

Clustering Ensemble Selection Considering Quality and Diversity

Roham Ranjbar, Hamid Parvin*, and Farhad Rad

Department of Computer Engineering, Yasooj Branch,
Islamic Azad University, Yasooj,
Iran

parvin@iust.ac.ir

Abstract. Information clustering means classifying information or partitioning some samples in clusters such that samples inside each cluster have maximum similarity to each other and maximum distance from other clusters. As clustering is unsupervised, selecting a specific algorithm for clustering of an unknown set may fail. As a consequence of problem complexity and deficiencies in basic clustering methods, most of studies have focused on ensemble clustering methods in recent years. Diversity in initial results is one of the most important factors which may affect final quality of the results. Moreover, the quality of primary results affects the quality of final results. Both factors have been investigated in recent studies on clustering. Here, a new framework is proposed which is used for improving clustering efficiency and it is based on use of a subset of initial clusters. Selection of this subset plays a significant role in performance of the scheme. The subset is selected using two intelligent methods. The main idea in these methods is utilizing stable clusters through intelligent search algorithms. Two stability factors are utilized for cluster evaluation. One of these two stability factors is based on mutual information and the other one is based on Fisher measure. Finally, the selected clusters are added using several final combining methods. Practical results of several standard data sets demonstrate that the proposed method may improve combination clustering method significantly.

Keywords: Clustering combination, local optimization, diversity.

1 Introduction

Clustering is a branch of unsupervised learning. It is an automatic process through which samples are divided into groups with similar members which are called clusters. Thus, cluster is a set of objects which are similar to each other while they are different from objects inside other clusters. Various criteria might be considered for

*Corresponding author.

similarity, for instance, distance could be used for clustering and objects which are closer could be clustered as one cluster; it is called distance based clustering. In unsupervised methods no objective variable is defined and data-mining algorithm searches correlations and structures of all variables. Clustering is the most prominent example of unsupervised data mining.

As mentioned before clustering is putting similar objects together; however, it must be found out how a clustering system is evaluated. As a matter of fact there is not any absolute measure for determining the best clustering method and it depends on the problem and user's opinion; nevertheless, there are various measures to determine a good clustering which may help the user to achieve a proper clustering. Some of these measures are explained in efficiency measures section.

In statistics and machine learning, clustering or cluster analysis is the procedure of grouping similar objects. The clustering problem might be introduced in two ways: 1) a $n \times n$ dissimilarity (similarity) matrix is given, 2) a $n \times d$ matrix is given where each row defines an object. The output of the algorithm could be in two forms: 1) grouping the objects to separated sets 2) hierarchical clustering which finds a tree for division of objects. The algorithms of first group are faster. It must be noticed that this paper does not deal with hierarchical clustering. Each clustering algorithm cluster the data in a unique way as it focuses on a specific aspect of the data. Thus, it is necessary to combine such algorithms, take advantage of a few algorithms and provide optimum results. Actually, the main goal of ensemble clustering is searching for the best clusters obtained via combining other algorithms [1, 2]. Combination clustering may provide better results from stability, flexibility and robustness perspectives [2-4]. To sum up, combination clustering includes the following steps: a) generating different subsets from whole given data, b) initial clustering based on applying various clustering algorithms on the subsets generated from main samples. c) combining the results obtained from primary clustering methods to achieve final clustering. There are two important issues regarding combination clustering: 1) diversity of various clustering algorithms such that each of them focuses on a specific characteristic of the data, 2) the combining algorithm which provides the final results. To address the first issue the following methods might be exploited: 1) using different clustering algorithms [5], 2) changing initial values or other parameters of the clustering algorithm [4, 6], 3) selecting some data features or generating new features [1, 3, 7], 4) dividing the main data to different and separated subsets [8, 9, 10, 11, 12, 13]. The second issue is vastly investigated to find algorithms for combining the results [14, 15, 16]. But the proposed methods have been static ones so far rather than dynamic methods. In this paper a dynamic approach is provided. Despite, information classification which has a supervisor and training set, in information clustering the data set is completely unknown. Lack of supervisor and training set makes it difficult to introduce modern and smart clustering methods with high efficiency. One of the methods which might be used to achieve smart information classification is considering diversity concept.

Diversity in information classification means that if one classifier has errors in some samples, we look for other classifier which has errors in samples different from errors of the first classifier. As a result the classifiers provide better results. Lack of training set has deprived information clustering methods from such technique. Here, it is tried to include diversity concept in information clustering [13, 21, 22]. Diversity

concept has been utilized widely in recent research works [3, 13, 23, 24]. The main goal of recent combination clustering methods is examining data set from different perspectives and it has not been investigated whether the generated diversity is useful or not. Indeed, it is difficult due to unsupervised nature of clustering problem. However, practical results have demonstrated that generating diversity in primary clusters usually leads to better results [25]. Azimi [26] has shown that in some data sets more diversity does not necessarily increase final precision. In this paper, diversity and clustering quality have been simultaneously emphasized.

Routing in computer networks has played a special role in recent years. The cause of this is the role of routing in a performance of the networks. The quality of service and security is one of the most important challenges in routing due to lack of reliable methods.

2 Heuristic Search Methods

Increasing complexity of optimization problems has necessitated novel search methods. To address this issue, heuristic methods have been developed in many fields as a powerful optimization and search tool in recent decades. Their wide ranged applications, simple use and the capability of obtaining near to optimum solution has made these methods successful ones. In this section two heuristic algorithms which are utilized in this paper are briefly explicated.

Genetic algorithm is a scheme which considers the natural evolution of creatures [4]. It tries to imitate evolution process using computer algorithms. The most essential principle of evolution is inheritance. John Holland innovated genetic algorithm for the first time during 70s according to evolution theory. The algorithm exploits the same principles that natural evolution uses [4] to improve solutions of an optimization problem. There are two important operators in genetic algorithms which make the solutions chaotic to exit probable local optimums. One of these operators is crossover through which genetic algorithm generates solutions. Another operator called mutation is able to provide new values for bits which do not exist in the parents. Mutation guarantees genetic diversity and pushes the search to new domains.

Simulated annealing is an optimization method which is similar to the process through which metals are heated and then slowly annealed [7]. It is suitable for simple objective functions with one local bound point (minimization or maximization problems). For complex functions (for example for minimization problems) the local optimum points might be completely different from global optimum point. In such cases the optimization model will not be able to provide optimum solution. SA utilizes stochastic release so that it exits local minima points.

SA procedure starts from a possible solution such as q_0 (a real vector which shows all decision variables) and its corresponding objective function $J_0 = J(q_0)$. A new solution q_1 with objective function $J_1 = J(q_1)$ is randomly selected and evaluated among the neighboring area of the initial solution. The amount of variation in decision variable is usually known. Random nature is due to direction or dimension of changes (e.g. x variation might be known but its direction might be unknown). If the new solution has lower objective function $J_1 < J_0$ (for minimization problem), it is accepted and the search process is transferred to point q_1 . If the new solution is not

better than current solution ($J_1 \geq J_0$) it may be selected or rejected which depends on the following acceptance probability.

$$p_{acc} = e^{-\frac{J_1 - J_0}{T}}$$

3 Literature Review

Combining clustering methods is more difficult than combining supervised classifications. In clustering data set is not known. It is difficult to propose high efficiency smart clustering methods due to lack of supervisor and training set. Combination clustering is a method of clustering which is resulted from combining different clustering methods. Two main steps of generating a combination of initial clustering methods are generating each clustering method and using a function or mechanism to combine their results to obtain the final results.

Since the final result is a combination of initial clustering results, the more different initial results lead to better final result. As a matter of fact, if data are investigated from more different aspects the final result would more precise. There are various methods for generating diversity in combination clustering including different clustering algorithms, changing initial values or other clustering parameters, selecting some data features or generating new features and dividing main data to different and separated subsets. In the presented methods the main goal is to examine data set from different points of view. They have not investigated whether the generated diversity is useful or not [17].

Usually, most combination clustering methods use k-mean algorithm as their initial clustering method [12, 13, 18]. It has been shown that in some cases other clustering methods might be more beneficial considering behavior of each data set [1, 3]; nonetheless, k-mean algorithm has been the first choice due to its simplicity and appropriate ability.

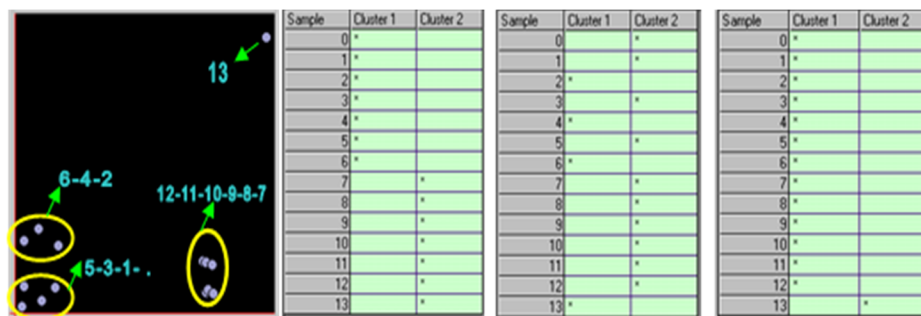


Fig. 1: Initial samples in k-mean algorithm. Figures from left to right: 1) space view of 14 samples, 2) results obtained using 1 and 8 initial samples, 3) results obtained from 2 and 3 initial samples and 4) results obtained from 1 and 13 initial samples

Another method for increasing diversity is changing initial parameters of clustering algorithms. For instance changing number of clusters in k-means or changing seed points significantly affect diversity [15]. In the following figure, the

effect of initial samples on final clustering is clear. In this figure, the distribution of samples is shown and the results of three different runs of algorithm with three different starts are depicted [1].

Feature selection might be considered as a method for adding diversity to combination clustering as well. Thus, another solution to increase diversity in combination clustering is using some features of total data set space or generating new features [13]. However, in information clustering selecting a subset of features has not been noticed due to the unsupervised nature of the problem. It is mostly tried to generate new features. There are several methods for feature generation in ensemble clustering [19] among which the simplest is data normalization. In fact it is shown that each data set achieves better behavior using one normalization method. As a result in many methods presented for information clustering, raw value of results are reported.

4 Consensus Function

Combining the first ensembles and obtaining the final result is one of the most important steps of combination clustering. There are diverse methods for combining the results of initial ensembles some of which are introduced in this section. Then, the proposed method is presented.

1- Hyper-graph based method

In hyper-graph partitioning, first off, the combination clustering problem is converted to a graph partitioning problem. Afterwards, the problem is solved using graph partitioning algorithms. Clusters are denoted by hyper edges of a graph. Graph vertexes correspond to samples which must be clustered. The problem is dividing this graph and generating k separated partitions each of which belongs to a cluster. There are three different algorithms in this group of methods including HGPA, CSPA and MCLA [1, 3].

1-1- CSPA

In CSPA feature space of data points is mapped to correlation feature space of hyper-graph. Then, a hyper graph minimum cut algorithm similar to METIS is applied to the data points. According to assumptions of this method more data points in one cluster in primary portioning means that data points are more probable to naturally belong to one cluster. CSPA is the simplest heuristic method. Its computational complexity is $O(kN^2M)$ where k is the number of clusters, N is the number of data points and M is the number of domains. The computational complexity of two other methods is lower than CSPA.

1-2- HGPA

HGPA consider vertexes as data points. Also, clusters which are resulted from initial portioning are assumed to be hyper edges. Now a hyper-graph minimum cut algorithm such as METIS is applied to hyper-graph for separating vertexes to k different components. Its computational complexity is $O(kNM)$ where k is the number of clusters, N denotes the number of data points and M is the number of domains.

1-3- MCLA

MCLA algorithm partitions the cluster resulted from initial partitioning. Afterwards, it utilizes a voting mechanism to generate set partitions. Clustering is

done using METIS. Its computational complexity is $O(k^2NM^2)$ where k , N and M are similar to previous methods. For more details regarding hyper graph based methods interested reader is referred to [3].

2- Voting method

This is actually majority of vote method. The cluster to which each sample belongs is determined according to majority votes. The main problem of this method is matching of cluster numbers in different runs which imposes heavy computational overhead on the algorithm. This computational overhead has caused this method to be unpopular among various consensus function methods [2, 8, 23].

3- Co-association matrix

Consider D as a data set consisting of N points (samples) in a d dimensional space. The input data might be considered either as a $N \times d$ pattern matrix or a dissimilarity $N \times N$ matrix. Assume that $X = \{X_1, X_2, \dots, X_{B_1}\}$ is a subset of available samples extracted from initial samples. All algorithms generate $P = \{P_1, P_2, \dots, P_{B_1}\}$ when they are applied to samples inside X . Each P_i is a set of clusters i.e. $P_i = \{C_1^i \cup C_2^i \dots \cup C_{k(i)}^i\}$ and $X_i = C_1^i \cup C_2^i \dots \cup C_{k(i)}^i$ such that $k(i)$ is the number of clusters in i th ensemble.

The first base algorithm which is utilized is k -means algorithm. At first step, k -means algorithm is executed on $X = \{X_1, X_2, \dots, X_{B_1}\}$ so that the co-association matrix could be derived as follows using generated P_i s.

$$Co - associatio n(xy) = \sum_{i=1}^{B_1} \lambda(P_i(x), P_i(y))$$

where

$$\lambda(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$

$\lambda(P_i(a), P_i(b))$ acquires 1 if in P_i combination a and b are located in the same cluster and it would be zero otherwise. B_1 denotes the number of subsets i.e. the number of times that k -mean base algorithm is repeated. When co-association matrix is obtained, final clusters are extracted from co-association matrix employing a simple hierarchical algorithm such as average link (AL).

5 Proposed Procedure

Clustering combination is more difficult than combining supervised classifications. Despite classification problem which has supervisor and training set, in clustering there is not any information available about data set. It is difficult to present high efficiency modern and intelligent methods due to lack of supervisor and training set. Furthermore, when labeled training data is not available, problem of correspondence between cluster labels in different partitions of a combination arises. Recent clustering methods mainly try to examine dataset from different perspective while they mostly ignore whether the diversity is useful or not. Indeed, it is difficult to do so as a result of unsupervised nature of clustering. Although experimental results have

shown that diversity improves clustering in most cases [29], Azimi [30] demonstrated that in some data sets more diversity does not necessarily increase precision. Since there is not any true labels (supervisor), clustering is one of the most difficult and ambiguous concepts in artificial intelligence; as all partitions of data might be considered to be correct. One of the problems is shortage of a precise and absolute measure for clustering which could be optimized to obtain the best clustering.

In this section a novel scheme is proposed which optimizes diversity while taking precision into account. For this purpose a set (combination) of initial clusters called reference set or RefSet is generated. The size of RefSet combination is |RefSet| which denotes the number of its elements. It is worth mentioning that RefSet_{*i*} denotes the *i*th member of this combination. Afterwards, another combination called main combination or combination is generated. It must be mentioned that combination_{*i*} denotes *i*th member of combination. Then, stability is calculated for each combination_{*i*} where *i* changes from 1 to B. The stability of partition combination_{*i*} is the average of its similarity in reference set. The similarity of two partitions is calculated via Fisher measure equation. This measure which is utilized here to assess a partition is called F-measure.

$$FM(P, L) = \max_{\tau} \sum_{i=1}^{K_P} \frac{2 \times N_i^P \times \left(\frac{N_{i\tau(i)}^{PL}}{N_i^P} \times \frac{N_{i\tau(i)}^{PL}}{N_{\tau(i)}^L} \right)}{N \times \left(\frac{N_{i\tau(i)}^{PL}}{N_i^P} + \frac{N_{i\tau(i)}^{PL}}{N_{\tau(i)}^L} \right)}$$

where K_P is the number of clusters in partition P; N_i^P denotes the number of data existing in *i*th cluster of partition P; N_j^L represents the number of data in *j*th cluster of partition L; N_{ij}^{PL} is the number of data which are in both *i*th cluster of P and *j*th cluster of L. N is the total number of data and τ is a permutation of numbers from 1 to N. If partition P and label L are completely similar, FM has its maximum value i.e. 1 and it is zero in case of complete dissimilarity.

The stability of partition combination_{*i*} is derived as follows.

$$Stability(\text{combination}_i) = \frac{1}{|RefSet|} \sum_{j=1}^{|RefSet|} FM(\text{combination}_i, RefSet_j)$$

Then, clustering ensembles are searched with respect to stability and diversity so that the most stable and diverse clustering is found.

When combinations are selected according to their stability, an evolutionary algorithm is employed to select a subset of initial combinations. This algorithm is explicated in the following. These evolutionary algorithms include a bit-string chromosome whose length is the total number of combinations inside second combination. Each genes of this chromosome may acquire 1 or 0. When a gene is 1 it means that the combination with corresponding number is selected and zero means the combination is not selected. In other words a 1 in *i*th gene means that SE_{*i*} is selected while a 0 is *m*th gene means that *m*th combination is not selected. To calculate fitness function of evolutionary algorithm, diversity of selected combinations needs to be derived.

To calculate efficiency of the chromosome, the following equation is used which demonstrates the amount of diversity.

$$FitnessFunction = 0.5 - \frac{\sum_x \sum_y abs(Co(x, y) - 0.5)}{N^2}$$

6 Simulation Results

In this paper combination clustering was investigated from other aspects. Contrary to previous methods which suggest a constant method with constant characteristics for all types of datasets (even if diversity was considered) [26], our proposed method changes its behavior dynamically according to the data set. A dynamic approach changes its behavior with respect to sample distribution in each dataset. As each clustering method has its pros and cons, it is not possible to choose a specific method for a specific data set [1]. Our proposed method tries to choose those initial combinations which lead to the best initial results for a specific data set. K-means algorithm is considered as base clustering algorithm. Besides, in different runs of this algorithm the number of clusters is assumed to be integer. Furthermore, this number is considered as a parameter. The experimental results provided in next section confirm the performance of our method.

Table 1. Results

	NMI		FM		AR	
	Proposed Ensemble	Full Ensemble	Proposed Ensemble	Full Ensemble	Proposed Ensemble	Full Ensemble
Glass	15.82	15.77	28.73	28.17	8.72	8.62
BreastCancer	39.90	35.58	48.39	47.34	44.26	40.69
Wine	21.97	21.44	35.71	35.48	18.58	18.56
Iris	38.91	37.1	46.01	44.61	39.3	35.82

In this section the results of applying the proposed algorithm to some datasets are reported. 4 datasets which are popular in literature are investigated; so, it would possible to compare the proposed method to other methods. The results are reported in Table 1.

7 Conclusion

The proposed method opens up new horizons in clustering algorithms. Including fuzzy concept in clustering combination is a promising idea. Data normalization is

necessary when Euclidean distance is exploited. Since using data normalization algorithm does not guarantee improvement in clustering, proposed clustering methods present their reports according to raw data. Therefore, another idea which might be considered in future studies, is finding a dynamic method for assigning proper normalization method to each data set. The most prominent factor which leads to considerable improvement in the proposed method is finding an intelligent method for generation of initial results. It must be able to generate initial results which cover deficiencies of other initial results.

Resources

1. Azimi J.: Investigating Diversity in combination Clustering, Master thesis, Iran University of Science and Technology (2007)
2. Alizadeh H.: Combination Clustering Based on a Subset of Initial Results. Master thesis, Faculty of Computer Engineering, Iran University of Science and Technology.
3. Jain A., Murty M. N., Flynn P.: Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323 (1999)
4. Faceli K., Marcilio C.P. Souto D.: Multi-objective Clustering combination, *Proceedings of the Sixth International Conference on Hybrid Intelligent Systems* (2006)
5. Strehl A., Ghosh J.: Cluster combination - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(Dec):583–617 (2002)
6. Melanie M.: *An Introduction to Genetic Algorithms*. A Bradford Book The MIT Press, Cambridge, Massachusetts. London, England, Fifth printing (1999)
7. Davis, T.: *The Handbook of Genetic Algorithms*. Van Nostrand Reinhold Co., New York, NY (1991)
8. Muni, D., Pal, N., Das, J.: Genetic programming for simultaneous feature selection and classifier design. *IEEE Trans Syst Man Cybern B Cybern* 36(1):106–117 (2006)
9. Aarts E. H. L., Korst J.: *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, Essex, U.K. (1989)
10. Fred, A., Jain, A. K.: Data Clustering Using Evidence Accumulation. In: *Proc. of the 16th Intl. Conf. on Pattern Recognition, ICPR02, Quebec City*, pp. 276–280 (2002)
11. Parvin H., Alizadeh H., Minaei-Bidgoli B.: A New Method for Constructing Classifier combination. *International Journal of Digital Content: Technology and its Application, JDCTA* (2009)
12. Parvin H., Alizadeh H., Minaei-Bidgoli B.: Using Clustering for Generating Diversity in Classifier combination. *International Journal of Digital Content: Technology and its Application, JDCTA*, Vol. 3, No.1, pp. 51–57 (2009)
13. Alizadeh H., Minaei-Bidgoli B., Amirgholipour S.K.: A New Method for Improving the Performance of K Nearest Neighbor using Clustering Technique. *International Journal of Convergence Information Technology, JCIT* (2009)
14. Topchy, A., Jain, A.K., Punch, W.F.: Combining Multiple Weak Clusterings. In: *Proc. 3d IEEE Intl. Conf. on Data Mining*, pp. 331–338 (2003)
15. Fred A., Lourenco A.: Cluster combination Methods: from Single Clusterings to Combined Solutions. *Studies in Computational Intelligence (SCI)*, 126, 3–30 (2008)

16. Ayad H.G., Kamel M.S.: Cumulative Voting Consensus Method for Partitions with a Variable Number of Clusters. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 30(1):160–173 (2008)
17. Minaei-Bidgoli B., Topchy A., Punch W.F.: Combinations of Partitions via Data Resampling. In: *Proc. Intl. Conf. on Information Technology, ITCC 04, Las Vegas (2004)*
18. Alizadeh H., Amirgholipour S.K., Seyedaghaee N.R., Minaei-Bidgoli B.: Nearest Cluster combination (NCE): Clustering combination Based Approach for Improving the performance of K-Nearest Neighbor Algorithm. In: *11th Conf. of the International Federation of Classification Societies, IFCS09, March 13–18 (2009)*
19. Mohammadi M., Alizadeh H., Minaei-Bidgoli B.: Neural Network combination using Clustering combination and Genetic Algorithm. In: *Intl. Conf. on Convergence and hybrid Information Technology, ICCIT08, Nov. 11-13, IEEE CS (2008)*
20. Barthelemy J.P., Leclerc B.: The median procedure for partition. In: *Partitioning Data Sets, AMS DIMACS Series in Discrete Mathematics, Cox, I. J. et al. (eds.), 19, pp. 3–34 (1995)*
21. Fern, X., Brodley, C. E.: Random Projection for High Dimensional Data Clustering: A Cluster combination Approach. In: *Proc. 20th Int. conf. on Machine Learning, ICML (2003)*
22. Dudoit S., Fridlyand, J.: Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, 19(9):1090–1099 (2003)
23. Fischer B., Buhmann J.M.: Bagging for path-based clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1411–1415 (2003)
24. Fred A., Jain A.K.: Robust data clustering. In: *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, USA, vol. II, pp. 128–136 (2003)*
25. Fred A.L., Jain A.K.: Combining Multiple Clusterings Using Evidence Accumulation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):835–850 (2005)
26. Fred A., Jain A.K.: Learning Pairwise Similarity for Data Clustering. In: *Proc. of the 18th Int. Conf. on Pattern Recognition (ICPR'06) (2006)*
27. Kuncheva L.I., Whitaker C. J.: Measures of diversity in classifier combinations. *Machine Learning (2003)*
28. Kuncheva L.I., Hadjitodorov S.: Using diversity in cluster combinations. In: *Proc. of IEEE Intl. Conference on Systems, Man and Cybernetics*, pp. 1214–1219 (2004)
29. Baumgartner R., Somorjai R., Summers R., Richter W., Ryner L., Jarmasz M.: Resampling as a Cluster Validation Technique in fMRI. *Journal of Magnetic Resonance Imaging* 11:228–231 (2000)
30. Law M.H.C., Topchy A.P., Jain A.K.: Multiobjective data clustering. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pp. 424–430, Washington D.C. (2004)*
31. Shamiry O., Tishby N.: Cluster Stability for Finite Samples. In: *21st Annual Conference on Neural Information Processing Systems (NIPS07) (2007)*
32. Breckenridge J.: Replicating cluster analysis: Method, consistency and validity. *Multivariate Behavioral research* (1989)
33. Fridlyand J., Dudoit S.: Applications of resampling methods to estimate the number of clusters and to improve the accuracy of a clustering method. *Stat. Berkeley Tech Report. No. 600 (2001)*
34. Levine E., Domany E.: Resampling Method for Unsupervised Estimation of Cluster Validity. *Neural Computation* 13:2573–2593 (2001)

35. Roth V., Lange T., Braun M., Buhmann J.: A Resampling Approach to Cluster Validation. In: Intl. Conf. on Computational Statistics, COMPSTAT (2002)
36. Roth V., Braun M.L., Lange T., Buhmann J.M.: Stability-Based Model Order Selection in Clustering with Applications to Gene Expression Data. In: ICANN 2002, LNCS 2415, pp. 607–612 (2002)
37. Lapointe F.J., Legendre P.: The generation of random ultrametric matrices representing dendrograms. *Journal of Classification* 8(2):177–200 (1991)