first FCM algorithm was proposed in (Dunn 1973). Its convergence was later improved in (Bezdek 1981). Both schemes, crisp and fuzzy, optimize a variance-criterion with respect to cluster center and point membership for the specified cluster number. The final clustering is given by a membership matrix, $U = [u_{ij}]$; $u_{ij}$ is the membership of $x_j$ in $u_i$. When $u_{ij}$ assumes values in {0,1} or [0,1], the matrix characterizes crisp or fuzzy partitions respectively.

It is common to define the pairwise similarities-dissimilarities in terms of distances which, in turn, give a structure i.e. the dataset underlying structure. The clustering algorithm, by processing the pairwise distances implicitly or explicitly, produces a structure, a partition. Its success is determined by the extent to which the produced partition aligns with the underlying structure, or more precisely, agrees with the pairwise distances. Supplying inconsistent values for $c$, forces the algorithm either to separate similar points or to group dissimilar points in the same cluster. Hence the issue of cluster number is crucial and largely affects clustering quality. Even by choosing features and a measure consistent with the end-use, the inherent number of clusters might be unknown. For example, in a topic-driven clustering application, terms that are significant to each possible topic or common theme might be universally known, but the number of topics

represented by documents in a particular dataset is unknown. Even if the cluster number is guessed correctly, there is the unfortunate possibility of obtaining a suboptimal clustering due to local optimum convergence. Besides, clustering of different types, crisp versus fuzzy, can be obtained on the same dataset. For "this and that kind" of reasons, cluster analysis is incomplete without the assessment of *clustering quality*. The issues of cluster number and quality are the main concerns of cluster validity.

This study reviews some fuzzy cluster validity indices then presents a generalization of silhouettes to fuzzy partitions. The performance of all reviewed indices is compared, with discussion, using two different datasets.

## Fuzzy *c*-Means Algorithm

FCM, described in (Bezdek, Ehrlich, & Full 1984), incorporates fuzzy membership values in its variance-based criterion as

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \,.\, d^2(x_j, v_i), \tag{1}$$

where $v_i$ is the center of cluster $u_i$. The clustering mechanism is carried as a Picard iterative process that alternates the application of

$$v_i = \frac{\sum_{j=1}^{n} u_{ij} \,.\, x_j}{\sum_{j=1}^{n} u_{ij}}, \tag{2}$$

and

$$u_{ij} = \left[\sum_{r=1}^{c}\left(\frac{d^2(x_j, v_i)}{d^2(x_j, v_r)}\right)^{1/(m-1)}\right]^{-1}. \tag{3}$$

The update rules are derived from the necessary conditions of (1) constrained by

$$\sum_{i=1}^{c} u_{ij} = 1;\ \forall j\,, \tag{4}$$

$$\sum_{j=1}^{n} u_{ij} > 0;\ \forall i\,. \tag{5}$$

FCM output ranges from crisp partitions, as produced by HCM, to the fuzziest possible partition for the specified number of clusters i.e. $\mathrm{U}_{c\,x\,n} = [1/c]$. Informally speaking, there are two sources for fuzziness in a produced partition. First is the amount of overlapping in the underlying structure; equation (3) assigns each point almost the same membership to overlapping clusters whose centers are within small proximity. Second is the exponent; the ratios in (3) become compressed around the value 1 when $m$ is too high, thereby, weakening the role of 'geometry' as a key factor in shaping membership values.

## Cluster Validity

Modeling the pairwise similarities-dissimilarities by a distance measure restates the goal of clustering as the search for optimally *compact* and *separated* clusters. One cluster is compact only if its member points are within small proximity from each other. Two clusters are well separated only if their member points are distant from each other. Accordingly, the variance-based criterion found in *c*-means can be thought of as a measure of compactness, which was shown to be equivalent to a measure of separation for the same number of clusters (Zhao & Karypis 2001). Hence, *c*-means is capable of producing partitions that are optimally compact and well separated, for the specified number of clusters. Note that better clustering might still be achieved by specifying different cluster numbers. Since clustering algorithms are supposed to optimize their output in compactness and separation, both should be assessed to find clustering quality.

One might need to distinguish between the *desired structure*, the *underlying structure*, and *candidate structures* produced by clustering algorithms. The desired structure is the ground truth clustering, mentioned earlier. What is known about this clustering might be vague or incomplete but it should drive the problem design. The underlying structure is the one shaped by the pairwise distances which suggests unique clustering (Fig. 1a), multiple clusterings (Fig. 1b) or no clustering due to the lack of any structure (Fig. 1c). A clustering algorithm produces different partitions for different configurations i.e. distance measure, parameters, etc. The best case scenario is when the pairwise distances structure the points into the desired grouping and an algorithm successfully produces a clustering that aligns with the underlying structure. Validating a produced clustering with respect to the underlying structure is possible by means of cluster validity indices.

### Partition Coefficient

The partition coefficient, *PC*, is defined as the Frobenius norm of the membership matrix, divided by the number of points, as in

$$PC(\mathrm{U}) = \frac{1}{n} \,.\, \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2. \tag{6}$$

The coefficient is bounded by $1/c$ and 1, if applied to FCM output. Its use as a measure of partition fuzziness was first investigated by Bezdek in his Ph.D. dissertation (Bezdek

1973). Although it can be used as a validity index with some success, it has been shown to be irrelevant to the problem of cluster validity (Trauwaert 1988). Clearly, the coefficient does not incorporate the pairwise distances that are necessary to the assessment of compactness and separation. Therefore, it is not reliable for the validation of any given partition, for example, one produced by random cluster assignment. Also, the coefficient assumes its upper value on any crisp partition, regardless of its clustering quality. Nevertheless, the coefficient does what it knows best, measuring fuzziness.

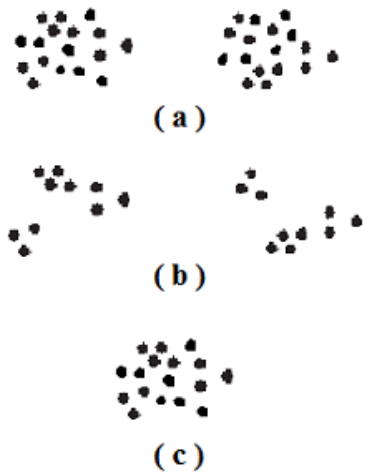


Figure 1. The structure of the dataset suggests (a) $c$ = 2, (b) $c$ = 2, 4 and 6, (c) $c$ = 1; no structure.

## Xie-Beni Index

An index that really measures compactness and separation was proposed by Xie and Beni, *XB* index (Xie & Beni 1991). *XB* takes the form of a ratio; the minimum center-to-center distance appears in its denominator and $J_2$ , as exactly as in FCM, is in its numerator but divided by $n$. Hence, *XB* is a measure of compactness divided by a measure of separation, given by

$$XB(\mathrm{U},V;\ X) = \frac{XB_{cmp}/n}{XB_{spr}} = \frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{2}\ .\ d^{2}(x_j,v_i)}{\min\{\ d(v_r,v_s)\mid r\ <\ s\}}. \tag{7}$$

In (7), $V = \{v_i \mid i = 1,\dots,c\}$ denotes the set of cluster centers. The authors define the variation of each cluster as the sum of point fuzzy deviations, *squared*. With respect to a cluster, a point deviation is its distance from the cluster center weighted by its membership. The total variation is the sum of all cluster variations that gives the compactness of the partition, when divided by $n$. This description explains why memberships and distances in (7) are squared. However, they suggest substituting $u_{ij}^{m}$ in place of $u_{ij}^{2}$ where $m$ is the same as the value used in FCM, justified by making the index 'compatible' with FCM. The final value of (1) can be directly plugged in (7), provided it is still available, or else recomputed. It is unclear how being compatible with FCM or raising membership values to powers different than 2 relates to the assessment of compactness and separation, or to the 'geometry' underlying the data.

## Fuzzy Hypervolume

The development of the index started as part of work that formulates clustering as a problem of maximum likelihood estimation (MLE) of a mixture of multivariate densities (Wolfe 1970); the dataset is assumed to be drawn from such a mixture. Bezdek and Dunn, in (Bezdek & Dunn 1975), give the MLE algorithm and FCM as well. The MLE algorithm solves for a composite parameter vector of densities' means, covariance matrices and the a priori probabilities. They describe the use of FCM to approximate the ML parameters. Substituting FCM-generated membership values for posterior probabilities computes the remaining ML parameters. A justification is given by comparing the behavior of two update rules in both algorithms. They produce small values when evaluated on data points that are distant from some density-cluster center relative to their distance from nearest center. However, they point out the fact that both algorithms compute different centers and distances.

Gath and Geva in (Gath & Geva 1989), first, give a similar description of FCM, and fuzzy MLE that is derived from FCM, as opposed to the separate treatment given by Bezdek and Dunn. Then, they suggest a 2-stage clustering scheme, FCM followed by MLE, justified by the unstable behavior of MLE as a standalone clustering scheme. As part of their work, some validity measures were proposed; among them is the *fuzzy hypervolume* measure (*FHV*). *FHV* is defined in terms of the covariance matrix determinants. The covariance matrix of cluster $u_i$ can be constructed using

$$\mathrm{F}_i = \frac{\sum_{j=1}^{n} u_{ij}\,.\,(x_j - v_i)\,(x_j - v_i)^{T}}{\sum_{j=1}^{n} u_{ij}}. \tag{8}$$

The *hypervolume* is then computed by

$$FHV(\mathrm{U},V;X) = \sum_{i=1}^{c} [\det(\mathrm{F}_i)]^{1/2}\,. \tag{9}$$

The determinants are functions of cluster spreads and point memberships. A clustering that is evaluated the smallest is assumed to be optimal. However, the following observations can be made:

- According to the authors, the index is sensitive to substantial overlapping in the dataset.

- It is unclear how the measure accounts for compactness and separation.
- Assuming that an MLE mixture has been successfully found, is it the best clustering in compactness and separation?
- Is the measure applicable to crisp partitions?
- The use of FCM as MLE requires setting m=2; how does the measure performs on partitions obtained using *m* different than 2?

## Pakhira-Bandyopadhyay-Maulik Index

Pakhira et el. proposed an index, referred here as *PBM*, that targets both fuzzy and crisp partitions (Pakhira, Bandyopadhyay, & Maulik 2004). Its fuzzy version is defined as

$$PBM(\mathrm{U}, V, c, m; X) = \left(\frac{\sum_{j=1}^{n} d(x_j, v)}{c} \cdot \frac{\max\{\, d(v_r, v_s) \mid r \neq s\}}{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \,.\, d(x_j, v_i)}\right)^2. \tag{10}$$

In (10), $v$ is the center of the whole dataset. The index can be factorized into a measure of compactness

$$PBM_{cmp} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \,.\, d(x_j, v_i), \tag{11}$$

a measure of separation

$$PBM_{spr} = \max\{\, d(v_r, v_s) \mid r \neq s\}, \tag{12}$$

and an artificial factor, irrelevant to compactness and separation,

$$PBM_x = \frac{\sum_{j=1}^{n} d(x_j, v)}{c}. \tag{13}$$

The whole term in (10) is raised to power two that is also irrelevant to the assessment of clustering quality. Larger values of the index are assumed to indicate better clustering. It can be noticed though that the quantity in (12) does not necessarily capture the separation between all pairs of clusters; an authentic separation measure should account for the poor separation found in partitions into large number of clusters. .

## Average Silhouette Index

Rousseeuw proposed an index for the validation of crisp partitions (Rousseeuw 1987). It is based on the notion of *silhouette*. A silhouette, constructed for each data point, measures the clustering quality for that point. The average over members of the whole dataset or an individual cluster is a measure of the set clustering quality. To illustrate silhouette construction, consider for the data point $x_k$ , the cluster to which $x_k$ has been assigned, $A$, and let $C$ be any cluster different than $A$. The silhouette $s_k = s(x_k)$ is defined in terms of a measure of compactness $a_k$ and a measure of separation $b_k$ . The average distance of $x_k$ to points in $A$ computes $a_k$, while $b_k$ is the minimum average distance from $x_k$ to all other clusters. Let $B$ denotes the cluster corresponding to $b_k$ (see Fig. 2).

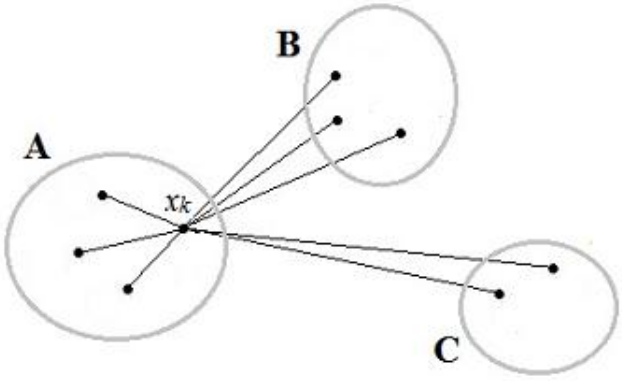


Figure 2. With respect to $x_k$, $a_k$ is the average length of lines within $A$ and $b_k$ is the average length of lines between $A$ and $B$.

Then the silhouette of $x_k$ is defined by

$$s_k = \frac{b_k - a_k}{\max\{a_k, b_k\}}. \tag{14}$$

Clearly, (14) evaluates to values in $[-1, 1]$. The average silhouette over a cluster $u_i$ or the whole dataset $X$ are given respectively by

$$Sil(u_i) = \frac{\sum_{x_j \in u_i} s_j}{|u_i|} = \frac{\sum_{j=1}^{n} u_{ij} \,.\, s_j}{\sum_{j=1}^{n} u_{ij}}, \tag{15}$$

$$Sil(X) = \frac{\sum_{j=1}^{n} s_j}{n}. \tag{16}$$

Note that the membership values in (15) are crisp and $u_i$ denotes the $i^{th}$ cluster, as a set.

From the data point perspective, the measure assumes positive values if the separation distance is larger than the compactness distance and negative values if vice versa. A value near zero indicates that the point is at clusters boundary region, of course in the context of clustering on hands. At the coarser cluster level, the average silhouette indicates weak structure if near zero, strong if near +1 and misclustering if near -1. Since a clustering algorithm cannot do any better than the underlying structure, an average close to +1 is attainable only in the presence of a strong structure.

The following appealing properties recommend the silhouette index:

- As opposed to other indices, it validates a given clustering at point level, providing thus the finest granularity.
- It is algorithm-independent.
- It takes as input only the pairwise similarities-dissimilarities and the membership matrix.
- As explained in the original work of Rousseeuw, it can be used to 'visualize' the clustering quality of a given partition.
- Its assessment of compactness and separation conforms literally to the stated goal of clustering; a relatively small $a_k$ compared to $b_k$ means that $x_k$ has been successfully grouped with its similar points in the same cluster in a way that separates from its dissimilar points.

### Extended Average Silhouette Index

The above construction of silhouettes is not directly applicable to fuzzy partitions since it requires crisp cluster boundaries, necessary to the computation of cluster average distances. Nevertheless, a fuzzy partition might be validated by silhouettes after being defuzzified, for example, by setting the maximum membership degree of each point to one and nullifying the rest. However, this discards cluster overlapping, defeating the reason of using FCM not HCM. An extension that integrates fuzzy values with silhouettes, computed from the defuzzified partition, into an average silhouette-based index was proposed in (Campello & Hruschka 2006). They suggest computing a weighted mean in which each silhouette is weighted by the difference in the two highest fuzzy membership values of the associated point. More precisely, if $p(j)$ and $q(j)$ denote cluster indices with the two highest membership values associated with $x_j$ then the index is given by

$$eSil(X) = \frac{\sum_{j=1}^{n}\left(u_{p(j)j} - u_{q(j)j}\right) \cdot s_j}{\sum_{j=1}^{n}\left(u_{p(j)j} - u_{q(j)j}\right)}, \qquad (17)$$

Therefore, points around cluster centers become significant to the computation of the index since they have higher weights, as opposed to the insignificant points found in overlapping regions. Clearly, such an assessment is not thorough since it tends to ignore the clustering of points in overlapping regions.

## Generalized Intra-Inter Silhouette Index

A generalization of silhouettes to fuzzy partitions is given in (Rawashdeh & Ralescu), based on the following central observations:

- A partition of a set of points into any number of clusters is essentially a clustering of the associated pairwise distances into *intra-distances* (within-cluster) and *inter-distances* (between-cluster).
- A strong structure, a good clustering, has small intra-distances and large inter-distances i.e. similar points are grouped together and dissimilar points are separated.
- In the context of a crisp partition, each distance is either intra-distance or inter-distance. This is modeled by intra-inter scores associated to a distance that assume the values 0 and 1, indicating distance membership.
- In the context of a fuzzy partition, two points belong to each cluster simultaneously and separately with some degree, intuitively suggesting the assignment of fuzzy intra-inter scores to the pairwise distances,

The original construction of silhouettes, which already incorporates the pairwise distances, is reformulated to employ intra-inter scores. The following is applicable to both crisp and fuzzy partitions, and it carries similar computation as in the original construction, provided that the partition is crisp. As input, the construction requires the pairwise distances $D_{n\,x\,n}$ and the membership matrix $U_{c\,x\,n}$.

**Step 1.** Given a partition into $c$ clusters, each distance $d_{jk}$, associated with $x_j$ and $x_k$ , is intra-distance with respect to either cluster and inter-distance with respect to any of the 2-combinations of $c$ clusters. The following constructs all of the $(n\,x\,n)$ intra-inter matrices:

$$\text{IntraDist}_i = \left[\, intra_i(d_{jk})\right];\ 1 \le i \le\ c,$$
$$intra_i(d_{jk}) =\ (u_{ij} \wedge u_{ik}). \qquad (18)$$

$$\text{InterDist}_{rs} = \left[inter_{rs}(d_{jk})\right];\ 1 \le r < s \le c,$$
$$inter_{rs}(d_{jk}) =\ (u_{rj} \wedge u_{sk}) \vee (u_{sj} \wedge u_{rk}). \qquad (19)$$

**Step 2.** With respect to each point $x_j$ , weighted means over the associated distances are computed, using intra-inter scores as weights; from which the compactness distance $a_j$ and the separation distances $b_j$ are selected. That is

$$a_j = \min\left\{\frac{\sum_{k=1}^{n}\text{IntraDist}_i(j,k)\ .\ d_{jk}}{\sum_{k=1}^{n}\text{IntraDist}_i(j,k)} \mid 1 \le i \le\ c\right\}, \qquad (20)$$

$$b_j = \min\left\{\frac{\sum_{k=1}^{n}\text{InterDist}_{rs}(j,k)\ .\ d_{jk}}{\sum_{k=1}^{n}\text{InterDist}_{rs}(j,k)} \mid 1 \le r < s \le c\right\}. \qquad (21)$$

**Step 3.** The silhouette of each point is found using (14).

Similar to the original average index, the average intra-inter silhouette, *gSil*, over members of the whole dataset is an assessment of its clustering quality. For each fuzzy cluster, a weighted mean using point membership values as weights, is a measure of its clustering quality.

## Experiments and Discussion

For further evaluation of the validity indices presented above, a few concrete examples are considered as follows:

### Example 1.

Clustering algorithms rely on pairwise distances to form their output and this should be taken into consideration when testing any proposed validity index. Consider the dataset given in Fig. 3. It is tempting to claim that $c = 2$ is the optimal number of clusters, however, this requires a similarity measure better to the task, than the Euclidean distance.

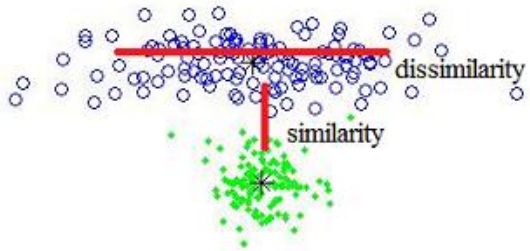


Figure 3. A dataset sampled from two Gaussians. The Euclidean distance is, somehow, inconsistent with the apparent clustering into 2 clusters.

Although HCM, using the Euclidean distance, successfully detects the two clusters, *XB*, *PBM*, *Sil*, *eSil* and *gSil* all score $c = 3$ better than $c = 2$ due to better compactness (Fig. 4). Only *FHV* gives $c = 2$ a better score, since it is based on fitting the data with a mixture of Gaussians. A single bi-Gaussian fits the overlapping clusters in Fig. 4b better than two Gaussians, assuming the crisp probability-membership values produced by HCM.

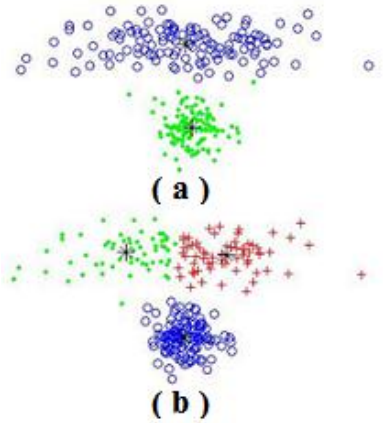


Figure 4. HCM clustering of the dataset from Fig. 3 to (a) 2 clusters and (b) 3 clusters.

### Example 2.

Different FCM partitions, using $c = 2, \ldots, 9$, were obtained on a dataset shown in Fig. 5.

Fig. 6 shows the performance of *PC*, *Sil*, *eSil* and *gSil*. *PBM*, *XB* and *FHV* are shown in Figs. 7, 8 and 9 respectively.

The extended index, *eSil*, scores the partition with $c = 3$ clusters (Fig. 5a) higher than the one with $c = 4$ clusters (Fig. 5b). A different ranking is suggested by the original index, *Sil*, and the generalized index, *gSil*. Both *Sil* and *eSil* incorporate the same silhouettes that are computed from the defuzzified membership matrix; clearly, the disagreement is caused by the weights in *eSil*. The points that occupy the undetected middle cluster (Fig. 5a) are not assigned high memberships to any of the three detected clusters; hence, they have low weights. The index *eSil* just ignores these points that are of insignificant weights and of approximately zero silhouettes. For the same reason, *eSil* always appears above the curve of *Sil*. The generalized index *gSil* can be naturally applied to both crisp and fuzzy partitions. It accounts for changes in the parameter $m$ and does not require any defuzzification of partitions. It scores highest the partition with clusters $c = 5$ (Fig. 5c).

The *PBM* index evaluates the clustering in Fig. 5d as the best. The separation measure, maximum center-to-center distance, does not decrease with $c$ even after reaching the cluster number that is sufficient for a good clustering of the dataset. In addition, the compactness measure decreases monotonically with $c$, provided a reasonable clustering algorithm is used. Therefore, *PBM* has a nondecreasing behavior with $c$ that can be easily exposed using small toy datasets. Moreover, it is not recommended to use any factor that is irrelevant to the assessment of compactness and separation, as part of a validity index.

The *XB* index also fails in its ranking; it scores $c = 3$ better than $c = 5$. The separation measure, minimum center-to-center distance, does not account for the spread of detected clusters: in Fig. 5a, the centers are well separated but there is overlapping among the clusters in the middle region. The separation measure is not thorough in its assessment as opposed to silhouette-based indices that make assessments at point level. Therefore, *XB* is not reliable to detect good cluster numbers, and to compare between any two partitions, in general; it is in disagreement with the silhouette-based indices in its scoring of the partitions with $c = 7$ and $c = 8$.

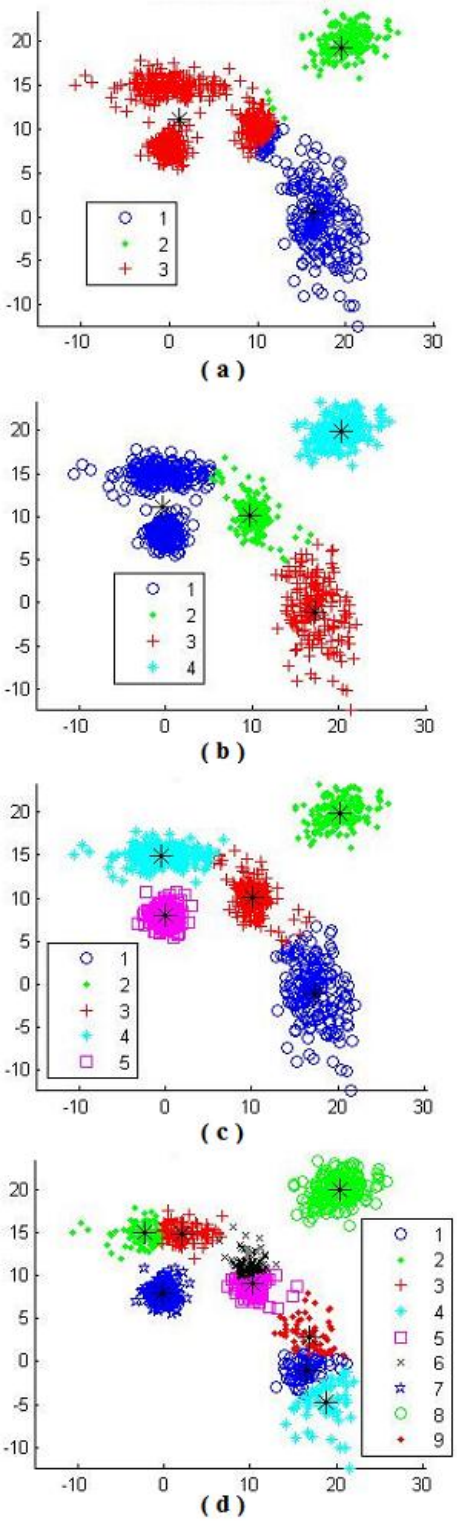


Figure 5. Showing FCM clustering, $m = 2$ and $c = 3,4,5,9$ obtained on a dataset of 1000 points, sampled from 5 bi-Gaussians, 200 each.

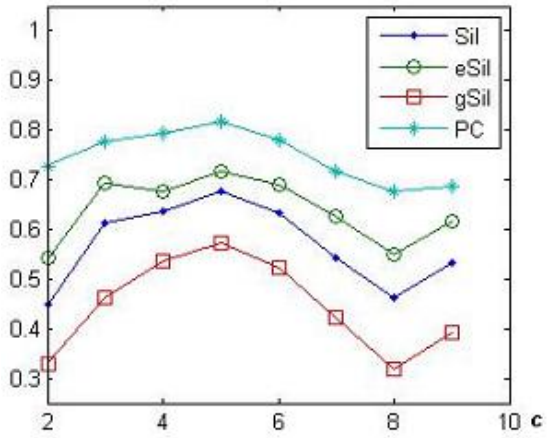


Figure 6. Silhouette-based indices and *PC* vs. $c$, of FCM applied to the dataset in Fig. 5.

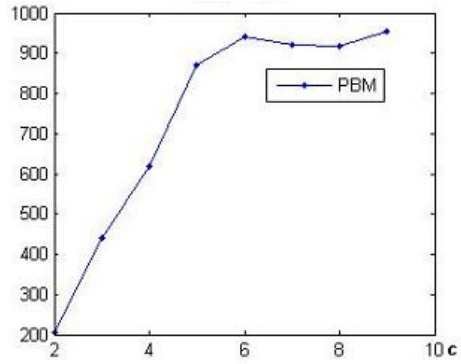


Figure 7. *PBM* vs. $c$, of FCM applied to the dataset in Fig. 5.

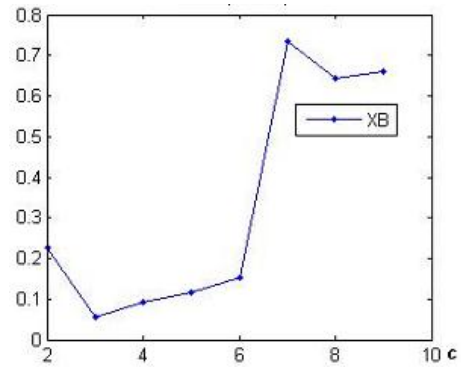


Figure 8. *XB* vs. $c$, of FCM applied to the dataset in Fig. 5.

Distance-based similarities and dissimilarities are inferred from how distance values compare with each other, not from distance magnitudes. Hence, the strength of the underlying structure is determined by distance values relative to each other. Since the quotient in (14) is just a difference in compactness and separation relative to their maximum, an average over the whole dataset measures the strength of a given clustering. Values close to +1, obtained from the average silhouette index, indicate good clustering and a strong underlying structure as well. It is worth noting that, silhouette-based indices are also scale-invariant that is, scaling the dataset by some factor, multiplying by 100

for example, does not affect their values since the structure is still the same. This is not the case for *FHV* and *PBM*. Hence, silhouette-based indices are easier to interpret.

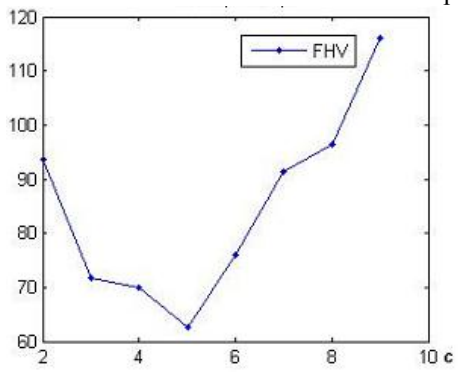


Figure 9. *FHV* vs. *c*, of FCM applied to the dataset in Fig. 5.

## Conclusion

A satisfactory or useful clustering requires careful selection of the features and the measure which, combined together, define the pairwise similarities-dissimilarities. Clustering algorithms, probably of different models, by varying model parameters, as in cluster number, produce partitions, candidate structures. The job of a validity index is to find the candidate that is best supported by the pairwise similarities-dissimilarities, in other words, the clustering that best aligns with the underlying structure. FCM is used mainly to model cluster overlapping in datasets, facilitated by partial cluster memberships assigned to the points, which also results in points in the same cluster taking different membership values. The generalized silhouette index is applicable to both approaches, crisp and fuzzy, of structure modeling to guide the search for the best structure in the dataset.

## References


Balcan, M.-F.; Blum, A.; Vempala, S. 2008. A Discriminative Framework for Clustering via Similarity Functions. Proceedings of the 40th annual ACM symposium on Theory of Computing (STOC).

Berkhin, P. 2002. Survey of Clustering Data Mining Techniques. Technical report, Accrue Software, San Jose, CA.

Bezdek, J. C. 1973. Fuzzy Mathematics in Pattern Classification. Ph.D. Diss., Cornell University.

Bezdek, J. C. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press.

Bezdek, J. C.; Dunn, J. C. 1975. Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions. IEEE Transactions on Computers 24(4): 835-838.

Bezdek, J. C.; Ehrlich, R.; Full, W. 1984. FCM: the Fuzzy *c*-Means Clustering Algorithm. Computers and Geosciences 10: 191-203.

Blienko, M.; Basu, S.; Mooney, R. 2004. Integrating Constraints and Metric Learning in Semisupervised Clustering. Proceedings of the 21st International Conference on Machine Learning. Banff, Canada.

Campello, R.; Hruschka, E. 2006. A Fuzzy Extension of the Silhouette Width Criterion for Cluster Analysis. Fuzzy Sets and Systems, 157: 2858-2875.

Dunn, J. C. 1973. A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Clusters. J. Cybernet 3: 32-57.

Gath, I.; Geva, A. 1989. Unsupervised Optimal Fuzzy Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7): 773-781.

Guyon, I.; von Luxburg, U.; Williamson, R. C. 2009. Clustering: Science or Art?. NIPS Workshop "Clustering: Science or Art".

Kleinberg, J. 2002. An Impossibility Theorem for Clustering. Proceedings of Advances in Neural Information Processing Systems 15: 463-470.

Jain, A.; Murty, M.; Flynn, P. 1999. Data Clustering: A Review. ACM Computing Surveys, 31(3),264-323.

MacQueen, J. 1967. Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematics. Statistics and Probability 2: 281-297. Berkeley, CA.

Pakhira, M.; Bandyopadhyay, S.; Maulik, U. 2004. Validity Index for Crisp and Fuzzy Clusters. Pattern Recognition 37: 481–501.

Rawashdeh, M.; Ralescu, A. Crisp and Fuzzy Cluster Validity: Generalized Intra-Inter Silhouette Index. Forthcoming.

Rousseeuw, P. J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics 20: 53-65. North-Holland.

Trauwaert, E. 1988. On the Meaning of Dunn's Partition Coefficient for Fuzzy Clusters. Fuzzy Sets and Systems 25: 217–242.

Xie, X.; Beni, G. 1991. A Validity Measure for Fuzzy Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 3(8): 841-846.

Xu, R.; Wunsch, D. I. I. 2005. Survey of Clustering Algorithms. IEEE Transactions on Neural Networks 16(3): 645-678.

Wolfe, J. H. 1970. Pattern Clustering by Multivariate Mixture Analysis. Multivariable Behavioral Research, pp. 329-350.

Zhao, Y.; Karypis, G. 2001. Criterion Functions for Document Clustering: Experiments and Analysis. Technical Report, CS Dept., Univ. of Minnesota.