Improving Multiset Canonical Correlation Analysis in High Dimensional Sample Deficient Settings

Nicholas Asendorf, Ph.D.
asendorf@umich.edu

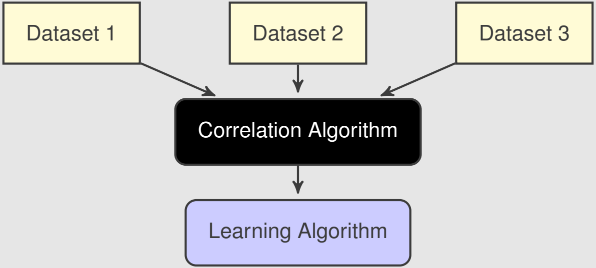
Prof. Raj Nadakuditi
rajnrao@umich.edu

Department of Electrical Engineering and Computer Science
University of Michigan

Asilomar Conference on Signals, Systems, and Computers

November 9, 2015

Motivation



Goal

Develop theoretically justified, robust correlation algorithms for multi-dataset fusion

A Myriad of Applications

Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

Multiple Datasets

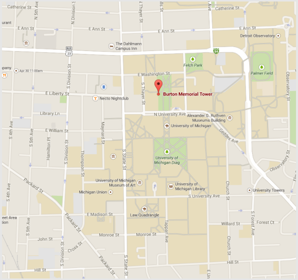
- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.



A Myriad of Applications

Multiple Datasets

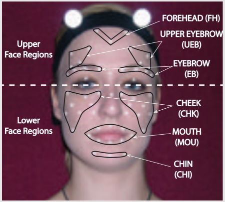
- Audio-Video
- Audio-Audio

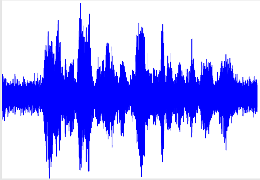
Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.





A Myriad of Applications

Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

A Myriad of Applications

Multiple Datasets

- Audio-Video
- Audio-Audio

Machine Learning

- emotion identification
- shopping predictions
- music genre classification

Medical Signal Processing

- MRI, fMRi, EEG, MEG, etc.

- disco influences
- danceable grooves
- repetitive melodic phrasing
- extensive vamping
- minor key tonality

MCCA - Multiset Canonical Correlation Analysis

Notation

- data: $y_1 \in \mathbb{C}^{d_1 \times 1}, \ldots, y_m \in \mathbb{C}^{d_m \times 1}$
- covariance matrices: $R_{ij} = \mathbb{E}\left[y_i y_j^H\right]$
- canonical vectors: $x_1, \ldots, x_m$
- canonical variates: $w_1, \ldots, w_m$, $w_i = x_i^H y_i$

$$\Phi(x) = E[ww^H] = \begin{bmatrix} x_1^H R_{11} x_1 & \ldots & x_1^H R_{1m} x_m \\ \vdots & \ddots & \vdots \\ x_m^H R_{m1} x_1 & \ldots & x_m^H R_{mm} x_m \end{bmatrix}$$

Optimization Problem

$$\begin{array}{ll} \underset{x}{\text{optimize}} & J(\Phi(x)) \\ \text{subject to} & h(x, R) \end{array}$$

MCCA - Multiset Canonical Correlation Analysis

Notation

- data: $y_1 \in \mathbb{C}^{d_1 \times 1}, \ldots, y_m \in \mathbb{C}^{d_m \times 1}$
- covariance matrices: $R_{ij} = \mathbb{E}\left[y_i y_j^H\right]$
- canonical vectors: $x_1, \ldots, x_m$
- canonical variates: $w_1, \ldots, w_m$, $w_i = x_i^H y_i$

$$\Phi(x) = E[ww^H] = \begin{bmatrix} x_1^H R_{11} x_1 & \ldots & x_1^H R_{1m} x_m \\ \vdots & \ddots & \vdots \\ x_m^H R_{m1} x_1 & \ldots & x_m^H R_{mm} x_m \end{bmatrix}$$

MAXVAR Optimization Problem

$$\begin{array}{ll} \underset{x_1,\ldots,x_m}{\text{argmax}} & \rho_{\text{mcca}} = \lambda_1\left(\Phi(x)\right) \\ \text{subject to} & x_i^H R_{ii} x_i = 1 \end{array}$$

MCCA Solution and Empirical Solution

Notation

- $R = [R_{ij}]_{i,j=1}^{m}$
- $R_D = \textbf{blkdiag}(R_{11}, \ldots, R_{mm})$
- $C_{\text{mcca}} = R_D^{-/12} R R_D^{-1/2}$

MCCA Solution

$$\boxed{\rho_{\text{mcca}}^{(j)} = \lambda_j\left(C_{\text{mcca}}\right)}$$

MCCA Solution and Empirical Solution

Notation

- $R = \left[R_{ij}\right]_{i,j=1}^{m}$
- $R_D = \textbf{blkdiag}(R_{11}, \ldots, R_{mm})$
- $C_{\text{mcca}} = R_D^{-/12} R R_D^{-1/2}$

MCCA Solution

$$\rho_{\text{mcca}}^{(j)} = \lambda_j \left(C_{\text{mcca}}\right)$$

Empirical MAXVAR

- Covariance matrices R_{ij} are typically unknown in practice
- Training data $Y_j = \left[y_1^{(j)}, \ldots, y_n^{(j)}\right]$, $j = 1, \ldots, m$
- Sample covariance matrices $\widehat{R}_{ij} = \frac{1}{n} Y_i Y_j^H$
- Form $\widehat{R}$ and $\widehat{R}_D$
- $\widehat{C}_{\text{mcca}} = \widehat{R}_D^{-/12} \widehat{R} \widehat{R}_D^{-1/2}$

Proposed correlation statistic

$$\widehat{\rho}_{\text{mcca}}^{(j)} = \lambda_j \left(\widehat{C}_{\text{mcca}} - I\right)$$

Empirical MCCA Insights

Empirical MAXVAR Equivalency

- Let $Y_j = \widehat{U}_j \widehat{\Sigma}_j \widehat{V}_j$ be SVD of dataset j
- Let $\widetilde{U} = \textbf{blkdiag}(\widehat{U}_1, \ldots, \widehat{U}_m)$
- Let $\widetilde{V} = [\widehat{V}_1, \ldots, \widehat{V}_m]$
- $\widehat{C}_{\text{mcca}} = \widetilde{U}\widetilde{V}^H\widetilde{V}\widetilde{U}$
- $\lambda_j\left(\widehat{C}_{\text{mcca}}\right) = \lambda_j\left(\widetilde{V}^H\widetilde{V}\right)$

Insights

- This uses *ALL* right singular vectors
- In many applications, low-rank setting
- In low-SNR, low-sample regime, singular vectors may be inaccurate
- We can use insights from random matrix theory (RMT) to quantify singular vector accuracy to improve MCCA
- Let $\widehat{k}_j$ be RMT estimates of rank of each dataset

Idea - Trim right singular vectors

- $\mathring{V}_i = \widehat{V}_i\left(:, 1 : \widehat{k}_i\right)$
- $\mathring{V} = \left[\mathring{V}_1, \ldots \mathring{V}_m\right]$

New IMCCA Matrix

$$\widehat{C}_{\text{imcca}} = \mathring{V}^H \mathring{V}$$

Proposed correlation statistic

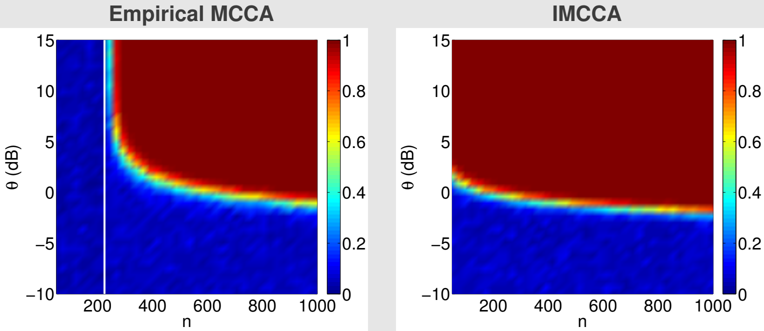
$$\widehat{\rho}^{(j)}_{\text{imcca}} = \lambda_j\left(\widehat{C}_{\text{imcca}} - I\right)$$

Insights

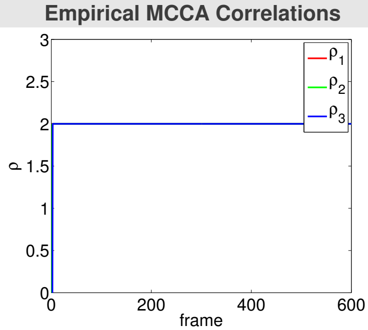
- Eigenvalue detects correlation
- Eigenvector reveals structure

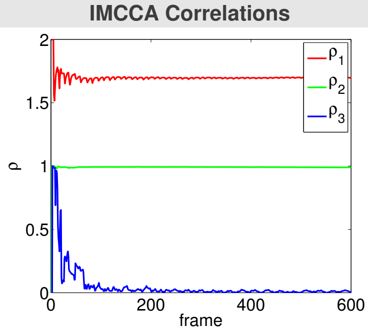
Numerical Simulation

- rank-1 setting, $m = 3$, $d_1 = d_2 = d_3 = 150$
- $y_j^{(i)} = U_j s_j^{(i)} + z_j^{(i)}$
- $z_j \sim \mathcal{CN}(0, I)$, $s_j^{(i)} \sim \mathcal{CN}(0, \theta)$
- $\mathbb{E}\left[s_j s_i\right] = 0.9$
- Plot KS-statistic of $\widehat{\rho}_{\text{mcca}}^{(j)}$ and $\widehat{\rho}_{\text{imcca}}^{(j)}$ for signal-plus-noise vs. noise only



Empirical MCCA and IMCCA Demonstration





Conclusion

Take Home Message

$\boxed{\text{Trim then fuse } \textbf{\textit{NOT}} \text{ fuse then trim}}$

Contributions

- Proposed IMCCA algorithm
- Proposed statistics to detect latent correlations in more than 2 datasets
- Acheived better performance in low-SNR, high dimensional setting