A GRAPH-BASED APPROACH FOR FEATURE EXTRACTION AND SEGMENTATION OF MULTIMODAL IMAGES

Geoffrey Iyer[1,2], **Jocelyn Chanussot**[1,2], **and Andrea Bertozzi**[1]

1. University of California, Los Angeles
2. Univ. Grenoble Alpes, GIPSA-lab

Contact Information:
Geoffrey Iyer
UCLA Mathematics Department
Box 951555
Los Angeles, CA 90095
Email: gsiyer@math.ucla.edu

Introduction

With the increasing availability of data we often come upon multiple datasets, derived from different sensors, that describe the same object or phenomenon. We call the sensors *modalities*, and because each modality represents some new degrees of freedom, it is generally desirable to use more modalities rather than fewer. For example, in the area of speech recognition, researchers have found that integrating the audio data with a video of the speaker results in a much more accurate classification [7, 9]. However, correctly processing a multimodal dataset is not a simple task [4]. Even the naive method of analyzing each modality separately still requires clever thinking when combining the results.

Here we consider the case where the datasets are co-registered (each modality contains the same number of points, and they share a common indexing). This often occurs in image processing problems, where the sets may be images of the same scene obtained from different sensors (as is the case in our experimental data), or taken at different times. Our method extracts features from the dataset by finding eigenvectors of the graph Laplacian, then uses standard data-segmentation algorithms on these features to obtain a final classification. In section 1 we give the general theory behind our method, and in 1 we show the results of the method applied to an optical/LIDAR dataset.

1 The Method

Graph Representation

We label the sets, $X^1, X^2, \dots, X^k$, with $\left|X^1\right| = \cdots = \left|X^k\right| = m$. Let and let d_i be the dimension of the data X^i. Have $X = (X^1, X^2, \dots, X^k) \subset \mathbb{R}^{n\times(d_1+\cdots+d_k)}$ be the concatenated dataset. We associate to X an undirected graph with (symmetric) weight matrix W. The nodes of the graph correspond to elements of X, and the edge weights $w_{ij} \geq 0$ represent the similarity between nodes, where large weights correspond to similar nodes, and small weights to dissimilar nodes. There are many different notions of "similarity" in the literature, and each has its own merits. In many applications, one defines

$$w_{ij} = \exp\left(-\left\|v_i - v_j\right\|/\sigma\right), \tag{1}$$

where σ is a scaling parameter. In this work we adapt this definition to apply to our multimodal dataset, as is explained below.

Multimodal Edge Weights

To calculate the weight matrix W, we first scale our sets $X^1, \dots, X^k$ to make distances in each set comparable. Let $X = (X^1, \dots, X^k) \subset \mathbb{R}^{n\times(d_1+\cdots+d_k)}$ be the concatenated dataset. Then for $\ell = 1, \dots, k$ define the scaling factor $\lambda_\ell = \text{stdev}\left(\left\|x_i^\ell - x_j^\ell\right\| \,;\, 1 \leq i, j \leq n\right)$. For a graph node $x \in X$, we define

$$\|x\| = \max\left(\frac{\left\|x^1\right\|}{\lambda_1}, \cdots, \frac{\left\|x^k\right\|}{\lambda_k}\right). \tag{2}$$

Then define the weight matrix W by letting $w_{ij} = \exp\left(-\left\|x_i - x_j\right\|\right)$.

Note that the $\|\cdot\|$ defined above is a norm on the concatenated dataset X. We specifically choose to use the maximum of the individual measurements to emphasize the unique information that each dataset brings. With this norm, two data points x_i, x_j are considered similar only when they are similar in each dataset.

The Graph Laplacian

Once we have created the weights, we define the *normalized graph Laplacian*.

$$L_{sym} = I - D^{-1/2}WD^{-1/2}, \tag{3}$$

Where D is the diagonal matrix consisting of degrees of nodes. For a thorough explanation of the properties of the graph Laplacian, see [6]. In this paper, we will use the connection between the graph Laplacian and the graph min-cut problem, as explained below.

Spectral Clustering

We rephrase the data clustering problem as a graph-cut-minimization problem of the similarity matrix W. A more detailed survey of the theory can be found in [10].

Given a partition of V into subsets $A_1, A_2, \dots, A_m$, we define the *graph N-cut*

$$\text{NCut}(A_1, \dots, A_m) = \frac{1}{2}\sum_{i=1}^{m}\frac{W(A_i, A_i^c)}{vol(A_i)}. \tag{4}$$

Heuristically, minimizing the N-cut serves to minimize the connection between distinct A_i, A_j, while still ensuring that each set is of a reasonable size.

Solving the graph min-cut problem is equivalent to finding an $n \times m$ indicator matrix u, where $u_{ij} = 1$ iff $x_i \in A_j$, and $u_{ij} = 0$ otherwise. Note that

$$\text{NCut}(A_1, \dots, A_m) = \text{Tr}\left(u^T L_{sym} u\right). \tag{5}$$

It has been shown in [3] that minimizing this energy exactly this problem is an $O(|V|^{m^2})$ process. As this is infeasible, we relax the problem to allow u to be an arbitrary orthogonal matrix. That is, we find

$$\text{argmin}_{u\in\mathbb{R}^{n\times m}}\text{Tr}\left(u^T L_{sym} u\right) \quad \text{where } u^T u = I. \tag{6}$$

This problem is solved by choosing the columns of u to be the m eigenvectors of L_{sym} corresponding to the m smallest eigenvalues. Using the eigenvectors u we define a map $X \to \mathbb{R}^m$. For each graph node $x_i \in X$ we get a vector $y_i \in \mathbb{R}^m$ given by the ith row of u. These y_i give the solution to the relaxed min-cut problem, and as such can be thought of as features extracted from the original dataset X.

To obtain a solution to the original min-cut problem, we then implement some classification algorithm on the y_i. Specifically, for spectral clustering we use k-means on the eigenvectors u to create a final classification into m classes.. Although k-means is unlikely to give an optimal classification, it is quite easy to implement, and the final results are strong enough to give a proof-of-concept.

Nyström Extension

Calculating the full graph Laplacian is computationally intensive, as the matrix contains n^2 entries. Instead we use Nyström's extension to find approximate eigenvalues and eigenvectors with a heavily reduced computation time [2, 5, 11].

Choose a subset $A \subset X$ of "landmark nodes", and have B its complement. Up to a permutation of nodes, we can write the weight matrix as

$$W = \begin{pmatrix} W_{AA} & W_{AB} \\ W_{BA} & W_{BB} \end{pmatrix}, \tag{7}$$

Nyström's extension approximates W as

$$W \approx \begin{pmatrix} W_{AA} \\ W_{BA} \end{pmatrix} W_{AA}^{-1} \begin{pmatrix} W_{AA} & W_{AB} \end{pmatrix}. \tag{8}$$

where the error of approximation is determined by how well the rows of of W_{AB} span the rows of W_{BB}. It is possible to find $|A|$ approximate eigenvectors of W using only the matrices W_{AA}, W_{AB}. This allows us to Xcompute and store matrices of size at most $|A| \times |X|$, rather than $|X| \times |X|$.

Results

We test our algorithm on three different optical/lidar datasets.

Data Fusion Contest 2015 Data [1]

Figure 1: RGB Data

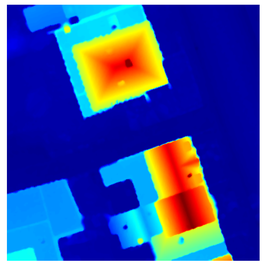

Figure 2: Lidar Data

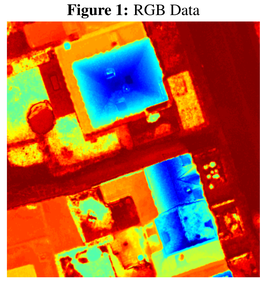

Figure 3: Example Eigenvector

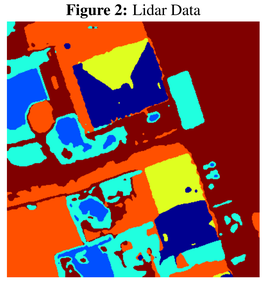

Figure 4: Segmentation Result

Umbrella Data [8]

Figure 5: RGB Data

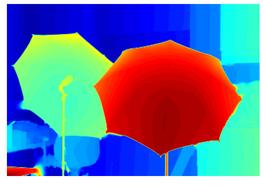

Figure 6: Lidar Data

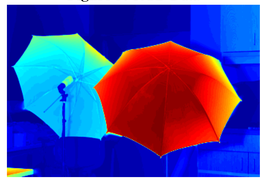

Figure 7: Example Eigenvector

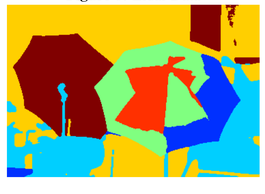

Figure 8: Segmentation Result

Jade Plant Data [8]

Figure 9: RGB Data

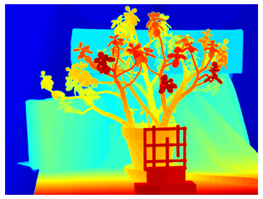

Figure 10: Lidar Data

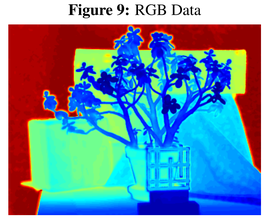

Figure 11: Example Eigenvector

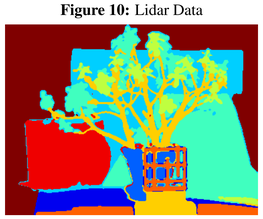

Figure 12: Segmentation Result

Conclusions

In conclusion, graph-based methods provide a straightforward and flexible method of combining information from multiple datasets. By considering the similarity between points in each individual dataset, we reduce the information from each modality into something directly comparable. This gives us a model that is more data-driven. Therefore the same algorithm could be applied in many different scenarios, with different types of data.

Once we have calculated and compared the different weight matrices, we can then create the graph Laplacian of the data and extract features in the form of eigenvectors. These features can then be used as part of many different data-segmentation algorithms. For this paper, we use k-means on the eigenvectors as a simple proof-of-concept, but other methods can easily be implemented. The main computational bottleneck is in calculation of the eigenvectors. Once we have these, there are many different viable classifications in the literature.

Our next area of interest is to further generalize the method by removing or weakening the co-registration assumption. In section 1 we only consider cases where the two images are of the same underlying scene, where pixels correspond exactly between images. We could not, for example, process two images taken from different angles. In image processing problems, co-registration is usually a reasonable assumption. However, if we remove this assumption the algorithm could be applied to data fusion problems across a much larger number of fields.

Acknowledgments

This work was supported by NSF grant DMS-1118971, ONR grant N00014-16-1-2119, NSF grant DMS-1417674, European Research Council (Grant no. 320684 - CHESS project), and CNRS (Grant no. PICS-USA 263484)

References

[1] M. Campos-Taberner, A. Romero-Soriano, C. Gatta, G. Camps-Valls, A. Lagrange, B. Le Saux, A. Beaupre, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, M. Ferecatu, M. Shimoni, G. Moser, and D. Tuia. Processing of extremely high-resolution lidar and rgb data: Outcome of the 2015 ieee grss data fusion contest #8211;part a: 2-d contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12):5547–5559, Dec 2016.

[2] 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.

[3] Olivier Goldschmidt and Dorit S. Hochbaum. A polynomial algorithm for the k-cut problem for fixed k. *Mathematics of Operations Research*, 19(1):24–37, 1994.

[4] Dana Lahat, Tülay Adali, and Christian Jutten. Challenges in Multimodal Data Fusion. In *22nd European Signal Processing Conference (EUSIPCO-2014)*, pages 101–105, Lisbonne, Portugal, September 2014.

[5] 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.

[6] Bojan Mohar. The laplacian spectrum of graphs. *Graph Theory, Combinatorics, and Applications*, 2:871–898, 1991.

[7] Gerasimos Potamianos, Chalapathy Neti, Guillaume Gravier, Ashutosh Garg, and Andrew W. Senior. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*, 91(9):1306–1326, Sept 2003.

[8] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *Proceedings of the 36th German Conference on Pattern Recognition*, september 2014.

[9] Farnaz Sedighin, Massoud Babaie-Zadeh, Bertrand Rivet, and Christian Jutten. Two Multimodal Approaches for Single Microphone Source Separation. In *24th European Signal Processing Conference (EUSIPCO 2016)*, pages 110–114, Budapest, Hungary, September 2016.

[10] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[11] 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.