OBJECT BASED VALIDATION ALGORITHM AND ITS APPLICATION TO CONSENSUS CLUSTERING

Rui Fa[1], *Basel Abu-Jamous*[1] *and Asoke K Nandi*[1,2]

[1]Department of Electronic and Computer Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom.

[2] Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland.

Email: {Rui.Fa, Basel.AbuJamous, Asoke.Nandi}@brunel.ac.uk

ABSTRACT

In this paper, we propose a new object based validity index using linear discriminant analysis (OVI-LDA). In OVI-LDA, each object is assigned an index value which is the log ratio of between-group distance to within-group distance. Unlike another object based validity index – Silhouette, OVI-LDA is suitable for both crisp and fuzzy clustering. Furthermore, its object based feature and dual-type capability lead to a consensus clustering by aggregating multiple crisp and fuzzy clustering results. For the demonstration purposes, we study a set of benchmark datasets with a variety of signal-to-noise ratio (SNR) levels and compare the results with other well known indices. The results show that the proposed OVI-LDA possesses not only all capabilities that other validity indices have, but also the capability to guide consensus clustering. The consensus clustering is also validated using a third-party validity index.

Index Terms— Clustering validation, Fisher linear discriminant analysis, Consensus clustering

1. INTRODUCTION

Clustering, also known as unsupervised learning, always needs to be judged by clustering validation because of two following reasons: the first is that due to the unsupervised nature, a metric, which is able to measure the goodness of clustering results, is required for comparison purposes; the second is that due to lack of *a prior* knowledge, clustering validation method should have the ability to indicate the true number of clusters in the datasets. The most popular way to use the clustering validation methods in the clustering analysis is comparing index values over a set of clustering results with varying numbers of clusters to find the best number of clusters [1–3].

There are mainly three classes of clustering validation methods if we only consider the cases that there is no *a prior* knowledge about the memberships and the number of clusters (we exclude the cases that clustering algorithms are assessed by *a prior* knowledge in some simulated datasets). The first class is known as model based algorithms. There are many examples of such criteria, including minimum description length (MDL) [4], Bayesian inference criterion (BIC) [5], minimum message length (MML) [6]. The second class is re-sampling based methods, like figure of merit (FOM) [7], which is similar to cross validation.

The third class is distance ratio based validations, also known as validity indices, which are the focus of this research. These indices measure the within-group distance to the between-group distance ratio and regard the clustering result with the minimum index value as the best, or measure the reciprocal to look for the maximum value. They are different with their different methods to calculate the within-group and between-group distances. In this class, some validity indices are specific to crisp clustering algorithms, namely Calinski-Harabasz (CH) index [8], Davies-Bouldin (DB) index [9], *I* index [10], Silhouettes [11], and some others [12–15]; some indices are specific to fuzzy clustering algorithms, namely partition coefficient (PC), partition entropy (PE) and so on [16, 17]; some others can be used for both fuzzy and crisp algorithms, namely Fukuyama-Sugeno (FS) index, Xie-Beni (XB) index (XB is equivalent to DB when validating crisp clustering) and Kwon's extended XB (KEXB) index [17, 18]. All of these validity indices except Silhouette index consider the clustering as a whole to answer how good the clustering rather than tell if each object is appropriately grouped into a certain cluster. Silhouette index is an object based validity index, but only suitable for crisp clustering. Thus, the research on a general object based validity index for both crisp and fuzzy clustering algorithms remains unaddressed.

In this paper, we propose a new object based validity index using linear discriminant analysis (OVI-LDA), which can be used to validate both crisp and fuzzy clustering algorithms. In OVI-LDA, each object is assigned an index value which is the log ratio of the distance between it and the centroid of the closest neighbouring cluster to the distance between it and the centroid of the cluster where the object is located. Thus the quality of clustering result can be assessed by averaging the

The project (Ref. NIHR-RP-PG-0310-1004-AN) is supported by National Institute for Health Research (NIHR), UK. A. K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

OVI-LDA over all objects and the quality of individual cluster can be assessed by averaging the OVI-LDA over the objects within the cluster. Also one can tell if individual object is appropriate or not in its cluster in terms of its OVI-LDA value. Additionally, a consensus clustering is proposed as an application of OVI-LDA to aggregate multiple clustering results. For the demonstration purposes, we study the quadratic phase shift key (QPSK) datasets with a variety of signal-to-noise ratio (SNR) levels and compare the results with other well known indices. The results show that the proposed OVI-LDA possesses not only all capabilities that other validity indices have, but also the capability to guide consensus clustering. The consensus clustering is also validated using third-party validity indices.

The rest of the paper is organized as follows: Sec. 2 describes the details of the OVI-LDA. Sec. 3 presents a consensus clustering guided by OVI-LDA. Sec. 4 briefly introduces the datasets explored in the paper and presents the numerical results. Finally, discussions and conclusions are given in Sec. 5.

2. PROPOSED VALIDITY INDEX

In this section, we detail the principle of the proposed object based validity index, which employs linear discriminant analysis (OVI-LDA). Suppose that we have many algorithms to partition the dataset $\mathbf{X} = \{\boldsymbol{x}_n | 1 \leq n \leq N\}$, where $\boldsymbol{x}_n \in \mathbb{R}^{M\times 1}$ denotes the n-th object, M is the dimension, and N is the number of objects. Thus, there are a set of clustering results $\boldsymbol{\mathcal{C}} = \{\boldsymbol{C}_{K,c,i} | K_{min} \leq K \leq K_{max},\ 1 \leq c \leq C,\ 1 \leq i \leq N_e\}$ provided by C different clustering algorithms on the same dataset, where K_{min} and K_{max} are the minimum and maximum expected number of clusters, and N_e is the number of experiments of clustering. For some algorithms with deterministic initialization, N_e is set to one. Each clustering result $\boldsymbol{C}_{K,c,i}$ is a $N \times K$ partition matrix, where each entry $u_{n,k}$ represents the membership of the n-th object in the k-th cluster. The properties of the partition matrix is mathematically expressed by

$$\begin{aligned}
&(1) \qquad u_{n,k} \in [0,1], \quad 1 \leq n \leq N,\ 1 \leq k \leq K,\\
&(2) \qquad \textstyle\sum_{k=1}^{K} u_{n,k} = 1, \quad 1 \leq n \leq N,\\
&(3) \qquad 0 \leq \textstyle\sum_{n=1}^{N} u_{n,k} \leq N, \quad 1 \leq k \leq K.
\end{aligned}$$

The crisp clustering partition matrix can be viewed as a special fuzzy clustering partition matrix with $u_{n,k} \in \{0,1\}$. The validation algorithms have to find out the best clustering result out of these clustering results and the best number of clusters K_{best} in terms of validity index value.

LDA, which is also known as Fisher's linear discriminant analysis, has been widely used in the feature reduction and classification [19]. It defines a linear classifier to characterize and separate two or more classes. Here, we introduce the LDA classifier for two classes into clustering validity index and measure the fitness of individual object in its cluster. Since clustering an individual object into a cluster is only challenged by its closest neighbouring cluster, we may simplify the clustering validation problem to a two-class classification problem. The OVI-LDA for each clustering result $\boldsymbol{C}_{K,c,i}$ is an array with the dimension of $(N \times 1)$. Each of its entry v_n represents the log ratio of the distance between n-th object and the centroid of the closest neighbouring cluster (between-group) to the distance between n-th object and the centroid of the cluster where the object is located (within-group), which can be written mathematically by

$$v_n = \log\left\{\frac{\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}'')}{\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}')}\right\}, \tag{1}$$

where $\mathcal{D}(\boldsymbol{a}, \boldsymbol{b})$ is the distance measure between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. $\boldsymbol{\mu}''$ denotes the centroid of the closest neighbouring cluster of object $\boldsymbol{x}_n$ and $\boldsymbol{\mu}'$ is the centroid of the cluster where $\boldsymbol{x}_n$ locates. The centroid of the k-th cluster $\boldsymbol{\mu}_k$ can be given by

$$\boldsymbol{\mu}_k = \frac{1}{l_k}\sum_{n=1}^{N} u_{n,k}^m \boldsymbol{x}_n, \tag{2}$$

where $l_k = \sum_{n=1}^{N} u_{n,k}$ is the weight (or number of members) of k-th cluster, and the fuzzifier m determines the level of cluster fuzziness. Thus, we can assess the whole clustering quality by averaging the OVI-LDA

$$V = \frac{1}{N}\sum_{n=1}^{N} v_n, \tag{3}$$

or we can assess the quality of the k-th cluster

$$V_k = \frac{1}{l_k}\sum_{\boldsymbol{x}_n \in \boldsymbol{C}_k} v_n, \tag{4}$$

In this work, we employ Mahalanobis distance for the distance measure. Thus between-group and within-group distances are given by

$$\begin{aligned}
\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}'') &= \sqrt{(\boldsymbol{x}_n - \boldsymbol{\mu}'')^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}'')},\\
\mathcal{D}(\boldsymbol{x}_n, \boldsymbol{\mu}') &= \sqrt{(\boldsymbol{x}_n - \boldsymbol{\mu}')^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}')},
\end{aligned} \tag{5}$$

where $(\cdot)^T$ is the transpose operator and $\boldsymbol{\Sigma}$ is the combined covariance matrix of n-th object locating cluster and its closest neighbouring cluster, which is written mathematically by

$$\boldsymbol{\Sigma} = \frac{(l' - 1)\boldsymbol{\Sigma}'' + (l'' - 1)\boldsymbol{\Sigma}'}{l' + l'' - 2} \tag{6}$$

where $\boldsymbol{\Sigma}''$ is the covariance matrix for the closest neighbouring cluster, $\boldsymbol{\Sigma}'$ is the covariance matrix for the local cluster, and l'' and l' are their weights respectively. The covariance matrix for the k-th cluster is obtained by

$$\boldsymbol{\Sigma}_k = \frac{1}{l_k - 1}\sum_{n=1}^{N} u_{n,k}^m (\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T. \tag{7}$$

Table 1. Comparison between OVI-LDA and many existing clustering validity indices.

Algorithms	crisp	fuzzy	Quality Assessment for			Consensus
			dataset	cluster	object	
CH	YES	NO	YES	NO	NO	NO
DB	YES	NO	YES	NO	NO	NO
I index	YES	NO	YES	NO	NO	NO
GI	YES	NO	YES	NO	NO	NO
VI	YES	NO	YES	NO	NO	NO
PVI	YES	NO	YES	NO	NO	NO
KVIs	YES	NO	YES	NO	NO	NO
Silhouette	**YES**	**NO**	**YES**	**YES**	**YES**	**N/Y***
PC	NO	YES	YES	NO	NO	NO
PE	NO	YES	YES	NO	NO	NO
XB	YES	YES	YES	NO	NO	NO
FS	YES	YES	YES	NO	NO	NO
OVI-LDA	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**

* Silhouette: **NO** for consensus cross crisp and fuzzy clustering, but **YES** for only consensus of crisp clustering.

Note that from estimation theory point of view, if there are not enough samples in one cluster, the estimation of its parameters would be unreliable. We set a threshold T_l for the number of members in clusters to regulate OVI-LDA: if $l_k < T_l$, then the k-th cluster will be discounted and the OVI-LDA values of the members in that cluster will be set not-a-number (NAN).

In Table 1, we list many existing validity indices in the literature and their capabilities, namely CH, DB, *I* index, GI, VI, PVI, kernel validity indices (KVIs), Silhouette, PC, PE, XB, FS and the proposed OVI-LDA. Each row of the Table represents one validity index and each column represents one capability type. Each cell of the Table is filled with YES or NO to indicate if the corresponding validity index has the corresponding capability or not. We consider six different capabilities, including the capability to assess crisp clustering (crisp, for short), the capability to assess fuzzy clustering (fuzzy), the capability to assess quality at the level of a dataset, the capability to assess quality at the level of an individual cluster, the capability to assess quality at the level of an individual object and the capability to guide consensus clustering. As shown in the Table, CH, DB, *I* index, GI, VI, PVI, KVIs and Silhouette are only suitable for crisp clustering. Except Silhouette, other seven crisp validity indices only have capability to assess the clustering treating the dataset as a whole and provide the estimated best number of clusters. Silhouette has capabilities to assess individual cluster and object because of its object-based nature. PC and PE are fuzzy-only validity indices; XB and FS can work for both fuzzy and crisp clustering. These four indices also only assess the clustering, treating the dataset as a whole. It is worth noting that the proposed OVI-LDA possesses all capabilities. We highlight two object based validity indices, i.e., Silhouette and OVI-LDA. The difference between these two indices is that the proposed OVI-LDA can deal with both fuzzy and crisp clustering. Due to the object-based nature and dual-type capability, the proposed OVI-LDA can guide a consensus clustering to aggregate multiple clustering results, which will be explored in the next section. Although Silhouette has the potential for combining crisp partitions, this application has not been explored in the literature.

Table 2. The pseudo-code for the routine of converting the consensus matrix to the partition matrix $\boldsymbol{P}$.

Initializing $\boldsymbol{P} = \emptyset$, $K = 0$;
for $n = 1 : N$ **do**
  if $\{\boldsymbol{P}$ is $\emptyset$ OR $\boldsymbol{P}(n, :)$ is zero vector$\}$ AND $\boldsymbol{E}^f(n, :)$ contains more than one entry greater than 0 **then**
    $K = K + 1$;
    $\boldsymbol{P}(:, K) = \boldsymbol{E}^f(n, :)^T$
  end if
end for
Output K and $\boldsymbol{P}$.

3. CONSENSUS CLUSTERING

In this section, we present an application of the proposed OVI-LDA to the consensus clustering. To reach a consensus clustering, there are several steps as follows:

STEP 1: Suppose that we have C clustering algorithms, each clustering algorithm will provide a set clustering results with variable number of clusters, say K_{min} to K_{max}. We calculate OVI-LDA for each object and average OVI-LDA of the dataset for each clustering result. For some randomly initialized algorithms, to make the clustering results representative to the clustering algorithm, we have to run N_e experiments and choose the best clustering results for every choice of K in terms of averaged OVI-LDA index value.

STEP 2: For each clustering algorithm, we compare the averaged OVI-LDA index values over K_{min} to K_{max} to find the best estimated K. Thus, we have C clustering results and they do not necessarily have same number of clusters.

STEP 3: We set a threshold, T_v to filter the objects whose OVI-LDA validity indices in any results below T_v. Thus we may obtain a list of objects which are certainly appropriate with the positions where they are in all clustering results.

STEP 4: We convert each clustering result for the same list of objects to connectivity matrix. Connectivity matrix was defined as a $(N \times N)$ sparse matrix $\boldsymbol{E}^c$ for the c-th clustering result and its entries are

$$E^c_{i,j} = \begin{cases} 1, & \text{if objects } i \text{ and } j \text{ in the same cluster} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

where $i, j = 1, ..., N$.

STEP 5: The consensus matrix can be obtained by

$$\boldsymbol{E} = \sum_{c=1}^{C} \boldsymbol{E}^c. \tag{9}$$

Then, we further truncate $\boldsymbol{E}$ to the final consensus matrix $\boldsymbol{E}^f$ by

$$E^f_{i,j} = \begin{cases} E_{i,j} & \text{if } E_{i,j} \geq \lceil C/2 \rceil \\ 0 & \text{if } E_{i,j} < \lceil C/2 \rceil, \end{cases} \tag{10}$$

Table 3. Results comparison shows the estimated best number of clusters by each validity index for each QPSK dataset and each clustering algorithm.

Data	Algorithms	CH	PE	XB	FS	Sil	proposed
2 dB	GG	-	2	6	4	-	4
	FCM	-	2	20	4	-	4
	K-Means	3	-	20	3	3	3
	KMeans-KA	3	-	19	3	3	3
	HC-Ward	4	-	20	4	3	4
	Consensus	-	-	-	-	-	**4**
5 dB	GG	-	2	4	4	-	4
	FCM	-	2	4	4	-	4
	K-Means	4	-	20	4	4	4
	KMeans-KA	4	-	4	4	4	4
	HC-Ward	4	-	4	4	4	4
	Consensus	-	-	-	-	-	4
8 dB	GG	-	2	4	4	-	4
	FCM	-	4	4	4	-	4
	K-Means	4	-	4	4	4	4
	KMeans-KA	4	-	4	4	4	4
	HC-Ward	4	-	4	4	4	4
	Consensus	-	-	-	-	-	4

where $\lceil \cdot \rceil$ is the ceil operator. As soon as we get the final consensus matrix $\boldsymbol{E}^f$, we can easily convert it back to the partition matrix $\boldsymbol{P}$. The pseudo code of the converting routine is shown in Table 2, where $\boldsymbol{P}(n,:)$ is the operation of getting the n-th row from the matrix $\boldsymbol{P}$ and $\boldsymbol{P}(:,K)$ is the operation of getting the K-th column. For some applications, for example gene discovery, the produced partition matrix can be the clustering output, since only small portion of objects forming very tight clusters should be preferably investigated.

STEP 6: For the applications which need a complete partition, we can employ the produced partition matrix $\boldsymbol{P}$ as a guideline, which defines the core for each cluster and helps the estimation of the parameters including the centroids and the covariance matrices. Finally, we can obtain the final partition matrix $\boldsymbol{P}_f$ by employing the estimated parameters,.

4. NUMERICAL RESULTS

Here, we present the performance comparison between the proposed OVI-LDA and some existing validity indices, as well as the performance of the proposed consensus clustering. In this wok, we study one benchmark case using quadratic phase shift key (QPSK) datasets with the signal-to-noise ratio (SNR) from 1 dB to 10 dB (ten datasets). We employ five clustering algorithms, including two fuzzy algorithms, namely fuzzy c-means (FCM) [20] and Gath-Geva (GG) [21], and three crisp algorithms, namely normal KMeans, hierarchical clustering with ward linkage (HC-Ward), KMeans with Kaufman approach initialization [22] (KMeans-KA). Since GG, FCM and KMeans are randomly initialized, we set the number of experiments to $N_e = 1000$ for one dataset, while for deterministic algorithms like HC-Ward and KMeans-KA, only one experiment is enough. Each dataset in each experiment is clustered using each algorithm with the number of clusters from K_{min} to K_{max}. Thus, one can find the best clustering for each K among all N_e experiments in terms of different validity indices, and then find the best clustering (the best K) for the dataset cross K_{min} to K_{max}.

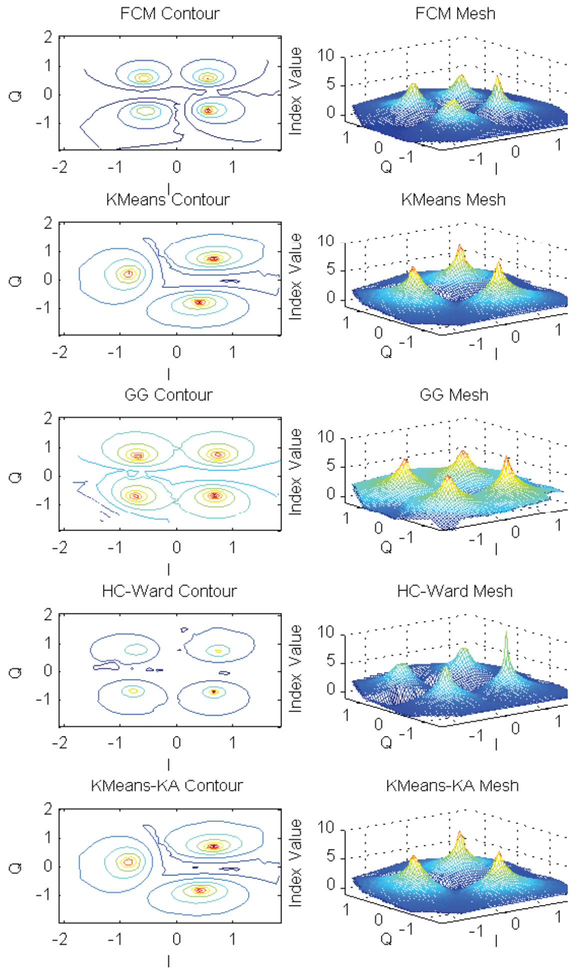

Fig. 1. Demonstration of OVI-LDA for each clustering algorithm, QPSK 2dB dataset.

Due to the limitation of the space, we choose three QPSK datasets with low SNR (2dB), moderate SNR (5 dB) and high SNR (8 dB) to display the result comparison. In Table 3, the estimated best numbers of clusters by each validity index for each dataset and each clustering algorithm are shown. It shows that for the moderate and high SNR datasets, most of validity indices work properly. But for the low SNR dataset, the majority of the estimated best Ks are wrong. In this case, it is difficult to judge which clustering is better and how many clusters in the dataset. Thus, the consensus clustering appears to be extremely useful. Note that the OVI-LDA indicates that KMeans and KMeans-KA have the best K equal to three and other three have $K = 4$. The contour and the mesh plots of OVI-LDA of the best clustering results by all clustering algorithms for QPSK 2dB dataset are shown in in Fig. 1. These clustering results are fed into consensus clustering and final

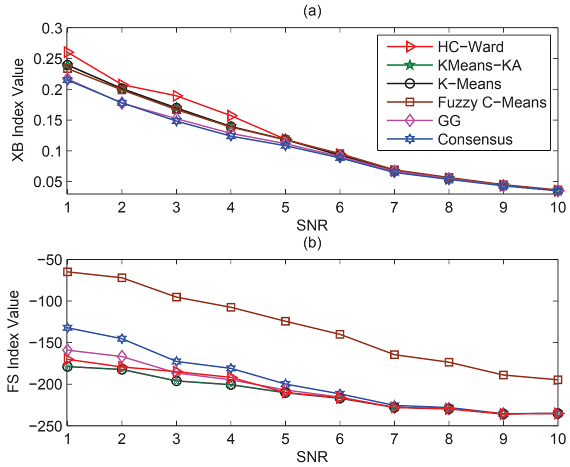

Fig. 2. Comparison of the validity index values over the SNR level between consensus clustering and other clustering algorithms using (a) XB index, (b) FS index.

consensus result is four clusters, which is highlighted in the table. We also validate the consensus clustering result using both XB and FS algorithms, which are third party algorithms. Both of them consider the minimum values as the best. The results are shown in Fig. 2. Interestingly, two validity indices do not show consistency: the XB index values of the proposed consensus clustering is the lowest and slightly lower than GG; while the FS index values of consensus clustering is only lower than FCM but higher than others. However, we also notice that the FS index values of the crisp algorithms are always lower than fuzzy ones. It turns out that FS validity index is sensitive to the fuzziness and issue low value to crisp or near-crisp algorithms. Thus, in this case, XB endorses that the proposed consensus clustering has the best performance.

5. CONCLUSIONS

We proposed a new object based validity index using linear discriminant analysis (OVI-LDA). Most validity indices in the literature, namely CH, DB, I index, GI, VI, PVI, KVIs, PC, PE, FS, XB, and KEXB, are not object-based indices, except Silhouette. Only Silhouette is capable of indicating how fit an individual object is in its cluster. However, Silhouette is only suitable for crisp clustering results. The proposed OVI-LDA possesses not only all capacities that other validity indices have, but also the capability to guide consensus clustering. The results showed that most validity indices worked properly at moderate or high SNR levels but there were great discrepancies among all validity indices at low SNR levels. A consensus clustering guided by OVI-LDA appeared to be useful to aggregate multiple clustering results. It was validated to be the best clustering by XB, which is a third-party algorithm.

6. REFERENCES

[1] D. X. Jiang, C. Tang, and A. D. Zhang, “Cluster analysis for gene expression data: A survey,” *IEEE Trans. Know. and Data Eng.*, vol. 16, no. 11, pp. 1370–1386, 2004.

[2] R. Xu and D. II Wunsch, “Survey of clustering algorithms,” *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645-678, 2005.

[3] R. Xu and D. II Wunsch, “Clustering Algorithms in Biomedical Research: A Review,” *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 120–154, 2010.

[4] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol.14, no. 5, pp. 465 - 471, 1978.

[5] C. Fraley and A. Raftery, “How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis,” *Technical Report 329*, Dept. Statistics, Univ. Washington, Seattle, WA, 1998.

[6] A. D. Lanterman “Schwarz, Wallace, and Rissanen: Intertwining Themes in Theories of Model Order Estimation,” *International Statistical Review*, v. 69, no. 2, pp. 185-212, 2001.

[7] K. Y. Yeung, D. R. Haynor and W. L. Ruzzo, “Validating clustering for gene expression data,” *Bioinformatics*, v. 17, no. 4, pp. 309-318, 2001.

[8] T. Calinski and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics - Theory and Methods*, vol. 3, no. 1, pp. 1-27, 1974.

[9] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224-227, 1979.

[10] U. Maulik and S. Bandyopadhyay, “Performance evaluation of some clustering algorithms and validity indices,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650-1654, 2002.

[11] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, 20, pp. 53 - 65, 1987.

[12] B. S. Y. Lam and H. Yan, “Assessment of microarray data clustering results based on a new geometrical index for cluster validity,” *Soft Computing*, vol. 11, no. 4, pp. 341–348, 2007.

[13] S. A. Salem and A. K. Nandi, “Development of assessment criteria for clustering algorithms,” *Pattern Analysis and Applications*, vol. 12, no. 1, pp. 79-98, 2009.

[14] R. Fa and A. K. Nandi, “Parametric validity index of clustering for microarray gene expression data,” in *IEEE Int. Workshop Machine Learning for Sig. Process. (MLSP)*, 2011.

[15] R. Fa, A. K. Nandi and B. Abu-Jamous, “Development and evaluation of kernel-based clustering validity indices,” in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* Aug. 2012

[16] J. C. Bezdek, “Cluster validity with fuzzy sets,” in *J. Cybernet*, v. 3, pp. 5873, 1974.

[17] W. Wang and Y. Zhang, “On fuzzy cluster validity indices,” in *Fuzzy Sets and Systems*, v. 158, pp. 2095-2117, 2007.

[18] X. Xie and G. Beni, “A validity measure for fuzzy clustering” *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol: 13 no: 8, p. 841-847, 1991.

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th Edition, Academic Press, 2008.

[20] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[21] I. Gath and A. B. Geva, “Unsupervised optimal fuzzy clustering” *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. I, no. 7, 1989.

[22] P. J. Rousseeuw and L. Kaufman, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, 1990.