Developing Additive Spectral Approach to Fuzzy Clustering

Boris Mirkin[1,2] and Susana Nascimento[3]

[1] Department of Computer Science, Birkbeck University of London, London, UK
[2] School of Applied Mathematics and Informatics,
Higher School of Economics, Moscow, RF
[3] Department of Computer Science and Centre for Artificial Intelligence
(CENTRIA), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa,
Caparica, Portugal

Abstract. An additive spectral method for fuzzy clustering is presented. The method operates on a clustering model which is an extension of the spectral decomposition of a square matrix. The computation proceeds by extracting clusters one by one, which allows us to draw several stopping rules to the procedure. We experimentally test the performance of our method and show its competitiveness.

In spite of the fact that many relational fuzzy clustering algorithms have been developed already [1,2,3,4,12], most of them are ad hoc and, moreover, they all involve manually specified parameters such as the number of clusters or threshold of similarity without providing any guidance for choosing them. We apply a model-based approach of additive clustering, combined with the spectral clustering approach, to develop a novel relational fuzzy clustering method that is both adequate and supplied with model-based parameters helping to choose the right number of clusters.

We assume the data in the format of what is called similarity or relational data, that is a matrix $W = (w_{tt'}), t, t' \in T$, of similarity indexes $w_{tt'}$, between objects t, t' from a set of objects T. We further assume that this similarity values are but manifested expressions of some hidden relational patterns which can be represented by fuzzy clusters. We propose to formalize a relational fuzzy cluster as represented by two items: (i) a membership vector $\mathbf{u} = (u_t)$, $t \in T$, such that $0 \leq u_t \leq 1$ for all $t \in T$, and (ii) an intensity $\mu > 0$ that expresses the extent of significance of the pattern corresponding to the cluster. With the introduction of the intensity, applied as a scaling factor to $\mathbf{u}$, it is the product $\mu\mathbf{u}$ that is a solution rather than its individual co-factors. Given a value of the product μu_t, it is impossible to tell which part of it is μ and which u_t. To resolve this, we follow a conventional scheme: let us constrain the scale of the membership vector $\mathbf{u}$ on a constant level, for example, by a condition such as $\sum_t u_t = 1$ or $\sum_t u_t^2 = 1$, then the remaining factor will define the value of μ. The latter normalization better suits the criterion implied by our fuzzy clustering method and, thus, is accepted further on.

S.O. Kuznetsov et al. (Eds.): RSFDGrC 2011, LNAI 6743, pp. 273–277, 2011.

To make the cluster structure in the similarity matrix sharper, we apply the spectral clustering approach to pre-process a raw similarity matrix W into A by using the so-called normalized Laplacian transformation as related to the popular clustering criterion of normalized cut [6]. This criterion relates to the minimum non-zero eigenvalue of the Laplacian matrix. To change this to the maximum eigenvalue, we further transform this to its pseudo-inverse matrix, which also increases the gaps between eigenvalues.

Our additive fuzzy clustering model follows that of [11,7,10] and involves K fuzzy clusters that reproduce the pseudo-inverted Laplacian similarities $a_{tt'}$ up to additive errors according to the following equations:

$$a_{tt'} = \sum_{k=1}^{K} \mu_k^2 u_{kt} u_{kt'} + e_{tt'}, \tag{1}$$

where $\mathbf{u}_k = (u_{kt})$ is the membership vector of cluster k, and μ_k its intensity.

The item $\mu_k^2 u_{kt} u_{kt'}$ is the product of $\mu_k u_{kt}$ and $\mu_k u_{kt'}$ expressing participation of t and t', respectively, in cluster k. This value adds up to the others to form the similarity $a_{tt'}$ between topics t and t'. The value μ_k^2 summarizes the contribution of the intensity and will be referred to as the cluster's weight.

To fit the model in (1), we apply the least-squares approach, thus minimizing the sum of all $e_{tt'}^2$. Within that, we attend to the one-by-one principal component analysis strategy for finding one cluster at a time by minimizing

$$E = \sum_{t,t' \in T} (b_{tt'} - \xi u_t u_{t'})^2 \tag{2}$$

with respect to unknown positive ξ weight (so that the intensity μ is the square root of ξ) and fuzzy membership vector $\mathbf{u} = (u_t)$, given similarity matrix $B = (b_{tt'})$.

At the first step, B is taken to be equal to A. After each step, the found cluster is subtracted from B, so that the residual similarity matrix for obtaining the next cluster is equal to $B - \mu^2 \mathbf{u}\mathbf{u}'$ where μ and $\mathbf{u}$ are the intensity and membership vector of the found cluster. In this way, A indeed is additively decomposed according to formula (1) and the number of clusters K can be determined in the process.

The optimal value of ξ at a given $\mathbf{u}$ is proven to be

$$\xi = \frac{\mathbf{u}' B \mathbf{u}}{(\mathbf{u}'\mathbf{u})^2} \tag{3}$$

which is obviously non-negative if B is semi-positive definite.

By putting this ξ in equation (2), we arrive at $E = S(B) - \xi^2 (\mathbf{u}'\mathbf{u})^2$, where $S(B) = \sum_{t,t' \in T} b_{tt'}^2$ is the similarity data scatter.

Let us denote the last item by

$$G(\mathbf{u}) = \xi^2 (\mathbf{u}'\mathbf{u})^2 = \left(\frac{\mathbf{u}' B \mathbf{u}}{\mathbf{u}'\mathbf{u}} \right)^2, \tag{4}$$

so that the similarity data scatter is the sum $S(B) = G(\mathbf{u}) + E$ of two parts, $G(\mathbf{u})$ explained by cluster $(\mu, \mathbf{u})$, and E, unexplained. Therefore, an optimal cluster is to maximize the explained part $G(\mathbf{u})$ in (4) or its square root

$$g(\mathbf{u}) = \xi \mathbf{u}'\mathbf{u} = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'\mathbf{u}}, \tag{5}$$

which is the celebrated Rayleigh quotient: its maximum value is the maximum eigenvalue of matrix B, which is reached at its corresponding eigenvector, in the unconstrained problem.

This shows that the spectral clustering approach is appropriate for our problem. According to this approach, one should find the maximum eigenvalue λ and corresponding normed eigenvector z for B, $[\lambda, z] = \Lambda(B)$, and take its projection to the set of admissible fuzzy membership vectors.

Our clustering approach involves a number of model-based criteria for halting the process of sequential extraction of fuzzy clusters. The process stops if either is true:

1. The optimal value of ξ (3) for the spectral fuzzy cluster becomes negative.
2. The contribution of a single extracted cluster to the data scatter becomes too low, less than a pre-specified $\tau > 0$ value.
3. The residual data scatter becomes smaller than a pre-specified ϵ value, say less than 5% of the original similarity data scatter.

The described one-by-one Fuzzy ADDItive-Spectral cluster extraction method is referred to as FADDIS. We have experimentally compared FADDIS with other approaches, specifically, with those used at (a) ordinary graphs for revealing community structure, (b) affinity similarity data derived from feature based information, (c) small real-world benchmark dissimilarity datasets, and (d) genuine similarity data [8]. In this paper we describe one of the experiments - in comparing the performance of FADDIS with various versions of fuzzy c-means algorithm [1]. In this, we carry on the experiment described in [2]. This experiment concerns a two-dimensional data set, that we refer to as Bivariate4, comprising four clusters generated from bivariate spherical normal distributions with the same standard deviation 950 at centers (1000, 1000), (1000,4000), (4000, 1000), and (4000, 4000), respectively (see Fig. 1).

This data was analyzed in [2] by using the matrix D of Euclidean distances between the generated points. Five different fuzzy clustering methods have been compared, three of them relational, by Roubens [9], Windham [12] and NERFCFM [4], and two of the fuzzy c-means with different preliminary pre-processing options of the similarity data into the entity-to-feature format, FastMap and SMACOF [2]. Of these five different fuzzy clustering methods, by far the best results have been obtained with the fuzzy c-means method applied to a five-feature set extracted from D with FastMap method [2]. The adjusted Rand index [5] of the correspondence between the generated clusters and those the best is equal on average, of 10 trials, 0.67 (no standard deviation is reported in [2]).

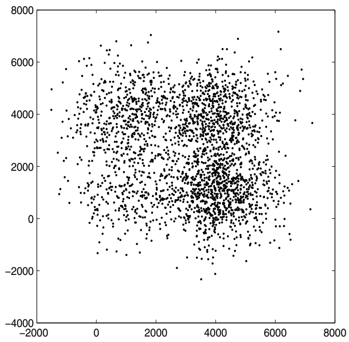

Fig. 1. Bivariate4: the data of four Gaussian bivariate clusters [2]

In our computations, five consecutive FADDIS clusters have been extracted for each of randomly generated ten Bivariate4 datasets. The algorithm halts at stop condition (2): 'cluster's contribution is too small'. Then the very first approximate cluster is discarded as reflecting just the general connectivity, and the remaining four are defuzzified into partitions so that every entity is assigned to its maximum membership class. The average values of the adjusted Rand index (ARI) in these experiments are 0.70 (0.03) at 500-and 1000-strong datasets, whereas ARI=0.73 (0.01) at 2500-strong generated datasets (the standard deviations are reported in the parentheses). This favorably compares with ARI value of 0.67 reported in [2] as the best achieved with fuzzy c-means.

Acknowledgments. This work has been supported by grant PTDC/EIA/69988/2006 from the Portuguese Foundation for Science & Technology. The support of the Laboratory for Analysis and Choice of Decisions at the National Research University Higher School of Economics, Moscow RF to BM is acknowledged.

References

1. Bezdek, J., Keller, J., Krishnapuram, R., Pal, T.: Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Kluwer Academic Publishers, Dordrecht (1999)
2. Brouwer, R.: A method of relational fuzzy clustering based on producing feature vectors using FastMap. Information Sciences 179, 3561–3582 (2009)
3. Davé, R., Sen, S.: Robust fuzzy clustering of relational data. IEEE Transactions on Fuzzy Systems 10, 713–727 (2002)
4. Hathaway, R.J., Bezdek, J.C.: NERF c-means: Non-Euclidean relational fuzzy clustering. Pattern Recognition 27, 429–437 (1994)
5. Hubert, L.J., Arabie, P.: Comparing partitions. Journal of Classification 2, 193–218 (1985)

6. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17, 395–416 (2007)
7. Mirkin, B.: Additive clustering and qualitative factor analysis methods for similarity matrices. Journal of Classification 4(1), 7–31 (1987)
8. Mirkin, B., Nascimento, S.: Analysis of Community Structure, Affinity Data and Research Activities using Additive Fuzzy Spectral Clustering. Technical Report 6, School of Computer Science, Birkbeck University of London (2009)
9. Roubens, M.: Pattern classification problems and fuzzy sets. Fuzzy Sets and Systems 1, 239–253 (1978)
10. Sato, M., Sato, Y., Jain, L.C.: Fuzzy Clustering Models and Applications. Physica-Verlag, Heidelberg (1997)
11. Shepard, R.N., Arabie, P.: Additive clustering: representation of similarities as combinations of overlapping properties. Psychological Review 86, 87–123 (1979)
12. Windham, M.P.: Numerical classification of proximity data with assignment measures. Journal of Classification 2, 157–172 (1985)