Main Topic

Geoffrey Iyer

University of California

gsiyer@math.ucla.edu

April 3, 2017

Overview

1. Introduction: Multimodal Data
   - Multimodality
   - Manifold Alignment
   - Synthetic Example
2. Current Work: Graph-Based Multimodal Segmentation
   - Problem setup
   - Background: The Graph-Laplacian
   - Nyström
   - Results
3. Future Work
   - Manifold Alignment

Multimodal datasets

With the increasing availability of data, many applications involve data drawn from more than one source (called *modalities*).

Ex: In medicine, a patient is often tested and monitored in several different ways. This data is then synthesized (usually by a doctor) to make the final diagnosis and treatment decision.

The existence of multimodal data raises a new topic: how to solve standard machine learning problems with these sets?

Examples from the literature

Placeholder slide. Will fill in later.

I'm interested in citing [Song et al, 2012] and [Correa et al, 2010] and [Sedighin et al, 2016].

For each paper, give a short explanation of how multimodality is important.

Challenges in multimodality

Currently, most multimodal methods are developed specifically for one problem. The wide variety of possible data makes it difficult to create a general multimodal processing algorithm.

[Lahat et al, 2015]: "... a solution that is based on a sufficiently data-driven, model-free approach may turn out to be useful in very different domains."

Manifold alignment

One way to address multimodal data in a general sense is via manifold alignment.

For each modality, view the data as a manifold (have sets $X^1, X^2, \ldots, X^n$. $n =$ number of modalities). Correlate the different manifolds using any manifold alignment technique.

Often this consists of creating a *latent space* Y and maps $X^i \to Y$. Any comparisons between data can then be done by working with the set images in Y.

In this talk we are interested in segmentation/classification problems. We can accomplish this by running standard clustering algorithms on the set images in the latent space.

Manifold alignment: Methods from the literature

Some examples from the literature:

- Canonical Correlation Analysis (either linear or with nonlinear kernel) [Yeh et al, 2014]
- Graph-based methods, linear projection maps $X^i \to Y$ [Wang et al, 2013]
- Similar graph-based methods with an added nonlinear kernel [Tuia et al, 2016]

Manifold alignment: Methods from the literature

Common theme: We expect the different modalities to contain redundant information (likely expressed in different ways in the raw data). Create the latent space by finding and correlating these redundant parts.

Add a nice picture from [Tuia et al, 2016] here.

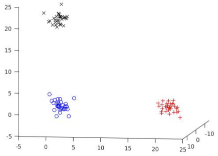
Manifold alignment: Our goal

Our goal: We believe that these methods do not make full use of the data. Along with the redundant information, we expect each modality to contain some unique information that is not present in the other sets.

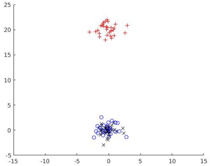
We want to create a general algorithm, similar to the above, that will process multimodal data in a way that accounts for both the redundant and the unique information from each source.

Synthetic example: Data

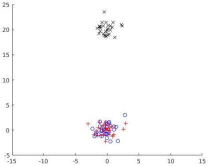
Synthetic example:
Ground truth $=$ 3 point clouds in $\mathbb{R}^3$ (20 points per cloud).
Modality 1 $=$ projection onto xy-plane.
Modality 2 $=$ projection onto xz-plane.



Ground Truth



Modality 1



Modality 2

Synthetic Example: Result of CCA

Result of CCA algorithm from [Yeh et al, 2014] applied to the data:

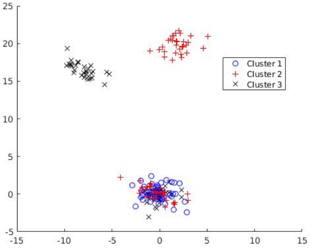


Image of clusters in latent space

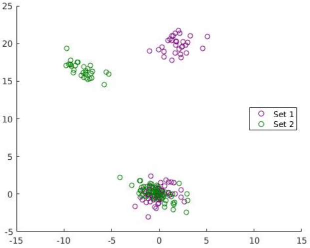


Image of data in latent space

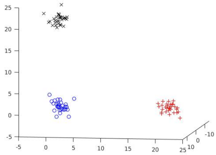
The algorithm identifies the redundancy (the blue clusters), but cannot properly use the unique information from each set.

Our work: Co-registration assumption

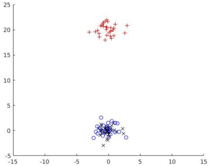
For our first attempt at solving this problem, we make the assumption that our data is *co-registered*.

That is, each modality contains the same number of observations, and there is a pre-understood correspondence between them.

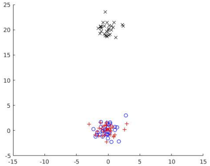
Ex: The synthetic example in slide 10. Each modality contains the same points, viewed in different ways.



Ground Truth



Modality 1

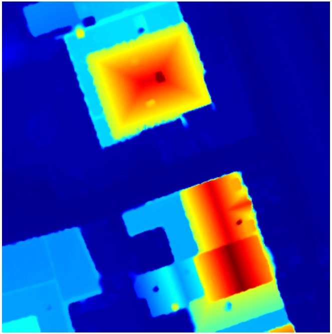


Modality 2

Our work: Co-registration assumption

Ex: RGB + Elevation map of residential neighborhood in Belgium. Found in [Bampos-Taberner et al, 2016].

RGB Data



Lidar Data

Notation

From each modality, we obtain a data set X^k.

n = number of observations.

d_j = dimension of set X^k. (So X^k can be viewed as an $n \times d_k$ matrix).

From co-registration assumption: i-th point in X^{k_1} corresponds to i-th point in X^{k_2}. From this we create the concatenated set $X = (X^1, X^2, \ldots, X^\ell) \subseteq \mathbb{R}^{n \times (d_1 + \cdots + d_\ell)}$.

x_i = element i from X. x_i^k = element i from X^k.

Weight Matrix: Background

For each pair of data points $x_i, x_j \in X$, define a *weight* w_{ij} that measures the similarity between the points.

A large weight corresponds to similar nodes. A small weight corresponds to dissimilar nodes.

Represents our data as an $n \times n$ weight matrix W.

Common similarity measure from the literature: RBF kernel

$$w_{ij} = \exp\left(- \|x_i - x_j\| / \sigma\right).$$

Our similarity measure is influenced by this idea.

Multimodal Weight Matrix

For each modality X^k, calculate the distance matrix E^k via

$$E_{ij}^k = \left\| x_i^k - x_j^k \right\|.$$

$\|\cdot\|$ may be chosen based on the details of the modality. In our work we use the 2-norm in all cases.

Scale each distance matrix based on the standard deviation

$$\bar{E}^k = \frac{E^k}{\mathrm{std}\left(E^k\right)}.$$

Multimodal Weight Matrix

Define

$$w_{ij} = \exp\left(-\max\left(\bar{E}^1_{ij}, \ldots, \bar{E}^k_{ij}\right)/\sigma\right).$$

Heuristics:

- Standard devation scaling allows us to directly compare $\bar{E}^{k_1}, \bar{E}^{k_2}$ with reasonable results.
- Because of the max, elements are similar under this measure only if they are similar in each modality.
- Could use 2-norm instead of max. This would have a smoothing property. Elements that are similar in almost every modality (but not all) could be considered similar when using 2-norm.

Graph min cut

Once we create W, we can rephrase the data clustering problem as a graph-cut-minimization problem.

Given a partition of X into subsets $A_1, A_2, \ldots, A_m$, we define the *ratio graph-cut*

$$\mathrm{RatioCut}(A_1, \ldots, A_m) = \frac{1}{2} \sum_{i=1}^{m} \frac{W(A_i, A_i^c)}{|A_i|}.$$

Where

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij}.$$

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

Nystrom

Placeholder slide. There will probably be a lot of slides for this.

Results

Placeholder slide. There will probably be a lot of slides for this.

Future Work

Placeholder slide. There will probably be a lot of slides for this.

References

M. Song and D. Tao and C. Chen and J. Bu and J. Luo and C. Zhang (2012)
Probabilistic Exposure Fusion
IEEE Transactions on Image Processing 21(1), 341-357

Correa, N. M. and Eichele, T. and Adali, T. and Li, Y. O. and Calhoun, V. D. (2010)
Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI
Neuroimage 50(4) 1438-1445

Sedighin, Farnaz and Babaie-Zadeh, Massoud and Rivet, Bertrand and Jutten, Christian (2016)
Two Multimodal Approaches for Single Microphone Source Separation
24th European Signal Processing Conference 110-114

Lahat, Dana and Adalı, Tülay and Jutten, Christian (2015)
Multimodal Data Fusion: An Overview of Methods, Challenges and Prospects
Proceedings of the IEEE 103(9), 1449-1477