Constrained Spectral Clustering under a Local Proximity Structure Assumption

Qianjun Xu and **Marie desJardins**
Dept. of CS&EE
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250

Kiri Wagstaff
Jet Propulsion Laboratory
Machine Learning Systems Group
4800 Oak Grove Dr., Pasadena CA 91109

Abstract

This work focuses on incorporating pairwise constraints into a spectral clustering algorithm. A new constrained spectral clustering method is proposed, as well as an active constraint acquisition technique and a heuristic for parameter selection. We demonstrate that our constrained spectral clustering method, CSC, works well when the data exhibits what we term *local proximity structure*. Empirical results on synthetic and real data sets show that CSC outperforms two other constrained clustering algorithms.

Introduction

This work focuses on incorporating pairwise constraints in spectral clustering to improve performance. Recently, there has been some research in the area of constrained spectral clustering. Yu & Shi (2004) formulated the problem as a constrained optimization problem. This method is appropriate when constraints can be propagated to neighboring items. Kamvar, Klein, & Manning (2003) integrated constraints by updating an affinity matrix. The affinity matrix is used to construct a Markov transition probability matrix whose eigenvectors are used to perform clustering. We call this algorithm KKM after the authors' initials. In this paper, we present a variation of KKM that outperforms the original algorithm on synthetic and real-world data sets.

The contributions of this paper are (1) a constrained spectral clustering method that performs well empirically, (2) a parameter selection heuristic to choose an appropriate scaling parameter for our algorithm, and (3) an active constrained clustering technique. Empirical results on synthetic and real data sets show that CSC outperforms two other constrained clustering algorithms.

Background

Spectral Clustering. Spectral clustering uses an affinity matrix, defined as:

$$A_{i,j} = exp((-\delta_{ij}^2)/(2\sigma^2)) \quad (1)$$

where δ_{ij} is the Euclidean distance between point i and j and σ is a free scale parameter. By convention, $A_{ii} = 0$. Meilă and Shi (2001) showed the connection of many spectral clustering methods to a random Markov chain model, as follows. Let D be the diagonal matrix with element $d_{ii} = \sum_{j=1}^{n} A_{ij}$. A new matrix P is derived using equation $P = D^{-1}A$. Each row of P sums to 1, so we can interpret entry P_{ij} as a Markov transition probability from i to j. The Markov chain defined by P will identify the subsets that have high intra-cluster transition probabilities and low inter-cluster probabilities.

Constrained Spectral Clustering. Inspired by the random walk model, Kamvar, Klein, & Manning (2003) proposed a method which is based on a different transition matrix, N, as follows:

$$N = (A + d_{max}I - D)/d_{max}, \quad (2)$$

where d_{max} is the largest element in the degree matrix D and I is the identity matrix.

Given a must-link constraint (i, j), KKM modifies the corresponding affinities so that $A_{ij} = A_{ji} = 1$. A cannot-link constraint (i, j) is incorporated by setting $A_{ij} = A_{ji} = 0$, preventing a direct transition between i and j. The nonzero diagonal entries in matrix N tend to generate singleton clusters when there are outliers in the data. To overcome this drawback, our CSC method uses the matrix $P = D^{-1}A$. After updating the affinity matrix, we derive the largest k eigenvectors, which we collectively term the *eigenspace*. K-means clustering is then applied to the rows in this eigenspace. A data point is labeled as belonging to cluster i if the ith row in the eigenspace is labeled with cluster i.

Constrained Spectral Clustering (CSC)

Local Proximity Structure. A data set is said to exhibit *local proximity structure* if the subclusters within the data set are locally convex: i.e., if each point belongs to the same cluster as its closest neighbors. However, this proximity structure may not hold globally: neighboring subclusters may not belong to the same cluster, and subclusters of a single cluster may be separated spatially by a subcluster of a different cluster. Figure 1 shows a synthetic XOR data set that exhibits local proximity structure. We index the items as follows: the bottom left (1 to 10), the bottom right (11 to 20), the top right (21 to 30), and the upper left (31 to 40).

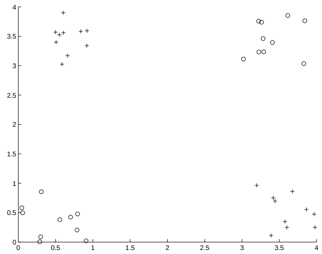

Figure 1: The XOR data set.

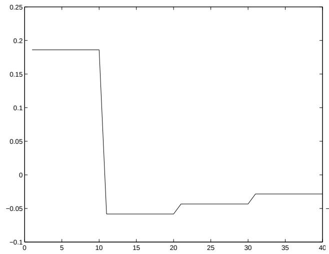

Figure 2: The 2nd eigenvector for the XOR data set with no constraints.

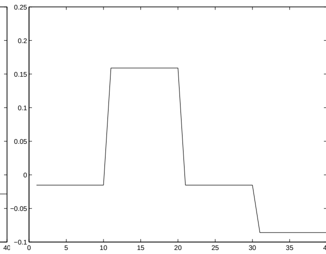

Figure 3: The 2nd eigenvector for XOR with one constraint.

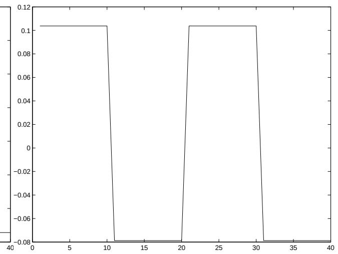

Figure 4: The 2nd eigenvector for XOR with two constraints.

Analyzing the Eigenvectors. Meilă & Shi (2001) defined a *piecewise constant eigenvector* as one where the items belonging to the same cluster have the same values in the eigenvector. If we choose σ (in Equation 1) appropriately, we can guarantee a piecewise constant eigenvector, as in Figure 2 for the XOR data set.

Instead of directly estimating σ, we introduce a new parameter, m. Let δ_i^m represent the Euclidean distance of point i to its mth closest neighbor. We choose σ so that $exp(\frac{-max_\delta^2}{2\sigma^2}) = \epsilon = 0.001$, where $max_\delta = \max_i\{\delta_i^m\}$. This ensures that each point in the data set will have affinity $\geq \epsilon$ to its m closest neighbors.

To choose m, we propose the following heuristic: (1) compute the distance δ_{ij} between all pairs of data points; (2) sort each row $\delta_{i.}$ in ascending order; (3) find the distance δ_{im} that has the largest gap with $\delta_{i,m+1}$; and (4) select the value for σ that maps this distance δ_{im} to affinity value ϵ (e.g., $\epsilon = 0.001$). The intuition is that step (3) identifies the largest natural "gap" in the data, and uses this to select a good cluster size m, leading to an effective choice for σ. Applying this heuristic on XOR yields $m = 9$, which is used to compute the eigenvectors shown in Figures 2, 3, and 4.

Active Clustering. Since a piecewise constant eigenvector identifies subclusters, we can pick any point from each subcluster as a representative. The constraints are obtained by querying the user to label these representatives. For example, for the XOR data set, we construct queries about two pairs of points: $(1, 21)$ and $(11, 31)$, which are both labeled as must-link constraints. Figure 3 shows the 2nd eigenvector after applying the first constraint. Figure 4 shows the eigenvector after applying both constraints. We can see that after imposing only two *actively selected* must-link constraints, the correct clusters for this hard problem are identified.

Experimental Results on Real-World Data

We have found experimentally that even with *randomly selected* constraints, CSC works well. We compare the performance of our CSC algorithm to that of KKM (Kamvar, Klein, & Manning 2003) and CCL (Klein, Kamvar, & Manning 2002), using the soybean and iris data sets from the UCI archive (Blake & Merz 1998). The Rand index (Rand 1971) (averaged over 100 runs) is plotted in Figures 5 and 6. CSC outperforms CCL and KKM on these data sets.

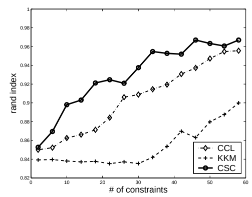

Figure 5: The Rand index for the soybean data set.

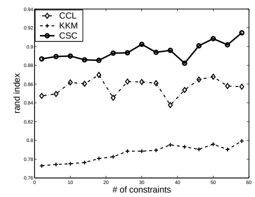

Figure 6: The Rand index for the iris data set.

Compared to KKM, we believe that P is a better choice than N for two main reasons. First, N can cause problems when there are outliers in data set. Second, and more importantly, it cannot identify piecewise constant eigenvectors for data sets that exhibit local proximity structure. The CCL method performs worse than our algorithm on real data sets, and cannot recognize subclusters easily when the data set exhibits local proximity structure.

References

Blake, C., and Merz, C. 1998. UCI repository of machine learning databases. `http://www.ics.uci.edu/~mlearn/MLRepository.html`.

Kamvar, S. D.; Klein, D.; and Manning, C. D. 2003. Spectral learning. In *Proceedings of IJCAI-03*, 561–566.

Klein, D.; Kamvar, S. D.; and Manning, C. D. 2002. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Proceedings of ICML-02*, 307–314. Morgan Kaufmann.

Meilă, M., and Shi, J. 2001. A random walks view of spectral segmentation. In *Proceedings of AI-STAT'01*.

Rand, W. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66:846–850.

Yu, S. X., and Shi, J. 2004. Segmentation given partial grouping constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(2):173–183.