Tensor Canonical Correlation Analysis (TCCA)

Anshul Bhatia (0801CS171011)

gs0801cs171011@sgsits.ac.in

December 13, 2020

Contents

Chapter 1

Introduction

1.1 What is a tensor?

A tensor is a multidimensional or N-way array. The array dimensionality, the value of N, specifies the tensor order or the number of tensor modes. A one-dimensional array a of p elements, that is a p x 1 vector, is considered a tensor of order one, and a two-dimensional array or matrix A of size p x q is a tensor of order two. Tensors of order three or more are generally called high-order tensors.

1.2 Canonical Correlation Analysis

Canonical correlation analysis (CCA), finds bases for two random variables (or sets of variables) so that the coordinates of the variable pairs projected on these bases are maximally correlated

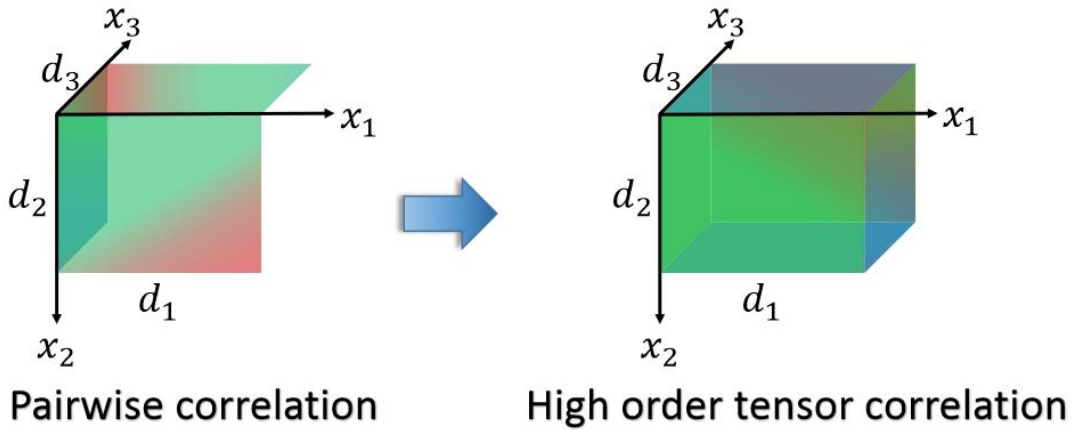
Given two sets of column vectors $x_{1n} \in R^{d1}$, $x_{2n} \in R^{d2}$, n = 1,. N. The objective of CCA is to and a pair of projections (usually called canonical vectors) h_1, h_2, such that correlations between the two vectors of canonical variables z_1 2 R^N and z_2 2 R^N with each $z_{1n} = x^T_{1n}h_1$, $z_{2n} = x^T_{2n}h_2$, are maximized. The optimization problem is thus given by

$$\underset{\mathbf{z}_1,\mathbf{z}_2}{\operatorname{argmax}} \rho = \operatorname{corr}(\mathbf{z}_1, \mathbf{z}_2) = \frac{\mathbf{h}_1^T C_{12} \mathbf{h}_2}{\sqrt{\mathbf{h}_1^T C_{11} \mathbf{h}_1 \mathbf{h}_2^T C_{22} \mathbf{h}_2}},$$

where $C_{11} = X_1X_1^T$, $C_{22} = X_2X_2^T$ are data variance matrices, and $C_{12} = X_1X_2^T$ is the covariance matrix. Here, X_1 2 R^{d1} ^{N} and X_2 2 R^{d2} ^{N} are the stacked data matrices.

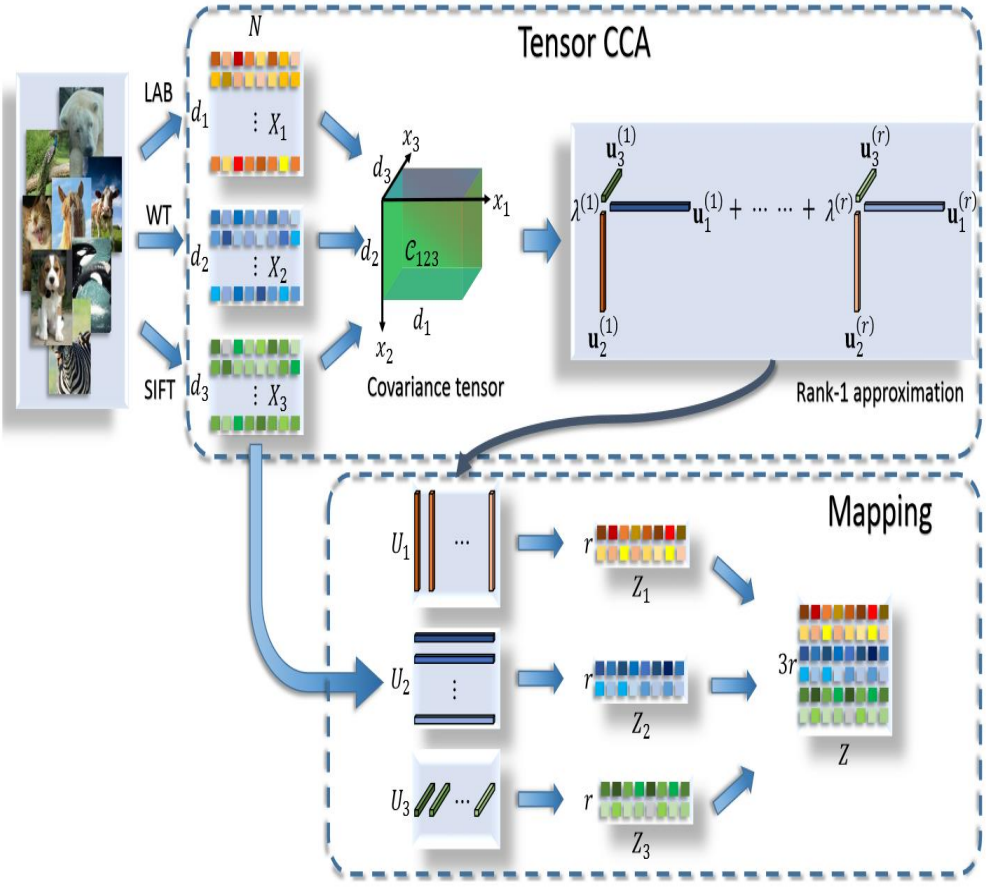
1.3 Tensor CCA

Tensor CCA (TCCA) naturally generalizes CCA to handle the data of an arbitrary number of views by analyzing the covariance tensor of the different views. TCCA aims to directly maximize the canonical correlation of multiple (more than two) views.



Chapter 2
Mathematical Formulation

A multiple feature matrix $\{X_p \in R^{dp \times N}\}^m_{p=1}$ .Here, m is set at 3 for intuitive illustration without loss of generality. The different sets of features are then used to calculate the data covariance tensor C_{123}, which is subsequently decomposed as a weighted sum of rank-1 tensors, i. e. $C_{123} \approx \sum^r_{k=1} \lambda^{(\kappa)} u_1^{(k)} o u_2^{(k)} o u_3^{(k)}$, where $r \leq \min(d1, d2, d3)$ is the reduced dimension and $\circ$ is the tensor (outer) product. The vectors $\{u_p^{(k)}\}^r_{k=1}$ are stacked as a transformation matrix U_p, which is used to map the original high dimensional features into the low dimensional common subspace. The projected features $\{Z_p\}^m_{p=1}$ are concatenated as the final representation of the instances.



Chapter 4
Algorithm

1. Input list containing n-dimensional arrays each representing a view.
2. Compute co-variance matrix of each view:

$$C_{pp} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_{pn}\mathbf{x}_{pn}^{T}, p = 1, \ldots, m,$$

3. Compute the covariance tensor among all views:

$$\mathcal{C}_{12\ldots m} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_{1n} \circ \mathbf{x}_{2n} \circ \ldots \circ \mathbf{x}_{mn},$$

4. Maximize ρ which is equivalent to computing p-mode tensor matrix product:

$$\underset{\{\mathbf{h}_p\}}{\operatorname{argmax}}\, \rho = \mathcal{C}_{12\ldots m} \times_1 \mathbf{h}_1^T \times_2 \mathbf{h}_2^T \ldots \times_m \mathbf{h}_m^T$$

$$\text{s.t. } \mathbf{h}_p^T C_{pp} \mathbf{h}_p = 1, p = 1, \ldots, m.$$

which is equivalent to

$$\underset{\{\mathbf{u}_p\}}{\operatorname{argmax}}\, \rho = \mathcal{M} \times_1 \mathbf{u}_1^T \times_2 \mathbf{u}_2^T \ldots \times_m \mathbf{u}_m^T$$

$$\text{s.t. } \mathbf{u}_p^T \mathbf{u}_p = 1, p = 1, \ldots, m,$$

where

$$\mathbf{u}_p = \tilde{C}_{pp}^{1/2} \mathbf{h}_p$$

$$\mathcal{M} = \mathcal{C}_{12\ldots m} \times_1 \tilde{C}_{11}^{-1/2} \times_2 \tilde{C}_{22}^{-1/2} \ldots \times_m \tilde{C}_{mm}^{-1/2}$$

$$\tilde{C}_{pp} = C_{pp} + \epsilon I.$$

5. Perform rank approximation of M.

6. We obtain the projected data for the p^{th} view Z_p as

$$Z_p = X_p^T \tilde{C}_{pp}^{-1/2} U_p.$$

Chapter 4

Documentation of API

4.1 Package organization

class TensorCCA.TensorCCA(max_iter)

Parameters

Views: List of numpy N-dimensional array, where each n-dimensional array represents a view

reduce_to_dim: int. Dimension to which each view will be reduced

max_iter: an integer, (default 500) the maximum number of iterations of the ALS Algorithm

4.2 Methods

fit(Views, reduce_to_dim): Fit method for given views

Returns list of numpy N-dimensional arrays. Each element represents a view.

views_hat (Views): Mean centering of data

Returns list of N-dimensional arrays. Each element represents a mean centered view.

cov_matrix(Views): Calculate covariance matrix of each view

Returns list of covariance of each matrix.

covariance_tensor(Views): Calculate covariance tensor of given views

Returns a numpy N-dimensional array

root_inverse(Cpp): Calculate root inverse of covariance matrix of a view

Returns numpy N-dimensional array.

ttm(cov_ten, var_matrix_inverse): Calculate tensor times matrix i.e. mode-i product

Returns numpy N-dimensional array.

tcca(Views, var_matrix, cov_ten, reduce_to_dim): Calculate canonical vectors

Returns a list of numpy N-dimensional array

transform(Views): Reduce dimensions of each view

Returns a list of numpy N-dimensional array

Chapter 5
Example

Example :

```
from TensorCCA import TensorCCA

U = [np.random.randint(1,4,(3,i)) for i in (5, 3, 4)]
tcca_object = TensorCCA(max_iter = 500)
fit_data = tcca_object.fit(U,reduce_to_dim = 2)
new_views = tcca_object.transform(U)
print(new_views)
```

Chapter 6

Learning Outcomes

- Learned about tensors and algebra of tensors.
- Concept of Canonical Correlation Analysis.
- Limitations of traditional canonical correlation analysis and need of Tensor Canonical correlation analysis (TCCA).
- Maximize correlation of n-views(together) (where n = 2 in case of traditional CCA).
- Study a research paper and related terminologies and implement it.

References

- Tensor Canonical Correlation Analysis for Multi-view Dimension Reduction https://arxiv.org/pdf/1502.02330.pdf

- Canonical Correlation Analysis (CCA) Based Multi-View Learning: An Overview by Chenfeng Guo and Dongrui Wu https://arxiv.org/pdf/1907.01693.pdf