

Estimation of the number of clusters using multiple clustering validity indices

Krzysztof Kryszczuk and Paul Hurley

IBM Zurich Research Laboratory, Switzerland

Abstract. One of the challenges in unsupervised machine learning is finding the number of clusters in a dataset. Clustering Validity Indices (CVI) are popular tools used to address this problem. A large number of CVIs have been proposed, and reports that compare different CVIs suggest that no single CVI can always outperform others. Following suggestions found in prior art, in this paper we formalize the concept of using multiple CVIs for cluster number estimation in the framework of multi-classifier fusion. Using a large number of datasets, we show that decision-level fusion of multiple CVIs can lead to significant gains in accuracy in estimating the number of clusters, in particular for high-dimensional datasets with large number of clusters.

Key words: clustering, clustering validity indices, multiple classifier

1 Introduction

Clustering algorithms are unsupervised machine learning techniques seeking to discover structure in unlabeled data sets. They analyze the similarities between the data and group it into similarity clusters [1, 2]. One of the properties of interest when analyzing unknown data is the number of such clusters. Discovering the number of clusters inherently present in the data has important practical applications, including computational biology [3], web text mining [4], etc.

To date no theoretically optimal method for finding the number of clusters inherently present in the data has been proposed. Existing algorithms include stability-based methods [5, 6], model-fitting-based algorithms [7], and methods based on *Clustering Validity Indices* (CVI) [1]. A CVI is a measure derived from the obtained clustering solution, which quantifies such properties of a clustering solution as compactness, separation between clusters, etc. The principal usage of CVIs is comparison of alternative clustering solutions. CVIs can be used for the task of cluster number estimation, if the compared clustering solutions use different number of clusters [8].

A recurring problem with the use of CVIs is that each CVI is designed to capture a specific aspect of the clustering solution that suggests how adequate that solution is. At the same time, other aspects can be inadequately represented or ignored altogether. For instance, a CVI that verifies how compact the obtained clusters are will frequently report that elongated clusters are not compact at all,

although such clusters may correctly represent the data. No CVI can a-priori be assumed better than its alternatives [8].

In this paper, we propose to overcome this problem by estimating the number of clusters by a fusion of multiple CVIs in order to estimate the number of clusters in the data. We notice an analogy between using multiple CVIs and constructing multiple classifier ensembles [9]. The motivation of our work is that although no single CVI can capture correctly the validity of any clustering solution, a proper conciliation of multiple CVIs should be able to cope with the task. In this work we treat the yield of each considered CVI as a feature that characterizes the quality of a clustering solution. We hypothesize that an ensemble of CVI features is a better predictor of clustering quality than any of the CVIs taken separately. To the best of our knowledge, the only attempt to combine validity indices into a compound index was reported in [10]. However, fusion of CVIs for cluster number estimation is not the main topic of [10]. Consequently, only simple score averaging of CVIs were used, and no comparison with other possible solutions was presented. In this paper, we compare score-, and decision fusion strategies of combining multiple CVIs using a large number of synthetic datasets. The results of our experiments suggest that decision-based multiple CVI fusion techniques offer highest accuracy in detecting the correct number of clusters in a dataset, provided that all used CVIs are individually competent. The result of this paper are particularly relevant to high-dimensional data analysis with multiple clusters, when visual inspection of the clustering solution is impractical or impossible.

The rest of the paper is structured as follows. Section 2 gives an overview of the CVIs used in this work. In Section 3 we provide the framework of cluster number estimation using multiple CVIs. The applied experimental methodology and the experimental results are described in Section 4. Section 5 gives a discussion of the results and concludes the paper.

2 Internal Clustering Validity Indices

In this work, we used three popular, internal clustering validity indices, the Davies-Bouldin, Calinski-Harabasz and Dunn indices, as in [11]. The choice of the CVIs was dictated by the fact that they are well-known measures, frequently used for the task of choosing the best clustering solution. The Davies-Bouldin and the Calinski-Harabasz indices were reported as performing well for the task of comparing clustering solutions [12].

Davies-Bouldin index is computed as

$$s_{DB} = \frac{1}{k} \sum_{i=1}^k \max_{j=1,\dots,k; i \neq j} \left(\frac{d_i + d_j}{d(c_i, c_j)} \right), \quad (1)$$

where k denotes the number of clusters. If i, j are cluster labels, then d_i and d_j are average distances of all patterns in clusters i and j to their respective cluster centroids, and $d(c_i, c_j)$ is the distance between these centroids [13]. Smaller value

of s_{DB} indicates a "better" clustering solution. In order to be consistent with other two clustering validity indices considered in this work, we used $1 - s_{DB}$ to have a maximum as an indicator of the preferred clustering solution.

Calinski-Harabasz index is computed by

$$s_{CH} = \frac{\text{trace}(S_B)}{\text{trace}(S_W)} \cdot \frac{n_p - 1}{n_p - k}, \quad (2)$$

where S_B is the between-cluster scatter matrix, S_W the within-cluster scatter matrix, n_p the number of clustered points, and k the number of clusters [14]. When using the CH index to compare clustering solutions, maximal value of s_{CH} identifies the preferred candidate.

Dunn index is defined as

$$s_{Dunn} = \frac{d_{min}}{d_{max}}, \quad (3)$$

where d_{min} is the minimum distance between two points belonging to different clusters, and d_{max} is the maximum distance between any two points selected from the same cluster [15]. The highest value of s_{Dunn} indicates the optimal clustering solution among the considered candidates.

3 Using multiple CVIs to compare clustering solutions

3.1 Estimation of the number of clusters using internal CVI

The application of a CVI to determine the number of clusters of data is straightforward. Given a set of alternative clustering solutions, the preferred one is found by comparing the CVIs calculated for several candidate clustering solutions [8, 16, 17]. In order to find the number of clusters, a clustering algorithm is used with the number of clusters equal to $k \in \Omega$, where Ω is the ordered set of candidate cluster numbers. The preferred clustering solution is obtained by finding the value of k_{est} that maximizes the function $CVI(k)$ over all values from Ω . Under the assumption that the chosen clustering algorithm is able to cluster the data properly, then $k_{true} = \arg_{k \in \Omega} \max(CVI(k))$ is the sought number of clusters.

There is no single theoretically optimal CVI that is guaranteed to work well with all datasets. The reported empirical comparisons of CVI performance argue in favor of different algorithms [12, 18, 17, 11, 19]. In the absence of theoretically proven superiority of any CVI over the rest of the pack, we conclude that all CVIs capture certain aspects of a clustering solution that helps in comparing it to its alternatives, but none does it universally better than others.

The situation is strikingly similar to the dilemma of comparing and selecting best classifiers in pattern recognition, where the *no free lunch theorem* rules that there is no universally best classifier [8]. Multi-classifier ensembles have been

shown to offer a powerful alternative to single-classifier solutions [20] as a particularly effective option when the considered individual classifiers are diversified and accurate [9].

In this paper we approach the problem of estimating the number of clusters from a multi-classifier fusion perspective. From this angle, estimation of the number of clusters is a classification problem, where each class corresponds to a particular number of clusters in the considered dataset. We consider individual CVI algorithms as feature extractors. Resulting features can be used individually for cluster number estimation, as it is the common case; or can be used together using a combining fusion classifier.

3.2 Fusion methods for Clustering Validity Indices

The multi classifier fusion can be realized by means of trained or rule-based combining classifiers. In general, trained fusion classifiers are considered to be more flexible and powerful than their heuristic, rule-based counterparts, given that an adequate training data corpus is available [21]. In the case of the CVI combination, it is not evident how to construct such a training corpus, and therefore heuristic fusion techniques are considered in this paper. In particular, we consider score-, and decision- level fusion schemes [20].

Score fusion-based methods A combined score s_F is computed using i normalized CVI scores $[s_1, s_2, \dots, s_i]$, as follows

- **SF-A:** $s_F = \frac{1}{i} \sum_{m=1}^i s_m$,
- **SF-G:** $s_F = (\prod_{m=1}^i s_m)^{\frac{1}{i}}$,
- **SF-H:** $s_F = \frac{i}{\sum_{m=1}^i \frac{1}{s_m}}$,
- **SF-Med:** $s_F = \text{median}(\{s_1, s_2, \dots, s_i\})$.

In order to perform a meaningful score-based fusion it is necessary to normalize the scores $[s_1, s_2, \dots, s_i]$ to a common range. We used a min-max normalization, which scaled the scores to a $0 \leq s_i \leq 1$ range. The normalization was performed within each clustered dataset, across all candidate $k \in \Omega$. We also evaluated other score normalization schemes, including z-norm across $k \in \Omega$, as in [10], and global z-norm, where the mean and variance of s_i were estimated using a large range of possible clustering solutions. We obtained the best results with the min-max normalization and this scheme was used in the experiments reported. An example of score-based fusion using the four considered strategies is shown in Figure 1(b).

Decision fusion-based methods Bezdek and Pal suggested the use of “some voting scheme” for CVIs [18]. Let $\Psi = \{1, 2, \dots, n\}$ be a vector of class labels, each corresponding to estimated cluster number from Ω . A set of class decisions $d_i \in \Psi$, originating from i individual CVIs is obtained. If $m \in \Psi$ is a class label, then $d_i = \arg_m \max(s_i(m))$. Consequently, the decision fusion strategies we considered were

- **DF-A:** $d_F = \text{round}(\frac{1}{i} \sum_{m=1}^i d_m)$,
- **DF-G:** $d_F = \text{round}((\prod_{m=1}^i d_m)^{\frac{1}{i}})$,
- **DF-Mod:** $d_F = \text{mode}(\{d_1, d_2, \dots, d_i\})$,
- **DF-Med:** $d_F = \text{median}(\{d_1, d_2, \dots, d_i\})$.

At first, computing the mean of decision values may seem counterintuitive. Indeed, in order for $DF - A$, $DF - G$ and $DF - Med$ to return meaningful results, the ordered set $\Omega = \{n_1, n_2, \dots, n_m\}$ must be constructed so that $n_1 < n_2 < \dots < n_m$. In this situation, the fusion algorithm will compensate for over- and underestimated k_{est} , given that enough single CVI decisions are involved. In our experiments $n_{i+1} - n_i = 1$ for all i , where i are indices of n in Ω . An example where decision fusion returns meaningful results is shown in Figure 1. In Figure 1(a), individual CVIs incorrectly return 9, 9 and 12 as the estimated number of clusters k_{est} , respectively. In Figure 1(b), the decision fusion strategies return $DF_A = DF_G = 10$ (correctly), and $DF_{Mod} = DF_{Med} = 9$.

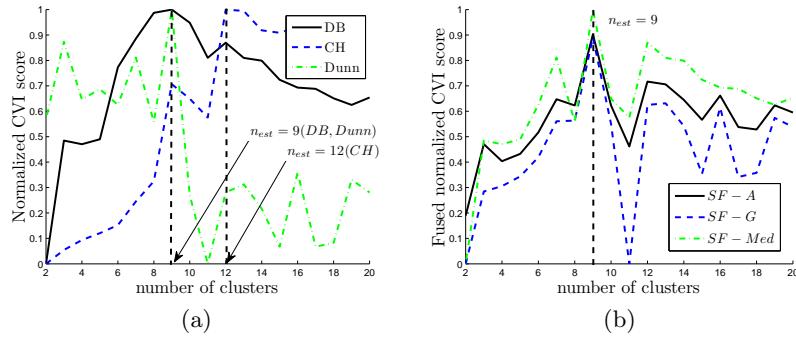


Fig. 1. Normalized CVI scores (a) and, fused scores using considered score-based fusion strategies (b), for $k_{true} = 10$ clusters. In the example shown, all of the CVIs (a) and fused scores (b) point out to the incorrect number of clusters $k_{est} = 9$. In this example, decision fusion strategies return $DF_A = 10$ and $DF_G = 10$, the correct number of clusters.

Decision rank fusion-based methods In our experiments we compared a number of decision rank fusion techniques. However, the results we obtained proved to be not competitive in comparison to other reported fusion strategies and we exclude them from this report due to space constraints.

4 Experimental evaluation of CVI fusion techniques

Data generation and normalization In order to evaluate the performance of the compared CVI fusion schemes we produced a large number of randomly-generated datasets. This choice allowed us to fully control the clustering tendency present in the clustered data, as well as the data properties such as dimensionality n_{dim} , number of data clusters k_{true} and number of data points per cluster n_p . We also used a clustering algorithm that suits the shape of the actual data clusters, in order to reduce the impact of incorrect clustering on the observed CVI performance. This level of control is hard to achieve using scarce real data.

For each generated dataset, its n_{dim} , k_{true} and n_p were set. For each cluster a centroid vector and a variance vector were generated from a uniform multivariate distribution. Each mean vector was then scaled using a parameter p . So obtained, randomly generated mixture model was accepted only if all distances between the cluster centroids were greater than a threshold t (in this work $t = 0.1$). This procedure ensured that two or more clusters are not randomly placed on top of one another, making it hard or impossible to estimate the true number of clusters. For each cluster, n_p data points were randomly generated from a Gaussian distribution with variance and mean defined by the corresponding centroid and variance vector. Finally, the data was scaled to fit into a unit hypercube with dimensionality equal to the dimensionality of the generated dataset n_{dim} . Figure 2 shows an example of a generated 2-dimensional dataset.

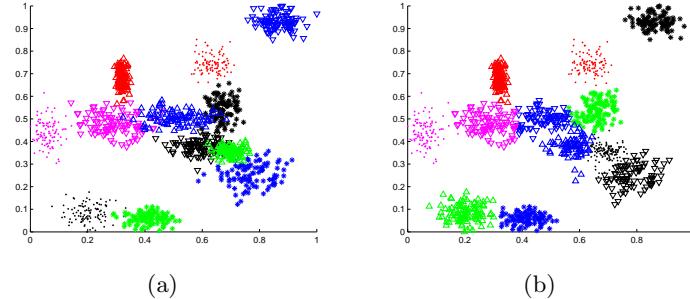


Fig. 2. Sample generated clustering problem for $p = 5$, $k_{true} = 12$ (a), and corresponding solution found by multiple k-means clustering with $k = k_{true}$ (b).

Evaluation method In the experiments, we compared the accuracy of the cluster number estimation procedures involving single CVIs with the procedures involving considered multi-CVI fusion approaches described in Section 3.2. For each fixed data dimensionality n_{dim} and the number of clusters k_{true} , we generated a batch of 100 datasets, as described in Section 4.

The normalized data from each set of the batch was clustered using the k-means algorithm with Euclidian distance measure [8], with the candidate number of clusters $k = [2, 3, \dots, 20]$. The clustering was repeated 10 times for each

value of k , each time using a random initialization of prototype cluster centroid positions. The clustering solution with the least mean within-cluster sum of point-to-centroid distances was selected as the best for the given k . Then, the three CVIs (s_{DB} , s_{CH} and s_{Dunn}) were computed for each k . Consequently, for each batch of data we obtained 19 sets of 3 CVI scores, which were normalized and fused using the methods described in Section 3.2. Finally, as the output for each clustered dataset we obtained a set of estimates k_{est} of the true number of clusters, as described in Section 3.1.

In the assumed classification framework, we compared the performance of the CVIs and CVI fusion methods by computing the accuracy of assigning a correct class label to the obtained CVIs. Figure 3 shows an example of the results obtained from one data batch of 100 random datasets, with dimensionality $n_{dim} = 10$, for $k_{true} = 12$. Each plot shows histograms of classification labels (decisions) k_{est} over 100 datasets, for single and fused CVIs. The histogram value corresponding to $k_{est} = k_{true}$ bin label is the recorded classification accuracy for a given method of cluster number estimation.

Experimental results We analyzed the performance of single CVIs versus CVI fusion-based methods for $k_{true} = [5, 6, \dots, 15]$, $n_{dim} \in \{2, 10, 20\}$, for constant value of $p = 5$ and constant number of points per cluster $n_p = 20$. The number of classes was 19: candidate $k = [2, 3, \dots, 20]$. The results of the experiment are given in Table 1.

5 Discussion and Conclusions

The results of the experiments reported in Section 4 show that combining CVIs is a viable way of increasing the robustness and accuracy of cluster number estimation in comparison with using a single CVI. In the reported experiments, the CVI fusion methods outperformed single CVIs in all cases, where all CVIs used in fusion were similarly accurate. An underperforming CVI often compromised the accuracy of the fusion results, as in Table 1 for $n_{dim} = 2$. This effect is due to the fact that the used heuristic fusion methods have no weighting mechanism that could identify and marginalize, or decrease the contribution of the underperforming CVI in the fusion process. The detrimental effect of using an underperforming CVI was particularly pronounced in the case of the score fusion techniques.

The decision-based fusion rules showed to be in general more robust and yielded higher accuracy than the score-based fusion methods. It can be explained by the fact that for all CVIs, the deviations from their corresponding maximal value $|s_i(k_{est}) - s_i(k \neq k_{est})|$ were of different scale and therefore not comparable quantitatively. However, in the decision-based fusion this deviation is comparable qualitatively: only the information of $\arg_{k \in \Omega} \max s_i(k)$ is preserved.

The overall best-performing scheme in the experiments reported was the mean-rule decision fusion scheme DF_A , closely followed by the DF_G scheme. This result can be attributed to the fact that the decision-based schemes are capable of

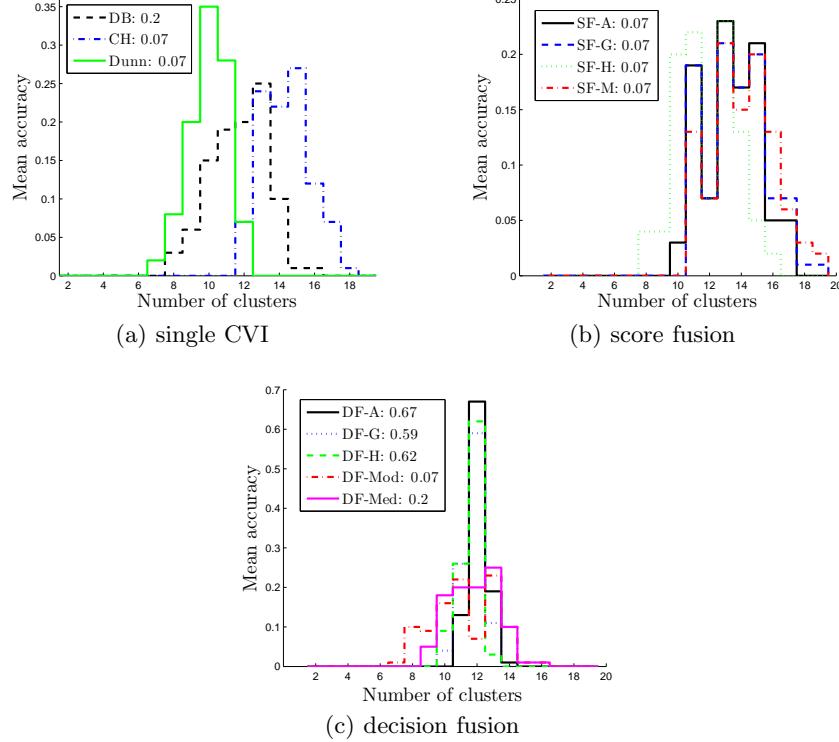


Fig. 3. Experimental results of a comparison of CVIs (a), and score- (b), and decision-based (c) CVI fusion methods, for $k_{true} = 12$ clusters, 19 classes ($k = [2, 3, \dots, 20]$), averaged over 100 randomly generated datasets.

compensating for CVIs that over- and underestimate k_{est} , while being insensitive to score normalization. This result shows that the individual CVIs indeed make different errors and therefore are diversified enough to merit an application of a multi-classifier fusion scheme. We anticipate that a larger number of fused CVIs would reinforce this effect.

In our experiments, the benefits of using the CVI fusion techniques became evident with the increase of the number of clusters within the hypercube, and with the increase of the dimensionality of the clustered data. In fact, as Table 1 shows, for low k_{true} (good class separation) the benefit of using fusion is immaterial. At the same time, for high-dimensional data ($n_{dim} = 20$) and high value of k_{true} , these benefits are significant, with the DF_A and $DF - G$ schemes yielding more than double the accuracy of the best individual CVI.

The experiments reported in this paper include only a limited number of examples to support the claim that using a fusion of multiple CVIs can result in higher accuracy of cluster number estimation than using a single CVI. Obviously, a single CVI that is well-selected for the problem at hand, can produce accuracy

Table 1. Results of the experimental evaluation of the compared CVIs and CVI fusion strategies, for $p = 5$, $n_p = 20$, $n_{dim} = \{2, 10, 20\}$, the actual number of cluster $k_{true} = [5, 6, \dots, 15]$, with 19 candidate number of clusters (classes), $k = [2, 3, \dots, 20]$. Maximal accuracy per k_{true} is marked in bold font.

k_{true}	s_{DB}	s_{CH}	s_{Dunn}	$SF - A$	$SF - G$	$SF - H$	$DF - A$	$DF - G$	$DF - Mod$	$DF - Med$
Dimensionality $n_{dim} = 2$										
5	0.41	0.32	0.32	0.4	0.26	0.3	0.21	0.32	0.34	0.42
6	0.23	0.44	0.15	0.2	0.23	0.42	0.24	0.24	0.16	0.26
7	0.29	0.41	0.17	0.26	0.2	0.41	0.22	0.23	0.24	0.3
8	0.2	0.41	0.14	0.23	0.12	0.39	0.15	0.19	0.2	0.27
9	0.14	0.45	0.06	0.14	0.11	0.43	0.15	0.14	0.12	0.18
10	0.14	0.28	0.09	0.13	0.06	0.27	0.2	0.11	0.11	0.17
11	0.11	0.25	0.05	0.11	0.04	0.24	0.13	0.05	0.06	0.13
12	0.08	0.21	0.04	0.08	0.05	0.19	0.12	0.13	0.05	0.09
13	0.19	0.16	0.04	0.13	0.05	0.22	0.08	0.06	0.08	0.21
14	0.09	0.14	0	0.05	0.02	0.13	0.04	0.04	0.03	0.11
15	0.12	0.16	0.03	0.05	0.04	0.16	0.07	0.06	0.04	0.16
Dimensionality $n_{dim} = 10$										
5	0.98	0.98	0.97	0.97	0.94	0.9	1	1	0.98	0.98
6	0.89	0.89	0.88	0.89	0.89	0.85	0.99	0.99	0.89	0.89
7	0.73	0.75	0.75	0.75	0.78	0.68	0.92	0.91	0.75	0.75
8	0.44	0.49	0.49	0.49	0.5	0.48	0.81	0.76	0.49	0.5
9	0.34	0.32	0.32	0.32	0.36	0.3	0.73	0.67	0.32	0.34
10	0.23	0.2	0.2	0.2	0.22	0.2	0.6	0.54	0.2	0.23
11	0.28	0.15	0.15	0.15	0.17	0.14	0.62	0.51	0.15	0.28
12	0.25	0.18	0.18	0.18	0.2	0.17	0.62	0.54	0.18	0.25
13	0.21	0.05	0.05	0.05	0.08	0.05	0.55	0.43	0.05	0.21
14	0.16	0.03	0.03	0.03	0.09	0.03	0.48	0.31	0.03	0.16
15	0.14	0.05	0.03	0.03	0.22	0.05	0.34	0.29	0.03	0.14
Dimensionality $n_{dim} = 20$										
5	0.96	0.96	0.96	0.96	0.96	0.91	1	1	0.96	0.96
6	0.8	0.8	0.8	0.8	0.82	0.77	0.97	0.97	0.8	0.8
7	0.71	0.71	0.71	0.71	0.77	0.7	0.93	0.91	0.71	0.71
8	0.46	0.46	0.46	0.46	0.51	0.44	0.87	0.87	0.46	0.46
9	0.36	0.33	0.32	0.32	0.45	0.33	0.85	0.75	0.32	0.36
10	0.22	0.16	0.16	0.16	0.22	0.16	0.71	0.66	0.16	0.22
11	0.24	0.15	0.16	0.16	0.19	0.15	0.78	0.74	0.16	0.24
12	0.18	0.08	0.08	0.08	0.08	0.08	0.64	0.54	0.08	0.18
13	0.17	0.03	0.03	0.03	0.06	0.03	0.61	0.51	0.03	0.17
14	0.17	0.05	0.04	0.04	0.09	0.05	0.47	0.38	0.04	0.17
15	0.18	0.06	0.02	0.02	0.15	0.05	0.45	0.45	0.05	0.18

competitive to, of higher than, the fusion result. This effect, well-known from rule-based classifier fusion, is particularly relevant if other used CVIs yield low accuracy. However, in many practical applications the shape and number of clusters is unknown, making a-priori selection of a suitable single CVI impossible. This is particularly the case if the data is of high dimensionality, cannot be easily visualized, and contains many clusters with few members each. Our results suggest that in such scenarios a fusion of CVIs for estimation of the number of clusters can be successfully applied. In the future, we intend to validate our findings using real datasets and using different clustering algorithms.

References

1. Jain, A., Dubes, R.: Algorithms for Clustering Data. Prentice-Hall (1988)
2. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: A review. *ACM Computing Surveys* **31**(3) (1999) 255–323
3. Dudoit, S., Fridlyand, J.: A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology* **3**(7) (2002) 0036.10036.21
4. Yao, Z., Choi, B.: Automatically discovering the number of clusters in web page datasets. In: Proceedings of the 2005 International Conference on Data Mining. (2005) 3–9
5. Ben-Hur, A., Elisseeff, A., Guyon, I.: A stability based method for discovering structure in clustered data. In: Pacific Symposium on Biocomputing. (2002) 6–17
6. Lange, T., Roth, V., Braun, M.L., Buhmann, J.M.: Stability-based validation of clustering solutions. *Neural Computation* **16**(6) (2004) 1299–1323
7. Zhang, J., Modestino, J.W.: A model-fitting approach to cluster validation with application to stochastic model-based image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(10) (1990) 1009–1017
8. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, Second Edition. Elsevier (2003)
9. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons (2004)
10. Machado, J.B., Amaral, W.C., Campello, R.: Design of obf-ts fuzzy models based on multiple clustering validity criteria. In: Proc. of the 19th IEEE International Conference on Tools with Artificial Intelligence. (2007)
11. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(12) (2002) 1650–1654
12. Milligan, G.W., Cooper, M.C.: An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50** (1985) 159–179
13. Davies, D.L., Bouldin, D.W.: A clustering separation measure. *IEEE Transactions on PAMI* **1** (1979) 224–227
14. Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. *Communications in Statistic* **3** (1974) 127
15. Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* **3** (1973) 32–57
16. Halkidi, M., Batistakis, Y., Vazirgiannis, M.: Cluster validity methods: part i. *SIGMOD Rec.* **31**(2) (2002) 40–45
17. Halkidi, M., Vazirgiannis, M.: Clustering validity assessment: Finding the optimal partitioning of a data set. In: Proceedings of the 2001 IEEE International Conference on Data Mining. (2001) 187–194
18. Bezdek, J.C., Pal, N.R.: Some new indexes of cluster validity. *IEEE Trans. on Systems, Man, and Cybernetics B* **28**(3) (1998) 301–315
19. Shim, Y., Chung, J., Choi, I.C.: A comparison study of cluster validity indices using a nonhierarchical clustering algorithm. In: Proc. of the 2005 Int. Conf. on Comp. Intelligence for Modelling, Control and Automation, and Int. Conf. on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC05). (2005)
20. Kittler, J., Hataf, M., Duin, R., Matas, J.: On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3) (1998) 226–239
21. Duin, R.P.W.: The combining classifier: to train or not to train? In: Proceedings of the International Conference on Pattern Recognition, Quebec, Canada (2002)