Spectral indexes of quality, diversity and stability in fuzzy clustering

Stefano Rovetta[1*] and Francesco Masulli[1,2]

1 Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Italy
2 Center for Biotechnology, Temple University, Philadelphia, USA
* Corresponding author – rovetta@disi.unige.it

Abstract. Quality assessment in clustering is a long-standing problem. In this contribution we describe some indexes to measure properties of clusterings, taking advantage of the added flexibility provided by fuzzy paradigms. We first present an approach to evaluate some indicators of quality of an individual clustering, by analyzing the co-association matrix. Then we describe a technique to evaluate the similarity of pairs of clusterings by comparing their respective co-association matrices by means of generalizations of well-known indexes of partition comparisons. Finally, we illustrate how some indexes borrowed from spectral graph theory can be used to evaluate clustering stability and diversity in ensembles of several clusterings.

Possibilistic clustering, Clustering quality, Fuzzy clustering

1 Introduction

Quality assessment in clustering is a long-standing problem, mostly due to the fact that clustering itself is not a well-defined problem.

Clusterings can be analyzed at several levels. An individual clustering can be assessed to provide an evaluation of its quality. Clustering quality measures come in a lot of varieties and flavors, as witnessed by the extensive literature available (see e.g. [18] for a survey). For pairs of clusterings, indexes of similarity based on the co-association matrix can be computed by several approaches [12,9,15]. Finally, when more than two clusterings are to be considered as a whole, pairwise comparisons can be combined to perform analyses on a wider scale [4,17].

When using fuzzy clustering models, it is possible to obtain a deeper insight in the quality and features of a given clustering, pair of clusterings, or clustering ensemble. In this contribution we describe some indexes to measure properties of clusterings, taking advantage of this added flexibility. The approach borrows from spectral graph theory, and is completely independent on the clustering model and on the data dimensionality. After reviewing a family of fuzzy clustering methods (Section 2), we first present an approach to evaluate some indicators of quality of an individual clustering, by analyzing the co-association matrix (Section 3). Then we describe a technique to evaluate the similarity of pairs of clusterings by comparing their respective co-association matrices by means of generalizations of well-known indexes of partition comparisons (Section

4). Finally, we extend the approach to ensembles of fuzzy clusterings, to evaluate clustering stability and diversity (Section 5). Some experimental results are then presented (Section 6) and commented (Section 7).

2 Fuzzy clustering

This section reviews some specific centroid-based fuzzy clustering methods, on which we based our present work. However, the techniques presented here do not use any model-specific information, and they are generally applicable to any clustering method.

In general terms, a clustering is a set $\mathcal{C} = \{X, Y, U\}$ where X is a set of n data items, Y is a set of c fuzzy sets (clusters), with $c < n$ (usually $c \ll n$), and U is the membership matrix relating X to Y, regarded as a set of n c-dimensional vectors u_l.

2.1 Centroid-based fuzzy clustering

In the case of centroid-based fuzzy methods, X and Y are sets of d-vectors, and

$$\mathbf{y}_j = \frac{\sum_{l=1}^{n} u_{lj}\mathbf{x}_l}{\sum_{m=1}^{n} u_{mj}}. \tag{1}$$

Since the memberships u_{lj} may be computed in different ways, the resulting centroids are not necessarily the same for different methods. However, all methods considered here use the following generalized equation:

$$u_{lj} = v_{lj}/Z_l, \tag{2}$$

where $v_{lj} = e^{-||x_l - y_j||/\beta}$, and Z_l characterizes a given method.

By applying an entropic penalty, in the Maximum Entropy (ME) approach [14] memberships are given by Eq. 2 with $Z_l = \sum_{l=1}^{c} e^{u_{lj}/\beta}$. These membership obey the so-called "probabilistic constraint", $\sum_{j=1}^{c} u_{lj} = 1 \; \forall l \in \{1, \ldots, n\}$.

The Possibilistic c-Means (PCM) [7,8] is based on removing any equality constraint on the sum of memberships. This is replaced by a set of inequalities:

$$u_{lj} \in [0,1] \;\; \forall l \; \forall j \qquad 0 < \sum_{l=1}^{n} u_{lj} < n \forall j \qquad \forall l \; \exists j : u_{lj} > 0 \tag{3}$$

which imply that no cluster be empty and that each pattern be assigned to at least one cluster. The interpretation of memberships is as "typicalities". We focus on the second formulation of the method [8], where the entropy of clusters is also considered, yielding:

$$u_{lj} = e^{-u_{lj}/\beta_j}. \tag{4}$$

The parameters β_j can be individually set for each cluster or considered as a global property, as in the ME approach; in this case we can set a single value β for all the β_j. This corresponds to Eq. 2 if $Z_l = 1 \; \forall l$.

2.2 The graded possibilistic model and clustering method

The graded possibilistic membership model presented in [11] assumes that events may be independent to a certain degree, but not completely. This "soft transition" between the probabilistic and the possibilistic cases is under the user's control.

Let $[\xi]$ be an interval variable defined as $[\,\underline{\xi}, \overline{\xi}\,]$. Memberships are subject to:

$$\sum_{j=1}^{c} u_{lj}^{[\xi]} = 1 \;\forall l. \tag{5}$$

This interval equality is satisfied for any point u_l such that there exist a scalar number $\xi^* \in [\,\underline{\xi}, \overline{\xi}\,]$ s.t. $\sum_{j=1}^{c} u_{lj}^{\xi^*} = 1$. Eq. 5 is equivalent to a set of two inequalities:

$$\sum_{j=1}^{c} u_{lj}^{\underline{\xi}} \geq 1 \qquad\qquad \sum_{j=1}^{c} u_{lj}^{\overline{\xi}} \leq 1.$$

This formulation includes as the two extreme cases the "probabilistic" assumption of ME: $[\xi] = [1,1] = 1$, $\sum_{j=1}^{c} u_{lj} = 1$, and the "possibilistic" assumption of PCM: $[\xi] = (0,\infty)$, $\sum_{j=1}^{c} u_{lj}^{\underline{\xi}\to 0} > 1$, $\sum_{j=1}^{c} u_{lj}^{\overline{\xi}\to\infty} < 1$, in the limit.

Several choices are possible for $\underline{\xi}$ and $\overline{\xi}$. The specific implementation that we adopt in this work is: $\underline{\xi} = \alpha$ and $\overline{\xi} = 1$, where $\alpha \in (0,1]$ controls the "possibility level". In this asymmetric implementation (AGPCM, *Asymmetric Graded Possibilistic c-Means*), memberships whose sum exceeds 1 are forbidden. Therefore clustering is effectively competitive among nearby centroids. However, for far-away centroids, the competition decreases with α. The simplest case where $[\xi] = [\,0,1\,]$ allows points with (almost) zero memberships to any cluster, and can be used to effectively implement automatic outlier rejection. Z_l is defined as:

$$\begin{aligned} Z_l &= \textstyle\sum_{j=1}^{c} v_{lj} && \text{if} \quad \textstyle\sum_{j=1}^{c} v_{lj} > 1 \\ Z_l &= \left(\textstyle\sum_{j=1}^{c} v_{lj}^{\alpha}\right)^{1/\alpha} && \text{if} \quad \textstyle\sum_{j=1}^{c} v_{lj}^{\alpha} < 1 \\ Z_l &= 1 && \text{else.} \end{aligned} \tag{6}$$

In [11] this particular model has been shown to possess robustness and outlier rejection properties, However, no criterion was given to set the values of α and β.

3 Evaluating fuzzy clustering

3.1 Co-association matrices for fuzzy clustering

Similarity of strings of crisp memberships can be measured by Hamming distance, the sum of the bits of the bitwise-AND between two words. To generalize this operation to the fuzzy domain we have to define the fuzzy conjunction connective AND. We adopt the product t-norm [13]: Given two fuzzy memberships/truth values μ_1 and μ_2,

the conjunction logical connective is simply defined as $\mu_1 \,\mathrm{AND}\, \mu_2 = \mu_1\mu_2$. The co-association of a given pair of data points to a given cluster $y_j \in \mathcal{C}$ is the conjunction of the respective point memberships to y_j, and the degree of similarity of two points is the average of these values. Therefore, for one clustering $\mathcal{C}$, the *co-association matrix* [4] (also termed *bonding relationship* in [1]) is the $n \times n$ Gram matrix S with the usual dot product in the space of membership vectors. It can be obtained as

$$S^{(l)} = U^{(l)}U^{(l)\prime}. \tag{7}$$

Note that in the crisp case this definition collapses to the proposition "in clustering $\mathcal{C}$, data items x_l and x_m are in the same cluster", by resulting in either 1 (for true) or 0 (for false); in the fuzzy case it is necessary to take all clusters into considerations because, in general, none of them will be exactly zero or one.

3.2 Analysis of the co-association matrix

We will use two indexes computed on the co-association matrix. The first is simply the *normalized trace of S*

$$\nu = \frac{1}{n}\,\mathrm{tr}(S) \tag{8}$$

Depending on the type of memberships there are different kinds of clustering models, and different resulting values for co-association. The co-association is maximum for a pair of data items clustered in exactly the same way by the given clustering. If $u_{lj} = u_{mj} \forall j$, then $S_{lm} = \sum_j u_{lj}u_{mj} = \sum_j u_j^2$ if we assume $u_{lj} = u_{mj} = u_j \;\forall j$, which is always true for the self-association S_{ll}. Therefore we have to study $\sum u_j^2$ as a function of the possible constraints on $\sum u_j$.

We can describe four cases, although they are really a continuum.

- Crisp clustering: $u_j \in \{0, 1\}\; \forall j, \quad \sum_j u_j = 1, \quad \sum_j u_j^2 = 1$. For crisp clustering, one element is 1, all others are 0, so co-association is 1 for equal memberships, always 0 for different memberships. These are the limiting values for all cases.
- Fuzzy "probabilistic": $u_j \in [0, 1]\; \forall j, \quad \sum_j u_j = 1, \quad \sum_j u_j^2 \leq 1$. Crisp clustering is a special, limit case of fuzzy "probabilistic" clustering. For equal memberships, co-association is 1 only in the crisp case (triangle inequality). If there are nonzero membership values for more than one cluster, the self-association value can be as low as $1/c$.
- Fuzzy possibilistic: $u_j \in [0, 1]\; \forall j, \quad \sum_j u_j < c, \quad \sum_j u_j^2 < c, \quad \sum_j u_j^2 > 0$. Probabilistic clustering is in turn a special case of possibilistic clustering. Here the self-association value can be as high as c or as low as $1/c$.
- The asymmetric graded possibilistic model: $u_j \in [0, 1]\; \forall j, \quad \sum_j u_j \leq 1, \quad \sum_j u_j^2 \leq 1, \quad \sum_j u_j^2 > 0$, In this case, the self-association value can as low as $1/c$, but it cannot exceed 1.

In the crisp case, ν gives no information. In all fuzzy cases, however, it provides an indication on the degree of fuzziness of the clustering under examination, a confidence evaluation: if $\nu \ll 1$ then most points have mutually equivalent memberships to all clusters, indicating a bad clustering. In the possibilistic case, ν may also exceed unity: then most points have high memberships to all clusters, an effect of too many overlapping clusters. With AGPCM, $\nu \leq 1$ always, and a strong clustering has $\nu \approx 1$.

3.3 A spectral criterion

The second index proposed in this work is an eigengap statistic borrowed from spectral graph theory [2,10]. It is well known that a pairwise proximity or similarity matrix can be interpreted as the weighted adjacency matrix of a graph, so that the tools of spectral graph theory can be used. Especially in the case of fuzzy clustering, the same approach can be used to analyze both individual clusterings and sets of clusterings. We are now focusing on the first case.

The co-association matrix S is a type of similarity measure, [17], a Gram matrix where similarity is provided by the fact that a given pair of data item belongs to the same cluster in a given clustering. If we view it as the adjacency matrix of a graph having all available n data items as vertices and all possible $n \times n$ clustering similarities as edge weights, then the degree matrix is the diagonal, $n \times n$ matrix D whose elements are $D_{hh} = \sum_{l=1}^{n} S_{lm}$. The Laplacian matrix of the data is $L = D - S$.

From spectral graph theory it is known [3,2] that the spectrum of a graph provides information about its connected components. In particular, the multiplicity of eigenvalue 0 is the number of connected components in the graph. Therefore, it is possible to apply an *eigengap criterion* to evaluate the degree of separatedness in a clustering. Given the n ordered eigenvalues λ_l of $\mathrm{diag}(\sum_{l=1}^{n} S_{lm}) - S$, compute the following eigengap index κ:

$$\kappa = \frac{1}{Q} \sum_{h=1}^{N-1} h \cdot e^{(\lambda_{h+1} - \lambda_h)/\gamma} \tag{9}$$

with $Q = \lambda_N \sum_{k=1}^{N-1} e^{(\lambda_{k+1} - \lambda_k)/\gamma}$.This is a "softmax-type" index. The maximum summand correspond to the maximum eigengap. The value of the index is exactly the index of the eigengap for a perfect (crisp) clustering, but may differ for less strong partitions. The gain term γ can be used to tune the sensitivity of the index; when it is low, the index simply points out the location of the maximum eigengap; when it is raised, the relative amplitude of the gap also plays a role, so that the index is also sensitive to the degree of structure of the clustering. In our experiments, we set $\gamma = 2$.

4 Comparing two fuzzy clusterings

4.1 Comparing partitions

In many cases, a clustering induces a partition of the data, so measuring the agreement or overlap between two clusterings $\mathcal{C}^{(h)}$ and $\mathcal{C}^{(k)}$ amounts to measuring the similarity between two partitions. There are several partition similarities available in the literature. The two main approaches include comparing matching clusters, and comparing co-association information. In [15] we concentrated on the second approach, defining a set of fuzzy pairwise indexes. In this work we focus on one of the most popular, the Jaccard index [6], in the generalized fuzzy version.

4.2 The fuzzy Jaccard similarity

The Jaccard index is defined as $\mathcal{J}(\mathcal{C}^{(h)}, \mathcal{C}^{(k)}) = |\mathcal{C}^{(h)} \cap \mathcal{C}^{(k)}| / |\mathcal{C}^{(h)} \cup \mathcal{C}^{(k)}|$.

This expression can be written in terms of S. This allows us to obtain a fuzzy version, by using a fuzzy co-association matrix. For the product t-norm we have as the associated disjunction operator the *probabilistic sum* t-conorm, so that $\mu_1 \,\mathrm{OR}\, \mu_2 = \mu_1 + \mu_2 - \mu_1\mu_2$. For a pair of clusterings $\mathcal{C}^{(h)}, \mathcal{C}^{(k)}$, the Jaccard similarity index is

$$\mathcal{J}\left(\mathcal{C}^{(h)}, \mathcal{C}^{(k)}\right) = \frac{\langle S^{(h)}, S^{(k)} \rangle_{\mathcal{F}}}{\|S^{(h)} + S^{(k)}\|_1 - \langle S^{(h)}, S^{(k)} \rangle_{\mathcal{F}}}, \tag{10}$$

where $\langle M, N \rangle_{\mathcal{F}}$ denotes the Frobenius inner product between the two matrices M and N of equal dimensions and $|M|_1$ indicates the entry-wise 1-norm (sum of entries) of matrix M. The Jaccard distance between $\mathcal{C}^{(h)}$ and $\mathcal{C}^{(k)}$ is $\mathcal{L} = 1 - \mathcal{J}\left(\mathcal{C}^{(h)}, \mathcal{C}^{(k)}\right)$.

4.3 Clusterings that are not partitions

A remark is in order regarding clustering techniques that are not (strictly) partitional. For instance, PCM clusters are overlapping, so that the resulting clusterings are not fuzzy partitions. In general, the main effect of this possibility is to yield values of similarity exceeding unity. In our specific case, we only concentrated on the modified, asymmetric graded possibilistic model, for which membership may sum up to less than unity, *but not more*. The effect of this choice is that we still have a maximum value of unity for similarity indexes (in this work $\mathcal{J}$, but the same would hold for other options), and, in the case of self-similarity, the difference can be used as an estimation of the level of fuzziness of a set of memberships.

5 Evaluating fuzzy clustering ensembles

In this section we propose a criterion to analyze the matrix H defined as:

$$H_{hk} = \mathcal{J}\left(S^h, S^k\right) \tag{11}$$

where $S^{(h)}$ and $S^{(k)}$ refer to $\mathcal{C}^{(h)}$ and $\mathcal{C}^{(k)}$, two clusterings from a collection of N.

5.1 Spectral indexes for clustering ensembles

Having introducing $\mathcal{J}$ as a measure of clustering similarity [15], we can apply the same arguments as in Section 3 to matrices measuring proximities between clusterings [16]. This approach provides tools to analyze the effect of varying a parameter, as for instance the number of centroids, to evaluate the stability of a set clustering, and to measure the diversity within a clustering ensemble, both as a global property, and on a local basis, to find families of mutually similar models which are distinct from other families.

For a set of clusterings $\mathcal{C}^{(h)}$, $h \in \{1, \ldots, N\}$, compute the affinity matrix H (again a Gram matrix, from the Mercer kernel $\mathcal{J}$ on the space of co-association matrices):

$$H_{hk} = \mathcal{J}\left(\mathcal{C}^{(h)}, \mathcal{C}^{(k)}\right) \tag{12}$$

for all possible pairs of clusterings $\mathcal{C}^{(h)}, \mathcal{C}^{(k)}$.

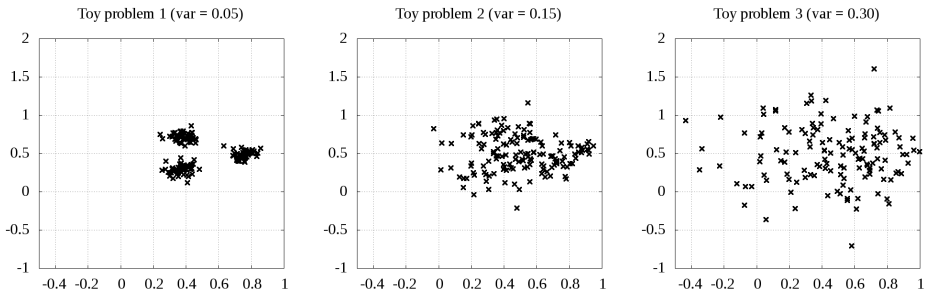

Fig. 1. Toy problems: Gaussian clusters with variance = 0.05 (left), 0.15 (center), 0.30 (right).

The eigengap index can now be computed for H, which is the adjacency matrix of a complete graph with all N clusterings as vertices and all $N \times N$ clustering similarities (measured by $\mathcal{J}$) as weights; the degree matrix is the $N \times N$ matrix Δ with $\Delta_{hh} = \sum_{k=1}^{N} H_{hk}$. The Laplacian matrix is $L = \Delta - H$ and κ is computed as in Section 3.

5.2 Parameter or model selection

One specific problem in clustering, due to the unavailability of prior information or a golden standard, is the ambiguity in the selection of parameters. A typical example is the selection of c. In [16] a visual procedure is described. The clustering methods used in this study also depend on other parameters in addition to c, namely, β and α. The latter is specific to (A)GPCM. It allows the user to introduce a bias into the representation of cluster; however, it may also cause convergence problems. Therefore an objective evaluation of the resulting effects is useful.

In all instances where a "sweep" of a given parameter is possible, the analysis may provide a quality indication to select the best value. Here we summarize two important goals: model/parameter selection and diversity analysis for ensemble clustering.

Stability analysis is often used as a model selection tool. The procedures described implement a stability analysis by inspection of the properties of matrix H, and differ by the type of parameter. For parameters with a smooth influence on the clustering results, a local analysis can be performed. This is the case for α and β. One possible technique for local analysis, based on visualization, is presented in [16]. The maximum stability is attained around the optimal value, thus allowing visual inspection. However, when a parameter has a complex influence on the clustering result (e.g., c), a visual analysis is not possible. In this case we use the eigengap index κ to point out the hidden structure.

Conversely, when the application is centered on ensemble clustering, a similar procedure can search for the *maximum* diversity, again as measured by κ.

6 Experiments

6.1 Experimental setup

With the aim to verify the behaviour of the proposed indexes, we performed experiments on two datasets. The first is a set of three toy problems (Fig. 1). In our case, we are not

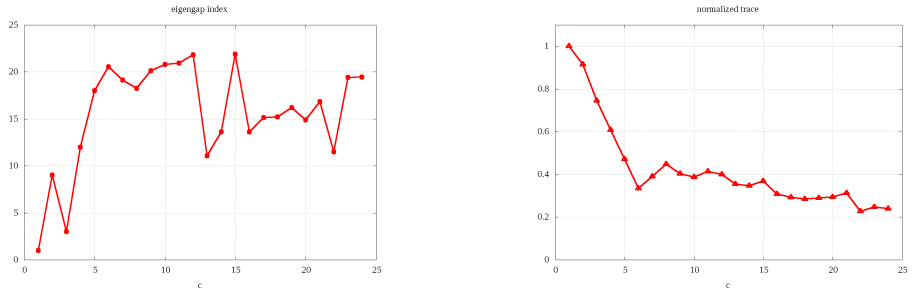

Fig. 2. κ and ν for individual clusterings with varying c for the Toy 1 dataset.

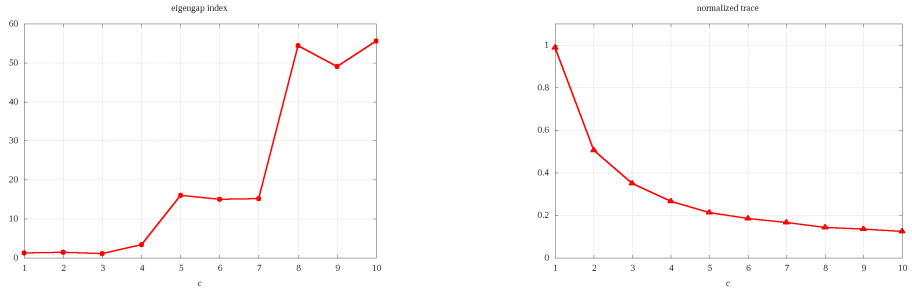

Fig. 3. κ and ν for individual clusterings with varying c for the Leukaemia dataset.

really interested in challenging a clustering algorithm. We need to degrade the resulting clusterings in a controlled way, to point out the relationship between the features of the problem and the corresponding values of ν and κ. We also tested the indexes on the Leukaemia data set from [5], a high-dimensional, real world problem.

The toy problems are generated by spreading points around three equi-spaced centroids (50 2-dimensional points per cluster, so $n = 150$) with three different values of spread: variance = 0.05 (trivial, well separated problem), variance = 0.15 (overlapped, non-separable clusters), and variance = 0.3 (structure completely lost).

For the Leukaemia problem we used the training set with $n = 38$ points in $d = 7129$ dimensions. It is known that these points can be clustered with $c = 3$ if only the 50 most relevant features are retained. Here instead we used the whole feature set.

The clustering method used was AGPCM with β and α selected appropriately for the problem at hand. Alternatively, selection of these parameters could also be performed automatically with the same procedure presented for c. In general, we kept the degree of possibility at a moderate level, with $\alpha = 0.8$ to 0.9.

6.2 Results and comments

The first experiment uses Toy 1 (the trivial dataset) to illustrate the use of ν and κ for the analysis of individual clusterings at various values of c. The experiments included

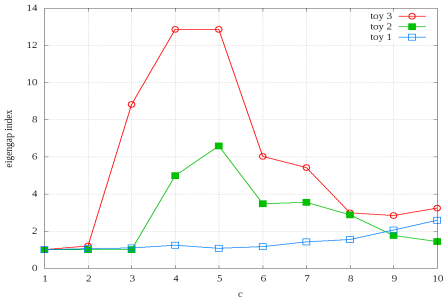

Fig. 4. Eigengap index for ensembles of 100 clusterings with varying c for the three toy problems.

10 random initializations, and the values presented in Fig. 2 are the averaged indexes. It is expected that for $c = 3$ all clusterings (except for pathological random initializations) reach a maximum for ν and a minimum for κ, indicating a strong clustering and a low variability. In fact, the optimal values are found for $c = 1$, where there is only one solution (the barycenter of the data). However, the first significant minimum for κ is for $c = 3$ as expected, where $\kappa = 3$, and for this value ν is quite high (about 0.93), indicating that on average the clusterings are very strong (apart from occasional local minima). There is another minimum for $c = 13$, but this is likely due to random variability: with growing c, the number of minima grows combinatorially, and just 10 random initializations are probably not enough. To validate this hypothesis, it is sufficient to check the value of ν, which is about 0.38, indicating weak clusterings.

A similar experiment was performed for the Leukaemia dataset. Here, of course, the results are much less clear (Fig. 3). However, there is a minimum of $\kappa = 1.11$ for $c = 3$, which is the value reported in [5]. The value of the eigengap index, however, is not 3 as expected, but only slightly larger than 1: this indicates that the clusters are not well defined in the native 7129-dimensional space. The value of ν is correspondingly low ($\nu = 0.35$), indicating a not-so-strong clustering result.

Finally, multiple clusterings were obtained by 100 random initializations, for c ranging from 1 to 10, for the three toy problems. The corresponding ensemble eigengap index κ is traced in Fig. 4. This experiment outlines the stability of clustering while varying a parameter (here c). The eigengap index, a measure of diversity, should be high when looking for varied clusterings to be used as ensemble members, while it should be low when searching for a good parameter value.

From the graph it is clear that the more difficult the problem, the higher the diversity. Only for the trivial problem (Toy 1) $\kappa = 1$ in almost every case. For Toy 2, spurious solutions are found as soon as $c > 3$, and for Toy 3, where the lack of structure probably enhances the dependence on the initial conditions, this happens even for $c > 2$. It is worth noticing that the behaviour of κ is not monotonic with c. This is probably due to the inherent ambiguity in the clustering problem, where different soultions may be found at different scales with comparable representation cost.

7 Remarks and future work

The proposed method is notably flexible, since it is agnostic with respect to cluster structure and representation. Other clustering techniques may be treated similarly to the centroid-based techniques used here. It may be integrated with a visual analysis, or embedded in an almost completely automatic clustering tool.

References

1. Brouwer, R.K.: Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions. J. Intell. Inf. Syst. 32(3), 213-235 (2009)
2. Chung, F.R.K.: Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92), American Mathematical Society (1997)
3. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Mathematical Journal 23(2), 298-305 (1973)
4. Fred, A.L.N., Jain, A.K.: Data clustering using evidence accumulation. Pattern Recognition, International Conference on 4 (2002)
5. Golub, T., et al.: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286(5439), 531-537 (1999)
6. Jaccard, P.: Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 547-579 (1901)
7. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems 1(2), 98-110 (1993)
8. Krishnapuram, R., Keller, J.M.: The possibilistic C-Means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems 4(3), 385-393 (1996)
9. Kuncheva, L.I., Vetrov, D.P.: Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Trans. PAMI 28(11), 1798-1808 (2006)
10. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. Tech. Rep. 134, Max Planck Institute for Biological Cybernetics (2004)
11. Masulli, F., Rovetta, S.: Soft transition from probabilistic to possibilistic fuzzy clustering. IEEE Transactions on Fuzzy Systems 14(4), 516-527 (2006)
12. Meilă, M.: Comparing clusterings–an information based distance. Journal of Multivariate Analysis 98(5), 873-895 (May 2007)
13. Menger, K.: Statistical metrics. Proceedings of the National Academy of Sciences of the United States of America 28(12), 535-537 (December 1942)
14. Rose, K., Gurewitz, E., Fox, G.: Statistical mechanics and phase transitions in clustering. Physical Review Letters 65, 945-948 (1990)
15. Rovetta, S., Masulli, F.: An experimental validation of some indexes of fuzzy clustering similarity. In: Gesù, V.D., Pal, S.K., Petrosino, A. (eds.) Fuzzy Logic and Applications, Italy, June 9-12, 2009. Lecture Notes in Computer Science, vol. 5571, pp. 132-139. Springer (2009)
16. Rovetta, S., Masulli, F., Adel, T.: Tuning graded possibilistic clustering by visual stability analysis. In: Gesù, V.D., Pal, S.K., Petrosino, A. (eds.) Fuzzy Logic and Applications, Italy, August 29-31, 2011, Proceedings. Lecture Notes in Computer Science, Springer, to appear
17. Strehl, A., Ghosh, J.: Cluster ensembles – a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3, 583-617 (2003)
18. Xu, R., Wunsch, D., I.: Survey of clustering algorithms. Neural Networks, IEEE Transactions on 16(3), 645-678 (2005)