Affinity matrix selection for spectral clustering

D. Peluffo, D. F. López, J. L. Rodríguez, G. Castellanos

Abstract—**Category 2. This work presents a comparative study of different affinity measures applied in spectral clustering algorithms commonly used and recommended by literature. Spectral algorithms for clustering are known to be useful in different areas of science and engineering such as image segmentation, load balance for intensive computing, circuits design, among others. In particular they have been proved to be effective in classification tasks where the classes are not linearly separable. Typically, spectral clustering involves calculating an affinity matrix from the data to be processed, which usually can be done by different standard approaches or adjusted to the requirements for the specific algorithm or classification task. Experiments results are carried out using an image database and clustering performance is measured by means supervised and unsupervised indices based on cluster coherence and objective function value. Results show a general idea of the effects of the affinity matrix selection in grouping process in terms of convergence time and accuracy.**

Keywords—**Affinity matrix, clustering, graphs, spectral analysis.**

I. Introduction

In the area of pattern recognition and classification, clustering methods based on graphs and spectral analysis are relatively new but have been applied successfully on several science fields. Such as circuits design [1], computational load balance for parallel applications [2], image segmentation [3], among others. These methods are discriminative, i.e., do not require prior information (assumptions about the nature of data) and the partitions are generated taking advantage of the information obtained from its own method heuristic. In this case, a global decision criterion that takes into consideration the estimated value of the probability of two elements are belonging to the same cluster [4]. Therefore, this kind of analysis can be easily studied from graph theory, where such probability corresponds to the *affinity* measure between a pair of nodes. Commonly, this criterion is applied on a new representation space whose dimension is less than that original data and then a dimensionality reduction procedure is accomplished, in that way the relation among elements are conserved as well as possible. Thus, eigenvectors and eigenvalues based analysis takes place.

Spectral methods are of great interest for classification tasks, mainly, in case of classes are not linearly separable. In particular, spectral clustering uses the information given by eigenspace (i.e, space generated by eigenvectors) because of eigenvectors are directly associated with the quality of clustering. Typically, clustering methods, due to their discriminative nature, require some prior initial parameters. In case of spectral clustering, it is necessary to establish a prior number of groups and the affinity matrix. There are several alternatives to choose the affinity measure and the selection of one of them is not a trivial task, but represents an important issue in spectral analysis.

In this work, a comparative study of different affinity measures is presented. For this end, it is applied a clustering algorithm commonly used and recommended by literature: normalized cuts based clustering [4]. Three different kind of affinity matrices: trivial, exponential and scaled exponential. Additionally, an estimation of number of groups is carried out. Experiments results are obtained over an image database corresponding to coffee beans. Such images are characterized using color spaces (HSV, YCC, LAB and LUV). Clustering performance is measured by means supervised and unsupervised indices based on cluster coherence and objective function value. Experiments show the effects of the affinity matrix selection in the clustering process in terms of convergence time and accuracy.

II. Materials and methods

A. Clustering method

A weighted graph can be represented as $\mathbb{G} = (\mathbb{V}, \mathbb{E}, \boldsymbol{W})$, where $\mathbb{V}$ is the set of nodes or vertices, $\mathbb{E}$ is the set of edges and $\boldsymbol{W}$ represents the relation among nodes, in other words, the affinity matrix. Given that w_{ij} represents the weight of the edge between *i*-th and *j*-th element, it must be a non-negative value. In addition, in a non-directed graph is evident that $w_{ij} = w_{ji}$. Therefore, matrix $\boldsymbol{W}$ must be a symmetric and positive semi-definite matrix. Fig. 1 shows an example of weighted graph.

Two measures are of interest: total weighted connections and degree.

Diego Peluffo dhpeluffoo@unal.edu.co, Daniel Fernando López dflopezm@unal.edu.co, José Luis Rodríguez jlrodriguezso@unal.edu.co, Germán Castellanos cgcastellanosd@unal.edu.co

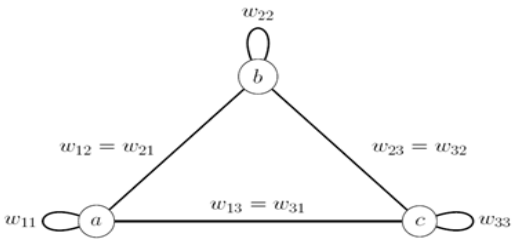


Fig 1. Weighted graph with three nodes

By letting $\mathbb{G} = (\mathbb{V}, \mathbb{E}, \boldsymbol{W})$ be a weighted graph and $\mathbf{A}, \mathbf{B} \subset \mathbb{V}$, total weighted connections between **A** and **B** can be computed as:

$$links(\mathbf{A}, \mathbf{B}) = \sum_{i \in \mathbf{A},\ j \in \mathbf{B}} w_{ij} \tag{1}$$

The degree of a set is simply the total links to all the nodes:

$$degree(\mathbf{A}) = links(\mathbf{A}, \mathbb{V}) \tag{2}$$

Using the degree as a normalization term, it can be defined the normalized links as:

$$linkratio(\mathbf{A}, \mathbf{B}) = \frac{links(\mathbf{A}, \mathbf{B})}{degree(\mathbf{A})} \tag{3}$$

In spectral analysis, term $\mathbb{V} = \{1, \dots, n\}$ represents the indices of data set to be grouped. The aim of spectral clustering is to decompose $\mathbb{V}$ into k disjoint subsets, then, $\mathbb{V} = \cup_{l=1}^{k} \boldsymbol{V}_l$ and $\boldsymbol{V}_l \cap \boldsymbol{V}_m = \emptyset, \ \forall l \neq m$, such decomposition is done, commonly, by using spectral information and orthonormal transformations.

Two special linkratios are considered for clustering: One is $linkratio(\mathbf{A}, \mathbf{B})$, which measures how many links stay within A itself [4]. The another one is $linkratio(\mathbf{A}, \mathbb{V})$, which is its complement and measures how many links escape from A. According to this, a good clustering can be achieved with both tight connections within partitions and loose connections between partitions. These two goals are captured in the K-way normalized associations ($knassoc$) and normalized cuts criteria ($kncuts$), as follows:

$$knassoc\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) = \frac{1}{k}\sum_{l=1}^{k} linkratio(\mathbb{V}_l, \mathbb{V}_l) \tag{4}$$

$$kncuts\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) = \frac{1}{k}\sum_{l=1}^{k} linkratio(\mathbb{V}_l, \mathbb{V} \backslash \mathbb{V}_l) \tag{5}$$

Because a normalization term was applied, it can be easily proved that $knassoc\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) + kncuts\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) = 1$. Thus, maximizing the associations and minimizing the cuts are achieved simultaneously. Therefore, the objective function to be maximized in order to obtain a good clustering is only:

$$\varepsilon\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) = knassoc\left(\Gamma_{\mathbb{V}}^{\mathrm{k}}\right) \tag{6}$$

This method is widely described in [4].

B. Affinity matrix

The relation degree among nodes, also called affinity measure, can be represented as a symmetric and positive semidefinite matrix as is discussed above. Therefore, being $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{nxd}}$ the data matrix, the affinity matrix can be chosen randomly as a matrix of the form $\boldsymbol{W_r = X_r X_r^T}$, where $\boldsymbol{X_r}$ is a $n \times d$ random matrix. Another form to select the affinity corresponds to the trivial case where affinity captures the inner products among rows of data matrix, i.e, $\boldsymbol{W_t = X\ X^T}$. This form is the most frequently used.

By the other hand, since the affinity is a similarity measure (i. e., is opposite to distance or dissimilarity), intuitively, can be established alternatives as $w_{ij} = 1/\mathrm{d}(x_i, x_j)$ for $(i \neq j)$ with $w_{ii} = \text{const}$, where $\mathrm{d}(\cdot,\cdot)$ is a distance measure. This measure presents an inverse relation with respect to distance and this could not convenient, in all cases, because it can generate strong changes on affinity matrix $\boldsymbol{W}$. Then, soft versions have been proposed. Exponential affinity matrix is one of them, which can be expressed as $w_{ij} = e^{-\mathrm{d}^2(x_i, x_j)}$ with $w_{ii} = \text{const}$.

The value of w_{ii} is set according to analysis conditions, when the itself affinity does not be of interest, e.g., it can be chosen as $w_{ii} = 0$. In [9], it is presented an alternative to the previous affinity measure called scaled affinity matrix which can be written as $w_{ij} = e^{-\mathrm{d}^2(x_i, x_j)/(\sigma_i \sigma_j)}$, with $w_{ii} = 0$, where $\sigma_i = \mathrm{d}(x_i, x_N)$ and x_N is the N-th nearest neighbor. The value of N is chosen regardless the scale parameter and it is established according to data dimension [1].

Table I shows the summary of affinity measures considered in this work.

TABLE I
Affinity measures

Affinity measure	Expression
Random	$\boldsymbol{W_r = X_r X_r^T}$, where $\boldsymbol{X_r}$ is a $n \times d$ random matrix
Trivial	$\boldsymbol{W_t = X\ X^T}$
Exponential	$w_{ij} = e^{-\mathrm{d}^2(x_i, x_j)}$ con $w_{ii} = cte$
Scaled exponential	$w_{ij} = e^{-\mathrm{d}^2(x_i, x_j)/(\sigma_i \sigma_j)}$, with $w_{ii} = 0$, where $\sigma_i = \mathrm{d}(x_i, x_N)$ and x_N is the N-th nearest neighbor

Fig 2. Weighted graph with three nodes

III. RESULTS AND DISCUSSION

The database is formed by high quality coffee beans images with controlled lighting. Each image corresponds to green and ripe coffee beans and some stems. Fig. 2 shows some examples from the database.

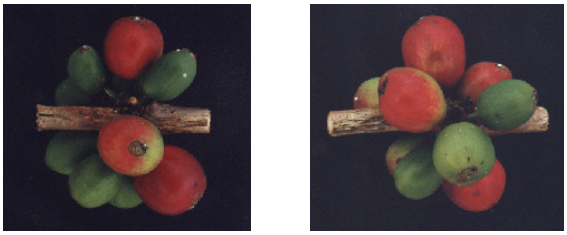
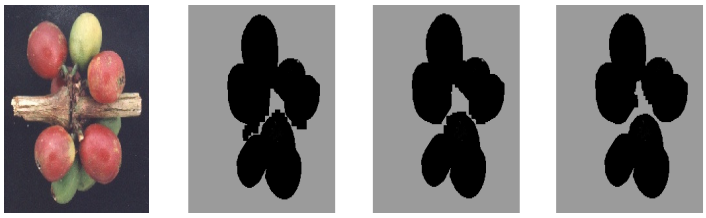


Fig 2. Samples from database

Images were characterized by using space colors such as HSV, YCC, LAB and LUV. The resultant features represent the data set: 45288 pixels (rows) and 12 features (columns). After, pixels were grouped by employing the clustering algorithm based on the normalized cuts criterion, taking into account different affinity measures and estimating the number of groups. The estimation of the number of groups is carried out by applying the eigenvalues based method described in [9].

Experimental results are shown in Fig. 3 and Table II. Clustering quality is measured in terms of clustering performance (CP) and a non-supervised measured generated by the objective function value ε (see eq (6)). Fig. 3 depicts graphically the segmentation results for an image from the database images, where the subfigures correspond to (a) original image, and performance of (b) trivial affinity, (c) exponential affinity and (d) scaled exponential affinity.



(a) (b) (c) (d)

Fig 3. Segmentation results

Table II shows the results obtained: CP, value ε and the number of groups k calculated for five images from data base. Values of mean (μ) and standard deviation (σ) are presented $(\mu - \sigma)$.

TABLE II
Clustering performance regarding to affinity measure

Affinity	CP (%)	ε_M	k
Trivial	90 – 3.71	0.77 – 2.1	8 – 3.12
Exponential	95.3 – 2.2	0.85 – 1.1	6 – 1.7
Scaled exponential	97.4 – 1.16	0.89 – 0.9	5 – 0.9

Exponential affinity shows better performance than trivial, as can be seen in Table II. By using this kind of affinity improves the CP, convergence value and estimated number of groups. This can be attributed to the local exploratory analysis (because of the scale factor) and the soft nature.

IV. CONCLUSIONS

Spectral clustering is a method that presents good performance, including when classes are not linearly separable, because it uses the information contained in eigenvectors and data transformations, however, its calculation represents a high computational cost. Therefore, it is necessary to design strategies for improving the convergence time without losing the effectiveness. To that end, affinity measures represent a good alternative. Also, it can be included an estimation of the number of groups stage based on spectral analysis in order to initialize the algorithm in a proper way.

Non-supervised measure associated with the coherence of clusters that is obtained through a partitioning spectral criterion, is a proper performance index because penalizes the number of groups and generates a value near to 1 when the grouping is correct and the affinity matrix is properly chosen.

As future wok, other properties and spectral clustering algorithms will be studied in order to develop an unsupervised system that will be less sensitive to initialization and reduce the computational cost; including stages for initialization, estimation of number of groups, feature selection and grouping.

V. ACKNOWLEDGMENTS

Authors thank to “Beca para estudiantes sobresalientes de posgrado de la Universidad Nacional de Colombia” and “Programa de financiación para Doctorados Nacionales de Colciencias”.

REFERENCES

[1] P. K. Chan, M. D. F. Schlag & J.T. Zien, (1994). K-way ratio cut partitioning and clustering. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*. 13:1088- 1096.

[2] B. Hendrickson and R. Leland, (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*. 16(2):452-469.

[3] H. Ishikawa and D. Geiger, (1998). Segmentation by grouping junctions. *In CVPR’98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Page 125, Washington, DC, USA. IEEE Computer Society.

[4] Yu, Stella X. and Shi, Jianbo. “Multiclass Spectral Clustering” In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. 2003. IEEE Computer Society.

[5] Jianbo Shi and J. Malik, (2003). Normalized cuts and image segmentation. *Proceedings of the Ninth IEEE International Conference on Computer Vision.* Page 313, Washington, DC, USA. IEEE Computer Society.

[6] Bui The Anh & D. D. X. Thanh, (2007). A perron-frobenius theorem for positive quasipolynomial matrices associated with homogeneous difference equations. *Journal of applied mathematics.*

[7] P. K. Chan, M. Schlag, & J. Y. Zien, (1993). Spectral k-way ratio-cut partitioning and clustering. *In Proceedings of the 1993 symposium on Research on integrated systems.* Pages 123–142, Cambridge, MA, USA. MIT Press.

[8] M. Maila and J. Shi, (2001). Learning segmentation with random walk. *In Neural Information Processing Systems (NIPS)*.

[9] L. Zelnik-manor and P. Perona, (2004). Self-tuning spectral clustering. *In Advances in Neural Information Processing Systems* 17. Pages 160-1608. MIT Press.

[10] S. Lee & M. H. Hayes, (2004). Properties of the singular value decomposition for efficient data clustering. *Signal Processing Letters.* IEEE, 11(11):862–866, Nov. 2004.

[11] A. Y. Ng, M. I. Jordan, & Y. Weiss. On spectral clustering: Analysis and an algorithm. *In Advances in Neural Information Processing Systems 14*. Pages 849-856. MIT Press, 2001.