Performance Evaluation of Some Clustering Algorithms and Validity Indices

Ujjwal Maulik, *Member*, *IEEE*, and
Sanghamitra Bandyopadhyay, *Member*, *IEEE*

Abstract—In this article, we evaluate the performance of three clustering algorithms, hard K-Means, single linkage, and a simulated annealing (SA) based technique, in conjunction with four cluster validity indices, namely Davies-Bouldin index, Dunn's index, Calinski-Harabasz index, and a recently developed index $\mathcal{I}$. Based on a relation between the index $\mathcal{I}$ and the Dunn's index, a lower bound of the value of the former is theoretically estimated in order to get unique hard K-partition when the data set has distinct substructures. The effectiveness of the different validity indices and clustering methods in automatically evolving the appropriate number of clusters is demonstrated experimentally for both artificial and real-life data sets with the number of clusters varying from two to ten. Once the appropriate number of clusters is determined, the SA-based clustering technique is used for proper partitioning of the data into the said number of clusters.

Index Terms—Unsupervised classification, Euclidean distance, K-Means algorithm, single linkage algorithm, validity index, simulated annealing.

1 Introduction

The purpose of any clustering technique [1], [2], [3], [4], [5] is to evolve a $K \times n$ partition matrix $U(X)$ of a data set X ($X = \{x_1, x_2, \ldots, x_n\}$) in $\mathcal{R}^N$, representing its partitioning into a number, say K, of clusters ($C_1, C_2, \ldots, C_K$). The partition matrix $U(X)$ may be represented as $U = [u_{kj}], k = 1, \ldots, K,$ and $j = 1, \ldots, n$, where u_{kj} is the membership of pattern x_j to clusters C_k. In crisp partitioning of the data, the following condition holds: $u_{kj} = 1$ if $x_j \in C_k$; otherwise, $u_{kj} = 0$. Clustering techniques broadly fall into two classes, partitional and hierarchical. K-Means and single linkage [1], [2] are widely used techniques used in the domains of partitional and hierarchical clustering, respectively.

The two fundamental questions that need to be addressed in any typical clustering system are: 1) How many clusters are actually present in the data and 2) how real or good is the clustering itself. That is, whatever the clustering method may be, one has to determine the number of clusters and also the goodness or validity of the clusters formed [6]. The measure of validity of the clusters should be such that it will be able to impose an ordering of the clusters in terms of its goodness. In other words, if $U_1, U_2, \ldots, U_m$ is m partitions of X and the corresponding values of a validity measure are $V_1, V_2, \ldots, V_m$, then $V_{k1} >= V_{k2} >= \ldots >= V_{km}$ will indicate that $U_{k1} \uparrow U_{k2} \uparrow \ldots \uparrow U_{km}$, for some permutation $k1, k2, \ldots, km$ of $\{1, 2, \ldots, m\}$. Here, $"U_i \uparrow U_j"$ indicates that partition U_i is a better clustering than U_j.

Milligan and Cooper [6] have provided a comparison of several validity indices for data sets containing distinct non-overlapping clusters while using only hierarchical clustering algorithms. Meilă and Heckerman provide a comparison of some clustering methods and initialization strategies in [7]. Some more clustering algorithms may be found in [8], [9]. In this paper, we aim to evaluate the performance of four validity indices, namely, the Davies-Bouldin index [10], Dunn's index [11], Calinski-Harabasz index [12], and a recently developed index $\mathcal{I}$, in conjunction with three clustering algorithms viz. the well-known K-means and single linkage algorithms [1], [2], as well as a recently developed simulated annealing (SA) [13], [14] based clustering scheme. The number of clusters is varied from K_{min} to K_{max} for K-means and the simulated annealing-based clustering algorithms, while, for single linkage algorithm (which incorporates automatic variation of number of clusters), the partitions in this range are considered. As a result, all the three clustering algorithms will yield ($K_{max} - K_{min} + 1$) partitions, $U^*_{K_{min}}$, $U^*_{K_{min}+1}$, $\ldots$, $U^*_{K_{max}}$, with the corresponding validity index values computed as $V_{K_{min}}$, $V_{K_{min}+1}$, $\ldots$ $V_{K_{max}}$. Let $K^* = \text{argmax}_{i=K_{min},\ldots,K_{max}}[V_i]$. Therefore, according to index V, K^* is the correct number of clusters present in the data. The corresponding $U^*_{K^*}$ may be obtained by using a suitable clustering technique with the number of clusters set to K^*. The tuple $< U^*_{K^*}, K^* >$ is presented as the solution to the clustering problem.

2 Clustering Algorithms

The three clustering algorithms considered in this article are the well-known K-Means and single linkage algorithms and a recently developed simulated annealing (SA) based clustering technique that uses probabilistic redistribution of points.

The K-Means algorithm [1], [2] is an iterative scheme that evolves K crisp, compact, and hyperspheroidal clusters in the data such that a measure

$$J = \sum_{j=1}^{n} \sum_{k=1}^{K} u_{kj} ||x_j - z_k||^2 \tag{1}$$

is minimized. Here, the K cluster centers are initialized to K randomly chosen points from the data, which is then partitioned based on the minimum squared distance criterion. The cluster centers are subsequently updated to the mean of the points belonging to them. This process of partitioning followed by updating is repeated until either the cluster centers do not change or there is no significant change in the J values of two consecutive iterations.

The single linkage clustering scheme is a noniterative method based on a local connectivity criterion,and is usually regarded as a graph theoretical model [2]. Instead of an object data set X, single linkages process sets of n^2 numerical relationships, say $\{r_{jk}\}$, between pairs of objects represented by the data. The number r_{jk} represents the extent to which object j and k are related in the sense of some binary relation ρ. It starts by considering each point in a cluster of its own. The single linkage algorithm computes the distance between two clusters S and T as

$$\delta_{SL}(S, T) = \underbrace{\min}_{x \in S, y \in T} \{d(x, y)\}.$$

Based on these distances, it merges the two closest clusters, replacing them by the merged cluster. The distance of the remaining clusters from the merged one is recomputed as above. The process continues until the a single cluster, comprising all the points, is formed.

The third clustering algorithm considered in this article for the purpose of comparison is a simulated annealing (SA) based scheme with probabilistic redistribution of the data points [13], [14]. The SA algorithm starts from a random initial configuration at high temperature T_{max}. A configuration, encoded as a set of cluster centers, represents a partitioning of the data based on the minimum squared distance criterion. The energy function $\mathcal{E}$ associated with configuration $\mathcal{C}$ is computed as $\mathcal{E} = \sum_{i=1}^{K} \sum_{\forall x_j \in C_i} ||x_j - z_i||^2$, where z_i is the center of cluster C_i. A new configuration $\mathcal{C}'$ with energy $\mathcal{E}'$ is generated from the old one, $\mathcal{C}$, by redistributing all the elements $x_i,\ i = 1, 2, \ldots, n_j$ in cluster C_j to cluster $C_k,\ k = 1, 2, \ldots, K,\ j \neq k$ with probability

$$\exp^{\left(\frac{-[D_{ik} - D_{ij}]+}{T_t}\right)},$$

- *U. Maulik is with the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019.*
E-mail: maulik@cse.uta.edu.
- *S. Bandyopadhyay is with the Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta, -700 108, India.*
E-mail: sanghami@isical.ac.in.

Manuscript received 27 June 2001; revised 8 Jan. 2002; accepted 9 May 2002.
Recommended for acceptance by V. Govindaraju.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114440.

0162-8828/02/$17.00 © 2002 IEEE

where $[x]+ = \max(x, 0)$ and $D_{ik} = ||x_i - z_k||$. T is the temperature schedule, which is a sequence of strictly positive numbers such that $T_1 \geq T_2 \geq \ldots T_t = 0 (\lim t \rightarrow \infty)$. The suffix t of T indicates the number of generations through the annealing process. The new configuration is accepted/rejected according to a probability

$$\frac{1}{1 + \exp\left(\frac{-(\mathcal{E} - \mathcal{E}')}{T}\right)},$$

which is a function of the current temperature and energy difference between the two configurations. The temperature is gradually decreased toward a minimum value T_{min} while the system settles down to a stable low energy state.

Note that, while the single linkage algorithm precomputes the distance between all pairs of points and subsequently uses them at each level of the hierarchy, both the K-Means and the SA-based algorithms compute the distance between the points to all the cluster centers in each iteration. Therefore, if n is the total number of data points, N is the dimensionality of the data, and K is the number of clusters being considered, then the complexity of the distance computation phase in single linkage will be $O(n^2 N)$, i.e., it is linearly dependent on N. Again, both the K-Means and the SA-based method will have complexity $O(KnN)$ in each iteration, i.e., linearly dependent on N as well.

3 Cluster Validity Indices

In this section, the four cluster validity indices that have been used in this article to evaluate the partitioning obtained by the above three techniques for different values of K are described in detail.

Davies-Bouldin (DB) Index: This index [10] is a function of the ratio of the sum of *within-cluster scatter* to *between-cluster separation*. The scatter within the ith cluster, S_i, is computed as $S_i = \frac{1}{|C_i|} \sum_{x \in C_i} \{||x - z_i||\}$ and the distance between cluster C_i and C_j, denoted by d_{ij}, is defined as $d_{ij} = ||z_i - z_j||$. Here, z_i represents the ith cluster center. The Davies-Bouldin (DB) index is then defined as

$$DB = \frac{1}{K} \sum_{i=1}^{K} R_{i,qt}, \tag{2}$$

where $R_{i,qt} = \max_{j, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right\}$. The objective is to minimize the DB index for achieving proper clustering.

Dunn's Index: Let S and T be two nonempty subsets of $\mathcal{R}^N$. Then, the diameter $\triangle$ of S is defined as $\triangle(S) = \underbrace{\max}_{x,y \in S} \{d(x, y)\}$ and set distance δ between S and T is defined as $\delta(S, T) = \underbrace{\min}_{x \in S, y \in T} \{d(x, y)\}$. Here, $d(x, y)$ indicates the distance between points x and y. For any partition, Dunn defined the following index [11]:

$$\nu_D = \underbrace{\min}_{1 \leq i \leq K} \left\{ \underbrace{\min}_{1 \leq j \leq K, j \neq i} \left\{ \frac{\delta(C_i, C_j)}{\underbrace{\max}_{1 \leq k \leq K} \{\triangle(C_k)\}} \right\} \right\}. \tag{3}$$

Larger values of ν_D correspond to good clusters, and the number of clusters that maximizes ν_D is taken as the optimal number of clusters.

Calinski Harabasz (CH) Index: This index [12] for n data points and K clusters is computed as

$$\frac{[trace\ B/(K-1)]}{[trace\ W/(n-K)]}.$$

Here, B and W are the between and within cluster scatter matrices. The maximum hierarchy level is used to indicate the correct number of partitions in the data. The trace of the between cluster scatter matrix B can be written as

$$trace\ B = \sum_{k=1}^{K} n_k ||z_k - z||^2,$$

where n_k is the number of points in cluster k and z is the centroid of the entire data set. The trace of the within cluster scatter matrix W can be written as

$$traceW = \sum_{k=1}^{K} \sum_{i=1}^{n_k} ||x_i - z_k||^2.$$

Therefore, the CH index can be written as

$$CH = \left[\frac{\sum_{k=1}^{K} n_k ||z_k - z||^2}{K - 1} \right] / \left[\frac{\sum_{k=1}^{K} \sum_{i=1}^{n_k} ||x_i - z_k||^2}{n - K} \right]. \tag{4}$$

Index $\mathcal{I}$: The index $\mathcal{I}$ is defined as follows:

$$\mathcal{I}(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^p, \tag{5}$$

where K is the number of clusters. Here,

$$E_K = \sum_{k=1}^{K} \sum_{j=1}^{n} u_{kj} ||x_j - z_k||,$$

and

$$D_K = \max_{i,j=1}^{K} ||z_i - z_j||.$$

n is the total number of points in the data set, $U(X) = [u_{kj}]_{K \times n}$ is a partition matrix for the data, and z_k is the center of the kth cluster. The value of K for which $\mathcal{I}(\mathcal{K})$ is maximized is considered to be the correct number of clusters.

As can be seen from (5), the index $\mathcal{I}$ is a composition of three factors, namely, $\frac{1}{K}$, $\frac{E_1}{E_K}$, and D_K. The first factor will try to reduce index $\mathcal{I}$ as K is increased. The second factor consists of the ratio of E_1, which is constant for a given data set, and E_K, which decreases with increase in K. Hence, because of this term, index $\mathcal{I}$ increases as E_K decreases. This, in turn, indicates that formation of more numbers of clusters, which are compact in nature, would be encouraged. Finally, the third factor, D_K (which measures the maximum separation between two clusters over all possible pairs of clusters), will increase with the value of K. However, note that this value is upper bounded by the maximum separation between two points in the data set. Thus, the three factors are found to compete with and balance each other critically. The power p is used to control the contrast between the different cluster configurations. In this article, we have taken $p = 2$.

Xie and Beni defined an index [15] that is a ratio of the compactness π of the fuzzy K-partition of a data set to its separation s. Mathematically, the Xie Beni (XB) index may be formulated as:

$$XB = \frac{\sum_{k=1}^{K} \sum_{j=1}^{n} u_{kj}^2 ||x_j - z_k||^2}{n \min_{i,j} ||z_i - z_j||^2}. \tag{6}$$

Here, u_{kj} is the membership of the jth point to the kth cluster and the XB index is independent of the algorithm used to obtain it. The XB index has been mathematically justified in [15] via its relationship to a well-defined hard clustering validity function, the Dunn's index (ν_D). In this section, we provide a theoretical justification of index $\mathcal{I}$ by establishing its relationship to the Dunn's index via XB index with the underlying assumption that $K \leq \sqrt{n}$, which is a practically valid assumption. From (5) and (6), we have

$$\begin{aligned} XB \times \mathcal{I} = {} & \frac{1}{n \times K^2} \times E_1^2 \times \frac{(\max_{i,j=1}^{K} ||z_i - z_j||)^2}{\min_{i,j=1}^{K} (||z_i - z_j||)^2} \\ & \times \frac{\sum_{k=1}^{K} \sum_{j=1}^{n} u_{kj}^2 ||x_j - z_k||^2}{(\sum_{k=1}^{K} \sum_{j=1}^{n} u_{kj} ||x_j - z_k||)^2}. \end{aligned} \tag{7}$$

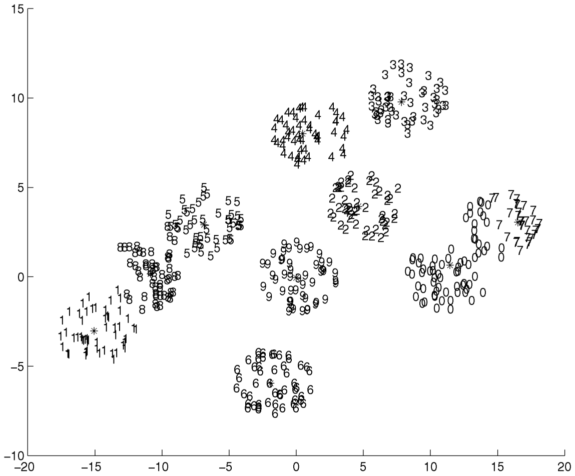

Fig. 1. AD_10_2 partitioned into 10 clusters by the SA-based clustering technique.

Note that

$$\frac{E_1^2}{\left(\sum_{k=1}^{K}\sum_{j=1}^{n} u_{kj}||x_j - z_k||\right)^2} \ge 1,$$

and

$$\frac{\left(\max_{i,j=1}^{K}||z_i - z_j||\right)^2}{\min_{i,j=1}^{K}\left(||z_i - z_j||\right)^2} \ge 1.$$

Therefore,

$$XB \times \mathcal{I} \ge \frac{1}{n \times K^2} \times \sum_{k=1}^{K}\sum_{j=1}^{n} u_{kj}^2||x_j - z_k||^2. \tag{8}$$

Since, in most real-life situations, we have $K \le \sqrt{n}$, so,

$$XB \times \mathcal{I} \ge \frac{\sum_{k=1}^{K}\sum_{j=1}^{n} u_{kj}^2||x_j - z_k||^2}{n^2}. \tag{9}$$

Let us assume that cluster k has n_k points, and the distances of these n_k points from the cluster center z_k are $d_{k1}, d_{k2}, \ldots, d_{kn_k}$. Note that $\sum_{k=1}^{K} n_k = n$. Let us define τ_k as $\frac{\sum_{i=1}^{n_k} d_{ki}^2}{n_k}$ and τ_{min} as $\min_k(\tau_k)$. Here, τ_{min} represents the minimum of the mean squared distances of the points from their respective cluster centers (or minimum of the mean squared error of the points in the respective clusters). In (9), $\sum_{k=1}^{K}\sum_{j=1}^{n} u_{kj}^2||x_j - z_k||^2$ can be written as $\sum_{k=1}^{K}\sum_{i=1}^{n_k} d_{ki}^2$. Since $\sum_{k=1}^{K}\sum_{i=1}^{n_k} d_{ki}^2 = \sum_{k=1}^{K} n_k\tau_k \ge n\tau_{min}$, so, $XB \times \mathcal{I} \ge \frac{\tau_{min}}{n}$. It is proven in [15] that $XB \le \frac{1}{\nu_D^2}$. Therefore,

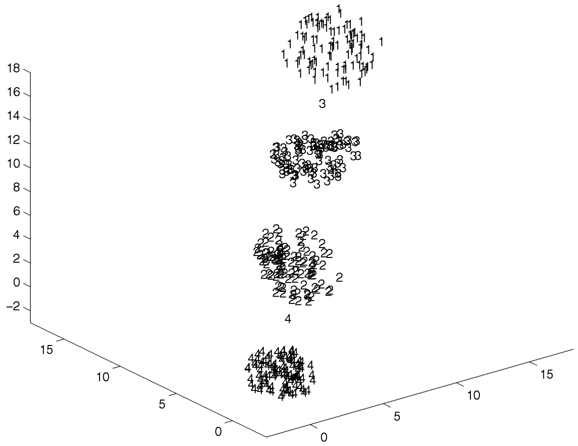

Fig. 2. AD_4_3N partitioned into four clusters by the SA-based clustering technique.

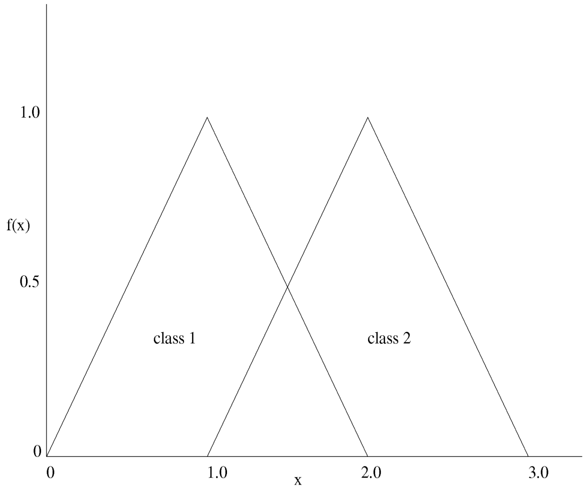

Fig. 3. Triangular distribution along the X axis for AD_2_10.

$$\mathcal{I} \times \frac{1}{\nu_D^2} \ge \frac{\tau_{min}}{n}. \tag{10}$$

Evidently, index $\mathcal{I}$ becomes arbitrarily large as ν_D grows without bound. It has been proven in [16] that, if $\nu_D > 1$, the hard K-partition is unique. Therefore, if the data sets have a distinct substructure and the clustering algorithm found it, then the corresponding $\mathcal{I} \ge \frac{\tau_{min}}{n}$.

4 Experimental Results

4.1 Data Sets and Implementation Parameters

The three artificial data sets that have been used in this article are *AD_10_2*, *AD_4_3N*, and *AD_2_10*. *AD_10_2* is a two-dimensional overlapping data set with 10 clusters, whereas *AD_4_3N* is a three dimensional with four clusters. Fig. 1 and Fig. 2 show the data sets *AD_10_2* and *AD_4_3N*, respectively (assuming the numbers to be replaced by data points). *AD_2_10* is an overlapping 10-dimensional data set generated using a triangular distribution of the form shown in Fig. 3 for two classes, one and two, both of which have equal a priori probabilities. It has 1,000 data points. The range for class one is $[0,2] \times [0,2] \times [0,2] \ldots 10$ times and that for class two is $[1,3] \times [0,2] \times [0,2] \ldots 9$ times.

Two real-life data sets considered are *Crude_Oil* and *Cancer*. *Crude_Oil* is an overlapping data [17] having 56 data points, five features, and three classes. The nine-dimensional Wisconsin breast cancer data (*Cancer*) (http://www.ics.uci.edu/ mlearn/MLRepository.html) is used for the purpose of demonstrating the effectiveness of the classifier in classifying high-dimensional patterns. It has 683 samples belonging to two classes: *Benign* (class 1) and *Malignant* (class 2). Table 1 presents the number of points, dimensions, and the number of clusters in each data.

In this article, the K-Means algorithm was executed for a maximum 100 iterations. The simulated annealing algorithm was implemented with the following parameters: $T_{max} = 100$, $T_{min} = 0.001$, $\alpha = 0.05$, and $N_T = 100$. Both the K-Means and the SA-based clustering algorithms were initialized with the same set of cluster centers for each K in order to make the comparison fair. Note that the single linkage clustering algorithm assumes that, initially, each point forms a cluster of its own. The values of K_{min} and K_{max} are chosen as two and $\sqrt{n}$ for all the algorithms, where n is the number of points in the data set.

TABLE 1
Description of the Data Sets

Name	# points	# clusters	# dimensions	Points per cluster
AD_10_2	500	10	2	50 per cluster
AD_4_3N	402	4	3	101,100,101,100
AD_2_10	1000	2	10	492,508
Crude_Oil	56	3	5	7,11,38
Cancer	683	2	9	444,239

4.2 Determining the Number of Clusters

The number of clusters provided by the three clustering algorithms in conjunction with the four validity indices for the different data sets is provided in Table 2. As can be seen from the table, the index $\mathcal{I}$ is able to indicate the correct number of clusters for all the data sets, irrespective of the underlying clustering technique. For AD_10_2, the values of $\mathcal{I}$ were found to be 299.288177, 295.271881, and 300.149780 when K-means, single linkage, and SA-based algorithms were used, respectively. In this context, one may note that the SA-based algorithm provides an improvement over that of K-means. As is widely known, the K-means algorithm often gets stuck at suboptimal values, a limitation that the SA-based method can overcome. The index for the single linkage algorithm may differ from those obtained using the other two clustering methods because of the difference of the underlying clustering principle. It may be noted from Table 2 that, irrespective of the clustering techniques that have been used in this article, none of DB, ν_D, or CH indices are able to find the appropriate number of clusters for AD_10_2. For AD_4_3N, it was found that both DB and CH (in addition to $\mathcal{I}$) were able to provide the exact number of clusters. On the contrary, ν_D failed to do so, irrespective of the clustering techniques that have been used here. Like $\mathcal{I}$, the CH index was found to provide the exact number of clusters for AD_2_10 for all three clustering techniques. However, the DB and ν_D indices failed to do so for this data. These are indicated in Table 2.

TABLE 2
Number of Clusters Provided by the Three Clustering Algorithms Using the Four Validity Indices for Different Data Sets

Data Set	Algorithm	Value of Index			
		DB	ν_D	CH	$\mathcal{I}$
AD_10_2	K-means	8	18	2	10
	S-Link	8	4	2	10
	SA	8	5	2	10
AD_4_3N	K-means	4	2	4	4
	S-Link	4	2	4	4
	SA	4	2	4	4
AD_2_10	K-means	30	14	2	2
	S-Link	30	14	2	2
	SA	30	14	2	2
Crude_Oil	K-means	2	6/7	2	3
	S-Link	3/4	9-13	2	3
	SA	4	2	2	3
Cancer	K-means	2	2	2	2
	S-Link	2	2	2	2
	SA	2	2	2	2

For Crude_Oil, apart from index $\mathcal{I}$, only the DB index provided the correct number of clusters when the single linkage algorithm was used. Even in this case, the minimum was not at a unique value of the number of clusters. As seen from Table 2, the number of clusters indicated in this case is three and four, when the DB index was found to attain the minimum value. The other two indices are unable to indicate the correct number of clusters. Cancer data has two classes which have only a small amount of overlap. As a result, all three clustering techniques, irrespective of the cluster validity index used, were found to provide the appropriate number of clusters for this data (Table 2).

4.3 Determining the Appropriate Clustering

The data set is partitioned into a number of clusters (K^*), whose value is obtained by noting the optimum of the validity index, as done in the previous section. Note that, for this purpose, we use index $\mathcal{I}$ since this is found to be the most reliable among the indices used. The corresponding $U^*_{K^*}$ is obtained by using this value of K^* in the simulated annealing-based clustering technique that uses a probabilistic redistribution of points. It is well-known that the K-means method has the limitation of getting stuck at suboptimal configurations, depending on the choice of the initial cluster centers. On the contrary, the simulated annealing based technique can overcome this limitation since it has the power of coming out of local optima. Fig. 1 and Fig. 2 demonstrate the results for AD_10_2 and AD_4_3N, respectively. Since the dimensionality of the other data sets is greater than three, their partitioning could not be demonstrated graphically.

5 Discussion and Conclusions

In this article, an extensive comparison of several cluster validity indices has been done for both artificial and real-life data sets, where the number of clusters and dimensions range from two to ten. In this regard, the performance of three crisp clustering algorithms, namely, hard K-Means, single linkage, and a simulated annealing-based clustering algorithm with probabilistic redistribution of data points are also studied. The validity indices are used to evolve the appropriate number of clusters. Subsequently, the simulated annealing-based clustering algorithm is utilized for appropriately partitioning the data into the said number of clusters.

In this context, a recently developed cluster validity index $\mathcal{I}$ is described in this article. This index is found to attain its maximum value when the appropriate number of clusters is achieved. Compared to the other validity indices considered, $\mathcal{I}$ is found to be more consistent and reliable in indicating the correct number of clusters. This is experimentally demonstrated for the five data sets, where $\mathcal{I}$ achieves its maximum value for the correct number of clusters, irrespective of the underlying clustering technique. A lower bound of the value of the index $\mathcal{I}$ is also theoretically estimated in order to get the unique hard K-partition when the data set has a distinct substructure. This is obtained as a result of establishing a relationship of the index $\mathcal{I}$ with the well-known Dunn's index and the Xie Beni index.

In addition to the experimental evaluation presented in this article, an extensive theoretical analysis comparing the validity indices needs to be performed in the future. Comparison with respect to the convergence speeds, as well as the effect of distance metrics, other than the Euclidean distance considered here, on the performance of the validity indices should be investigated. Note that the clustering algorithms and validity indices considered in this article are all crisp in nature. An extensive evaluation of fuzzy algorithms and validity indices needs to be carried out. In this regard, a fuzzy version of index $\mathcal{I}$ may also be developed.

ACKNOWLEDGMENTS

The authors sincerely acknowledge the reviewers for their valuable suggestions which helped in improving the quality of the article.

REFERENCES

[1] J.T. Tou and R.C. Gonzalez, *Pattern Recognition Principles.* Reading: Addison-Wesley, 1974.

[2] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data.* Prentice Hall, 1988.

[3] H. Frigui and R. Krishnapuram, "A Robust Competitive Clustering Algorithm with Application in Computer Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 1, pp. 450-465, Jan. 1999.

[4] B.S. Everitt, *Cluster Analysis,* Halsted Press, third ed., 1993.

[5] U. Maulik and S. Bandyopadhyay, "Genetic Algorithm Based Clustering Technique," *Pattern Recognition,* vol. 33, pp. 1455-1465, 2000.

[6] G.W. Milligan and C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika,* vol. 50, no. 2, pp. 159-179, 1985.

[7] M. Meilă and D. Heckerman, "An Experimental Comparison of Several Clustering and Initialization Methods," *Proc. 14th Conf. Uncertainty in Artificial Intelligence,* pp. 386-395, 1998.

[8] C. Fraley and A.E. Raftery, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," *The Computer J.,* vol. 41, no. 8, pp. 578-588, 1998.

[9] L.O. Hall, I.B. Ozyurt, and J. C. Bezdek, "Clustering with a Genetically Optimized Approach," *IEEE Trans. Evolutionary Computation,* vol. 3, no. 2, pp. 103-112, 1999.

[10] D.L. Davies and D.W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 1, pp. 224-227, 1979.

[11] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *J. Cybernetics,* vol. 3, pp. 32-57, 1973.

[12] R.B. Calinski and J. Harabasz, "A Dendrite Method for Cluster Analysis," *Comm. in Statistics,* vol. 3, pp. 1-27, 1974.

[13] S. Kirkpatrik, C. Gelatt, and M. Vecchi, "Optimization by Simulated Annealing," *Science,* vol. 220, pp. 671-680, 1983.

[14] S. Bandyopadhyay, U. Maulik, and M.K. Pakhira, "Clustering Using Simulated Annealing with Probabilistic Redistribution," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 15, no. 2, pp. 269-285, 2001.

[15] X.L. Xie and G. Beni, "A Validity Measure for Fuzzy Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, pp. 841-847, 1991.

[16] J.C. Dunn, "Well Separated Clusters and Optimal Fuzzy Partitions," *J. Cybernetics,* vol. 4, pp. 95-104, 1974.

[17] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis.* Prentice Hall, 1982.