MBO Details

1 Nystrom

[1–3] Let X be the set of graph nodes, and W the weight matrix. Let $A \subseteq X$ such that $|A| \ll |X|$, and let $B = X \backslash A$. Then up to a rearrangement of nodes, we can write

$$W = \begin{pmatrix} W_{AA} & W_{AB} \\ W_{BA} & W_{BB} \end{pmatrix}, \tag{1}$$

where the matrix $W_{AB} = W_{BA}^T$ consists of weights between nodes in A and nodes in B, W_{AA} consists of weights between pairs of nodes in A, and W_{BB} consists of weights between pairs of nodes in B. Nyström's extension approximates W as

$$W \approx \begin{pmatrix} W_{AA} \\ W_{BA} \end{pmatrix} W_{AA}^{-1} \begin{pmatrix} W_{AA} & W_{AB} \end{pmatrix}. \tag{2}$$

In particular, this approximates

$$W_{BB} \approx W_{BA} W_{AA}^{-1} W_{AB}.$$

A few words on the error of approximation. Suppose W is symmetric positive semidefinite (as it is in our example), then we can write $W = V^T V$ for some matrix V. It turns out that the Nystrom extension approximates the unknown part of V (corresponding to W_{BB} by projecting it orthogonally onto the known part (corresponing to W_{AB}). This explained much more in [1].

Here we use the normalized graph Laplacian

$$L = I - D^{-1/2} W D^{-1/2},$$

where D is the degree matrix. So to solve the eigenproblem on L, we solve it on the normalized version of W, which we calculate as follows.

$$d_X = \begin{bmatrix} W_{AA} & W_{AB} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{3}$$

$$d_Y = \begin{bmatrix} W_{BA} & W_{BA} W_{AA}^{-1} W_{AB} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{4}$$

Then define the normalized weights

$$\hat{W}_{AA} = W_{AA}./\left(\sqrt{d_X}\sqrt{d_X}^T\right), \tag{5}$$

$$\hat{W}_{AB} = W_{AB}./\left(\sqrt{d_X}\sqrt{d_Y}^T\right), \tag{6}$$

where ./ signifies componentwise division.

This is when I usually leave off. It turns out this is enough to calculate the eigenvectors of the approximate $\hat{W}$. Very exciting. In particular, we compute and store matrices of size at most $|A| \times |X|$ the entire time.

2 MBO algorithm

[2,4,5] Notation: let $N = |X|$, $m =$ number of classes. We'll keep track of our classification via an $N \times m$ *assigment matrix* u. Entry (i, j) of u stores the probability that element $x_i \in X$ belongs in class j. The final output matrix will contain exactly one 1 in each row (the rest are zero), but in the intermediate steps it can be anything that sums to one. Also, we like to label the i-th row of u as u_i for notational convenience.

Here we are minimizing the energy

$$E(u) = \epsilon\langle u, L_s u\rangle + \frac{1}{\epsilon}\sum_i W(u_i) + \sum_i \frac{\mu}{2}\lambda(x_i)\left\|u_i - \hat{u}_i\right\|_{L_2}^2. \tag{7}$$

The first term gives the graph-cut energy. The second term is the multiwell potential

$$W(u_i) = \prod_{k=1}^{m}\frac{1}{4}\left\|u_i - e_k\right\|_{L_1}^2. \tag{8}$$

I actually don't know why they use this exact potential. The main idea is clear though. The term encourages each u_i to be close to one of the simplex vertices e_k, i.e. close to completely classified. The last term is the fidelity. μ is an input parameter (generally as big as you can while maintaining stability), λ is 1 or 0 depending on if that x_i is supervised or not.

If we were to minimize by gradient descent, our update would be given by

$$\frac{\partial u}{\partial t} = -\epsilon L_s u - \frac{1}{\epsilon}W'(u) - \mu\lambda(x)(u - \hat{u}) \tag{9}$$

Instead we propose to minimize this via an MBO algorithm. Specifically, diffuse then threshold until we reach some stopping point. The diffuse step is given by

$$\frac{u^{n+\frac{1}{2}} - u^n}{dt} = -L_s u^n - \mu\lambda(x)(u^n - \hat{u}). \tag{10}$$

Then for thresholding you just let $u_i = e_r$ where r is the index of the biggest value in u_i. I.e. you threshold one number of u_i up to 1 and the rest to 0.

The important thing here is that we can use the eigenvectors of L_s (calculated earlier) to quickly do the diffuse step. Just change coords, so that the $L_s u^n$ step instead because multiplication by a diagonal matrix.

References

[1] Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2), February 2004. 1, 1

[2] Ekaterina Merkurjev, Tijana Kostic, and Andrea L Bertozzi. An mbo scheme on graphs for classification and image processing. *SIAM Journal on Imaging Sciences*, 6:1903–1930, October 2013. 1, 2

[3] J. T. Woodworth, G. O. Mohler, A. L. Bertozzi, and P. J. Brantingham. Non-local crime density estimation incorporating housing information. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 372(2028), 2014. 1

[4] Zhaoyi Meng, Ekaterina Merkurjev, Alice Koniges, and Andrea L Bertozzi. Hyperspectral video analysis using graph clustering methods. Note: This is a preprint, 2017. 2

[5] Huiyi Hu, Justin Sunu, and Andrea L. Bertozzi. *Multi-class Graph Mumford-Shah Model for Plume Detection Using the MBO scheme*, pages 209–222. Springer International Publishing, Cham, 2015. 2