ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Normalized Cuts Clustering with Prior Knowledge and a Pre-clustering Stage

D. Peluffo-Ordoñez[1], A. E. Castro-Ospina[1], D. Chavez-Chamorro[1],
C. D. Acosta-Medina[1], and G. Castellanos-Dominguez[1] *

1- Signal Processing and Recognition Group.
Universidad Nacional de Colombia, Manizales, Colombia.
{dhpeluffoo,aecastroo,dchavezc,cdacostam,cgcastellanosd}@unal.edu.co

Abstract. Clustering is of interest in cases when data are not labeled enough and a prior training stage is unfeasible. In particular, spectral clustering based on graph partitioning is of interest to solve problems with highly non-linearly separable classes. However, spectral methods, such as the well-known normalized cuts, involve the computation of eigenvectors that is a highly time-consuming task in case of large data. In this work, we propose an alternative to solve the normalized cuts problem for clustering, achieving same results as conventional spectral methods but spending less processing time. Our method consists of a heuristic search to find the best cluster binary indicator matrix, in such a way that each pair of nodes with greater similarity value are first grouped and the remaining nodes are clustered following a heuristic algorithm to search into the similarity-based representation space. The proposed method is tested over a public domain image data set. Results show that our method reaches comparable results with a lower computational cost.

1 Introduction

Clustering has been used in numerous applications, being preferred mainly in cases when data are partially or not labeled as well as when prior training is unfeasible. In particular, spectral approaches are of great interest since they are able to solve problems with highly non-linearly separable classes [1]. Perhaps, the most frequently used approach is the well-known normalized cuts clustering (NCC). The NCC methods can be easily understood from a graph theory point of view, where data points are seen as nodes [2]. Generally, the input parameters for NCC are the desired number of groups and the similarity matrix containing all the similarities among nodes. Most of approaches have been addressed to yield a relaxed problem formulation that is solved by means of an eigenvalues and eigenvectors decomposition, for instance, kernel k-means [3, 4]. Also, there exist approaches that solve the graph partitioning problem by means of a minimum cuts formulation [5] or by a quadratic problem [1]. Nonetheless, because of the high computational cost that often involves the computation of eigenvectors, some studies have concerned about getting alternatives for solving the normalized cuts clustering without using eigenvectors such as multilevel approaches with weighted graph cuts [6], and quadratic problem formulations with linear constrains [7, 8].

In this paper, we propose a method based on a heuristic search carried out over the representation space given by the similarities among data points. Our method is a lower

*This work has been supported by the project "Clasificación no supervisada de bioseñales orientada a detección de patologías" associated with the "Jóvenes Investigadores e Innovadores 2011" program by COLCIENCIAS

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

computational cost alternative to NCC methods based on eigenvector decompositions, named NCC based on a heuristic search (NCChs). The foundation of our method lies in deriving a new simple objective function by re-writing the quadratic expressions for Multi-Cluster Spectral Clustering (MCSC) [1] in such a way that we can intuitively design a searching process on the similarity representation space. Also, to guarantee the convergence, we propose to incorporate prior knowledge, i.e., taking advantage of the original labels to set initial data points for starting the grouping process. Our method's performance is compared with those obtained by kernel k-means [4], min-cuts [5] and MCSC [1]. For experiments, some images taken from Berkeley databases are considered [9]. Clustering performance is assessed in terms of segmentation quality. The proposed NCChs significantly outperforms conventional spectral clustering methods for solving the NCC problem in terms of speed and achieves comparable performance results as well. This paper is organized as follows: In section 2, the reference NCC formulation is presented. In section 3, we explain the proposed heuristic search-based clustering. Section 4 shows the results and discussion. Finally, some final remarks and conclusions are presented in section 5.

2 Normalized cuts based clustering (NCC)

Let $\boldsymbol{X} \in \mathbb{R}^{N\times d}$ be the data to be clustered representing the geometric coordinates of nodes, with $\boldsymbol{X} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_N^\top)^\top$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the i-th data point related to i-th node, and N is the number of nodes. In matrix representation terms, The aim of NCC is to determine a binary cluster indicator matrix $\boldsymbol{M} \in \{0,1\}^{N\times K}$ such that $\boldsymbol{M} = (\boldsymbol{m}^{(1)}, \ldots, \boldsymbol{m}^{(K)})$, where K is the number of groups and each vector $\boldsymbol{m}^{(k)} \in \{0,1\}^N$ is a column vector formed by a binary data point membership regarding cluster k [1]. Also, because each node can only belong to one cluster, the condition $\boldsymbol{M}\mathbf{1}_K = \mathbf{1}_N$ must be satisfied, where $\mathbf{1}_d$ is a d-dimensional all-ones vector. Then, the normalized cuts-based clustering (NCC), described in [1], can be written as:

$$\max_{\boldsymbol{M}} \varepsilon(\boldsymbol{M}) = \frac{1}{K}\frac{\mathrm{tr}(\boldsymbol{M}^\top \boldsymbol{\Omega} \boldsymbol{M})}{\mathrm{tr}(\boldsymbol{M}^\top \boldsymbol{D} \boldsymbol{M})}; \quad \text{s.t.} \quad \boldsymbol{M} \in \{0,1\}^{N\times K}, \quad \boldsymbol{M}\mathbf{1}_K = \mathbf{1}_N \tag{1}$$

where $\boldsymbol{\Omega} \in \mathbb{R}^{N\times N}$ is the similarity matrix, $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ is the degree matrix defined as $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{\Omega}\mathbf{1}_N)$, and $\mathrm{Diag}(\cdot)$ denotes a diagonal matrix built by its argument vector.

3 Solution of NCC problem via a heuristic search

First, take into consideration the following identities:

$$\mathrm{tr}(\boldsymbol{M}^\top \boldsymbol{\Omega} \boldsymbol{M}) = \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{\Omega} \boldsymbol{m}^{(k)} = \sum_{k=1}^{K}\sum_{s=1}^{N}\sum_{t=1}^{N} m_{tk}\Omega_{ts}m_{sk}, \quad \text{and} \tag{2}$$

$$\mathrm{tr}(\boldsymbol{M}^\top \boldsymbol{D} \boldsymbol{M}) = \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{D} \boldsymbol{m}^{(k)} = \sum_{k=1}^{K}\sum_{s=1}^{N} m_{sk}^2 d_{ss} = \sum_{s=1}^{N} d_{ss} \tag{3}$$

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

According to the previous statements and since $\sum_{s=1}^{N} d_{ss} = ||\boldsymbol{\Omega}||_{L_1}$ is constant, the term $\text{tr}(\boldsymbol{M}^\top \boldsymbol{\Omega} \boldsymbol{M})$ is the only one term to be of interest. By recalling equation (2) and since matrix $\boldsymbol{\Omega}$ is symmetric, it is possible to write a new cost function to be maximized, as follows:

$$\text{tr}(\boldsymbol{M}^\top \boldsymbol{\Omega} \boldsymbol{M}) = \text{tr}(\boldsymbol{\Omega}) + 2 \sum_{s>t} \Omega_{ts} m_{tt'} m_{ss'} = \text{tr}(\boldsymbol{\Omega}) + 2 \sum_{s>t} \Omega_{ts} \delta_{ts} \qquad (4)$$

Notice that δ_{ts} becomes 0 in case of the dot product between the rows t and s equals to 0, i.e., when such rows are linearly independent. Otherwise, it yields 1 pointing out that row vectors are the same -containing 1 in the same entry. The approach proposed here, named NCChs consists of using prior knowledge about the known data labeling and a pre-clustering stage to cluster heuristically the input data.

3.1 Prior knowledge and pre-clustering

According to eq. 4, since $\text{tr}(\Omega)$ is constant, the term to be maximized is plainly $\sum_{s>t} \Omega_{ts} \delta_{ts}$. Within this framework, the solution may lead to a trivial solution wherein all elements are belonging into the same cluster. In order to avoid this drawback, we propose to incorporate prior knowledge taking advantage of the original labels. Then, K data points (one per class) are arbitrarily chosen from the whole data set, in such a way K different membership values (rows of matrix $\boldsymbol{M}$) are known in advance. Therefore, clusters are assigned according to the maximum value of similarity but preserving the K initial seed nodes belonging to respective clusters. This can be easily done by setting the entry m_{ik} representing the prior nodes to be 1, and 0 for the remaining entries on the same row i. Also, in order to avoid wrongly assigning closer data points belonging to different clusters, we first carry out a pre-clustering process, where a relatively low percentage of the whole data set (ϵ) is added to the seed nodes whose value of similarity is maximum. Denote the indexes related to the initial nodes as $\boldsymbol{q} = (q_1, \ldots, q_K)$ where $q_k \in \{1, \ldots, N\}$. Then, the coordinates of seed nodes are $\{x_{q_1}, \ldots, x_{q_K}\}$. Once all the first assignments are done, we have the initial K seed nodes.

3.2 Heuristic search

Once the pre-clustering stage is done, we have K initial clusters. The remaining data points are assigned in accordance to the maximum similarity value between itself and any of the previously assigned data points. The proposed heuristic to form the final clusters work as follows. Each time that an entry Ω_{ij} is chosen as the maximum similarity in the actual iteration, it is then removed by setting $\Omega_{ij} = 0$ in order to avoid taking it into consideration for the next iteration, and so on. This assignment process is done until all the data points are belonging into any cluster, in other words, $\sum_{i=1}^{N} \sum_{k=1}^{K} m_{ik} = ||\boldsymbol{M}||_{L_1} = N$. Note that we can employ L_1-norm since all entries of $\boldsymbol{M}$ are positive. A graphic explanation of the search is shown in Fig. 1.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

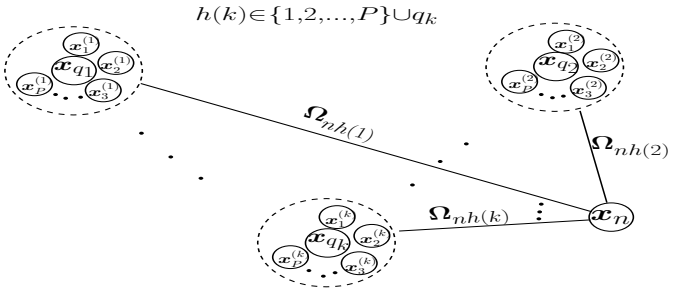

Fig. 1: Heuristic cluster assignment for a new data point $\boldsymbol{x}_n$

As can be appreciated, the initial data points assigned to each cluster corresponds to the coordinates given by the seed nodes: $\{\boldsymbol{x}_{q_1}, \ldots, \boldsymbol{x}_{q_K}\}$. After the seed nodes are assigned, pre-clustering is done by adding P data points to every seed node to form the initial clusters. Value P is selected as the integer closest to $\epsilon\%$ of the number of data N. Term $h(k)$ denotes the set of indexes related to cluster k including its corresponding seed node q_k, which is incremented when new data points are added to cluster k. In the example shown in Fig. 1, $\boldsymbol{x}_n$ is the new data point to be grouped. To cluster it, we compare the similarity between $\boldsymbol{x}_n$ and the actual formed clusters, i.e., $\boldsymbol{\Omega}_{nh(k)}, \forall k \in [K] = \{1, \ldots, K\}$. At the end, $\boldsymbol{x}_n$ is assigned to that cluster presenting maximum similarity value regarding node n, following the rule: it is assigned to cluster k such that $\arg\max_k \boldsymbol{\Omega}_{nh(k)}$, s.t. $k \in [K]$.

4 Results and discussion

To assess the performance of our method, we employ some images extracted from the free access Berkeley Segmentation Data Set [9]. Images are characterized by RGB color space and the xy position of each pixel. Due to memory usage restrictions, we resize the images at 20% of the original size. The proposed method is compared with kernel k-means (KKM) [4], min cuts (Min-cuts) [5] and multi-cluster spectral clustering (MCSC) [1]. All the methods are performed with a given number of clusters K set as shown in shown in Fig. 2 and using the scaled exponential similarity matrix as described in [4], setting the number of neighbors to be 9. To compare adequately the methods, we standardize the results by setting the same initial parameters (number of clusters K and a set of initial nodes $\boldsymbol{q}$) for all cases. Experiments were done using MatLab Version 7.12.0.635 (R2011a) on a computer with RAM 8Gb, and processor Intel(R) Xeon(R) CPU X5660 2.8GHz.

Image segmentation results are shown in Fig. 2. The segmentation performance is quantified by a supervised index noted as Probabilistic Rand Index (PR) explained in [10], such that $PR \in [0, 1]$, being 1 when regions are properly segmented. Proposed NCChs achieves in most cases the best performance, showing clearly the benefit of the use of a bit of prior knowledge, just K seed nodes are needed. Table 1 depicts the processing time employed by each method. Processing times are given as a proportion of the highest one. In this case, highest process times for all considered images are $T_p^{(1)} = 112.87s$, $T_p^{(2)} = 65.93s$, $T_p^{(3)} = 66.43s$, $T_p^{(4)} = 66.56s$ and $T_p^{(5)} = 66.75s$.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

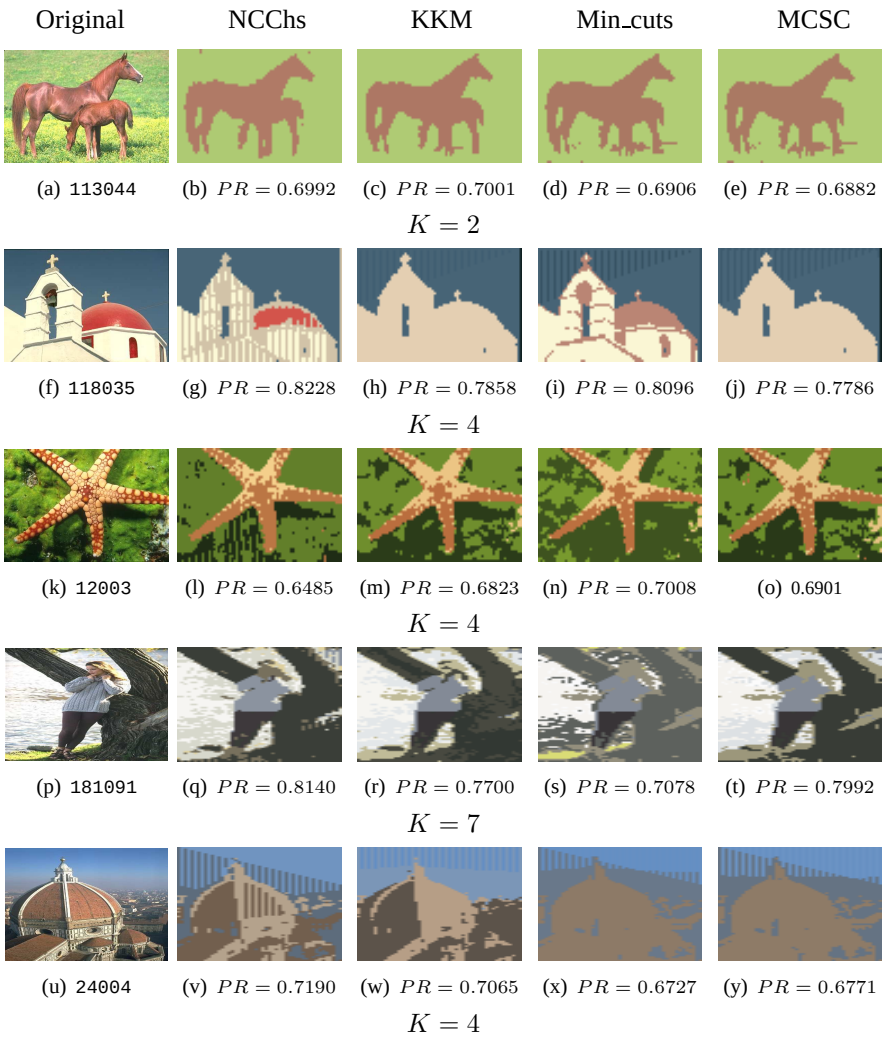

Fig. 2: Clustering performance on image segmentation along 10 iterations

Method	**Image**				
	113044	118035	12003	181091	24004
NCChs	$0.5553T_p^{(1)}$	$0.6486T_p^{(2)}$	$0.5643T_p^{(3)}$	$0.5371T_p^{(4)}$	$0.5980T_p^{(5)}$
KKM	$0.8538T_p^{(1)}$	$0.9339T_p^{(2)}$	$0.9244T_p^{(3)}$	$0.9327T_p^{(4)}$	$0.9342T_p^{(5)}$
Min-cuts	$0.7663T_p^{(1)}$	$0.8951T_p^{(2)}$	$0.7787T_p^{(3)}$	$0.7412T_p^{(4)}$	$0.8253T_p^{(5)}$
MCSC	$T_p^{(1)}$	$T_p^{(2)}$	$T_p^{(3)}$	$T_p^{(4)}$	$T_p^{(5)}$

Table 1: Average clustering processing time ratio along 10 iterations

Note that our method spends the least processing time. Index PR compares the

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

resultant segmentation with multiple manually labeled ground-truth images through a variability function regarding each pair of pixels in the ground-truth set. Then, we can say that, in terms of PR values, NCChs achieves a suitable trade-off between performance and computational cost in contrast with the other considered methods.

5 Conclusions and future work

Spectral clustering methods have been applied in several applications and have shown to be a powerful tool to solve grouping problems when data contains hardly separable classes. Nonetheless, since they involve the computation of eigenvectors, they can be prohibitive for clustering high-dimensional data. In this work, we introduced a new alternative to solve the normalized cuts problem for clustering without using eigenvectors. From the conventional formulation, we derived a new cost function that can be heuristically maximized by seeking for the nodes with maximum similarity. The heuristic search outcome is a binary cluster indicator matrix. Also, in order to avoid wrong assignments, we initialize the algorithm with some seed nodes and carry out a pre-clustering stage. Results show that proposed method reduces the computational cost in comparison with conventional spectral clustering methods, and achieves comparable performance as well.

For future works, we will focus on improving the heuristic search by both estimating properly the parameters for the pre-clustering stage and detecting outliers. As well, generalizing the NCChs to be capable of working properly in cases when there exist minority/unbalanced clusters and prior knowledge is not available.

References

[1] S. Yu and Shi. Jianbo. Multiclass spectral clustering. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 313, Washington, DC, USA, 2003. IEEE Computer Society.

[2] N. Fan and P.M. Pardalos. Multi-way clustering and biclustering by the ratio cut and normalized cut in graphs. *Journal of Combinatorial Optimization*, pages 1–28, 2012.

[3] Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Kernel k-means, spectral clustering and normalized cuts. 2004.

[4] Lihi Zelnik-manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems 17*, pages 1601–1608. MIT Press, 2004.

[5] Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, Aug 2000.

[6] I.S. Dhillon, Y. Guan, and B. Kulis. Weighted graph cuts without eigenvectors a multilevel approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(11):1944–1957, 2007.

[7] L. Xu, W. Li, and D. Schuurmans. Fast normalized cut with linear constraints. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2866–2873. IEEE, 2009.

[8] D. Peluffo, C. D. Acosta, and G. Castellanos. An improved multi-class spectral clustering based on normalized cuts. *XVII Iberoamerican Congress on Pattern Recognition - CIARP*, 2012.

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[10] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):929–944, 2007.