Weighted Co-regularization for Multiview Spectral Clustering

Jagadeesh Jagarlamudi, Abhishek Kumar, Hal Daumé III
{jags,abhishek,hal}@umiacs.umd.edu
Department of Computer Science
University of Maryland
College Park, MD 20742

Piyush Rai
piyush@cs.utah.edu
School of Computing
University of Utah
Salt Lake City, Utah 84112

Given a kernel matrix W, Spectral Clustering [3, 4] involves finding a vector $\mathbf{u}$ (which we refer as uni-representation) which minimizes the following objective function:

$$\arg\min_{\mathbf{u}} \frac{1}{2} \sum_{ij} W_{ij}(u_i - u_j)^2 \tag{1}$$

with an appropriate length constraint on u. The intuition is that, if a pair of points x_k and x_l are highly similar, *i.e.* W_{kl} is very high, then the objective function enforces the components u_k and u_l of the uni-representation vector to be close to each other (and hence two two points are more likely to belong to the same cluster). The objective function in Eqn. 2 can be rewritten as:

$$\arg\min_{\mathbf{u}} \frac{1}{2} \sum_{ij} W_{ij}(u_i - u_j)^2 = \arg\min_{\mathbf{u}} \mathbf{u}^T L \mathbf{u} \tag{2}$$

$$\text{s.t.} \;\; \mathbf{u}^T D \mathbf{u} = 1 \tag{3}$$

where $L = D - W$ is the unnormalized Laplacian matrix corresponding to W and D is the diagonal matrix with $D_{ii} = \sum_j W_{ij}$. It turns out that, this is same as solving the generalized eigenproblem $L\mathbf{u} = \lambda D\mathbf{u}$.

For multi-view data, spectral clustering can be naïvely applied on individual views and each view would give an independent representation of the data. However, performing clustering with each of these representations may potentially lead to clusterings that are different from each other and possibly different from the underlying true clustering. To overcome this problem, [2] proposes adding a co-regularization term to the objective function which minimizes the disagreement among the resulting clusterings across the multiple views. In this paper we propose a new co-regularization term which has the same form as the individual objective functions (Eqn. 2) and hence can be optimized easily.

Let $\mathbf{u}$ and $\mathbf{v}$ be the uni-representations of the observations in the individual views. Given the similarity matrices in both the views, $W^{(1)}$ and $W^{(2)}$, single view spectral clustering minimizes $\sum_{ij} W^{(1)}_{ij}(u_i - u_j)^2$ and $\sum_{ij} W^{(2)}_{ij}(v_i - v_j)^2$ for both the views independently (subject to length constraints of the form Eqn. 3). Since the individual similarity matrices can be noisy we use a new term $W^{(12)}$, called view-similarity matrix, which captures the notion of trustworthyness of the kernel matrices across views. A higher value of $W^{(12)}_{ij}$ for a pair indicates that, for this pair, we trust the similarity values provided by the individual similarity matrices. Likewise, setting all the entries of this matrix to a constant value indicates that we trust all the pairs equally. Then we propose to use the following cost function:

$$C^{(12)} = \sum_{ij} W^{(12)}_{ij} \Big((u_i - u_j) - (v_i - v_j) \Big)^2 = (\mathbf{u} - \mathbf{v})^T L^{(12)} (\mathbf{u} - \mathbf{v}) \tag{4}$$

where $L^{(12)}$ is the Laplacian corresponding to the view-similarity matrix $W^{(12)}$.

Folding the co-regularization term into the objective function, the final optimization problem becomes $\arg\min_{\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{u},\mathbf{v})$ where $\mathcal{L}(\mathbf{u},\mathbf{v})$ is given by:

$$\begin{aligned}&\mathbf{u}^T L^{(1)}\mathbf{u} + \mathbf{v}^T L^{(2)}\mathbf{v} + \alpha(\mathbf{u}-\mathbf{v})^T L^{(12)}(\mathbf{u}-\mathbf{v}) \\ &\text{s.t. } \ \mathbf{u}^T D^{(1)}\mathbf{u} + \mathbf{v}^T D^{(2)}\mathbf{v} = 1\end{aligned} \tag{5}$$

Where $L^{(1)}$ and $L^{(2)}$ are the unnormalized laplacians of both the views. Forming the Lagrangian and setting its first derivative with respect to $\mathbf{u}$ and $\mathbf{v}$ to zero yields the following generalized eigenproblem:

$$\begin{bmatrix} L^{(1)} + \alpha L^{(12)} & -\alpha L^{(12)} \\ -\alpha L^{(12)} & L^{(2)} + \alpha L^{(12)} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \lambda \begin{bmatrix} D^{(1)} & \mathbf{0} \\ \mathbf{0} & D^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \tag{6}$$

In our experiments on WebKB data set [1], we show that our method performs better than many baseline systems. Existing co-regularization based approaches [2] ignore the fact that example pairs which are deemed potentially similar across all the views, rather than across only some of the views, are more likely to belong to the same cluster. Our method provides a way to regularize certain pairs of points more than the rest.

References

[1] Mark Craven, Dan Dipasquo, Dayne Freitag, Andrew Mccallum, Tom Mitchell, and Kamal Nigam. Learning to extract symbolic knowledge from the world wide web. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI)*, pages 509–516. AAAI Press, 1998.

[2] Abhishek Kumar, Piyush Rai, and Hal Daumé III. Co-regularized spectral clustering with multiple kernels. In *NIPS 2010: Workshop on New Directions in Multiple Kernel Learning*, 2010.

[3] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2002.

[4] Ulrike von Luxburg. A Tutorial on Spectral Clustering. *Statistics and Computing*, 2007.

Topic: learning algorithms
Preference: oral