ORTHOGONAL CANONICAL CORRELATION ANALYSIS

(OCCA)

Contents

Chapter 1

Introduction

1.1 Multi-View Learning:

With this growing era of technology, the generation of data/ big data has taken a tremendous leap. The generated data often corresponds to different aspects or views of the same information and analyzing and fusing these interrelated multiple views together results in a more generic performance of any learning model. The whole idea behind multiview learning is to fuse the features from multiple aspects of a single information in such a manner which along with capturing the information from the individual datasets, also captures the information regarding the correlation between those multiple datasets.

The following paper is based on OCCA (Orthogonal Canonical Correlation Analysis), a variant of CCA(Canonical correlation Analysis) for fusion of information from multiple views.

1.2 Canonical Correlation Analysis

Canonical Correlation Analysis is a multivariate statistical analysis method for linear dimensionality reduction and feature fusion.It represents two dimensional views of a single instance as a linear representation of the individual correlations of each of these views and also interrelated correlation of the two datasets. Suppose, X and Y corresponds to two separate views of a single image, let say X corresponds to its front view and Y corresponds to left view, then CCA tries to measure the relationship among these two views in terms of the combine correlation of view X with X, view Y with Y and view X with Y and aims at retaining the maximum correlation between these two different views. For establishing this constraint, that there is maximum correlation between the different views/aspects, a new set of orthogonal bases or projection directions,known as orthogonal canonical projection pairs are computed along which if the views are projected have maximum correlation associated between them. An important feature of these canonical covariates or projection pairs are orthogonal to one another and also the information which is captured by prior covariate pairs is not contained or represented by the later projection pairs. These canonical projection pairs form a basis to represent the original aspects in a reduced dimensional space where maximum correlation is associated between each of these views. The whole process of feature fusion through CCA could be defined in three steps:

* Feature Extraction: Extract two different set of feature vectors from the original views.
* Canonical Projection Pair: Based on the extracted features, extract canonical projection pairs such that which maximizes the correlation between these views.
* Feature Fusion: Projecting the original features onto these canonical pairs and fusing them either through serial or parallel fusion.

1.3. Orthogonal Canonical Correlation Analysis:

Considering the traditional method or classical CCA, the generated canonical covariates are independent of any transformation and are less sensitive to data distribution and noises in the data. These inherit features of the canonical projection pairs due to the orthogonality of the covariates. The canonical covariates obtained from CCA are ensured to follow conjugate orthogonality property,i.e., pairwise orthogonality of each of the covariates, as they are obtained from the eigenvalue decomposition of maximization condition of the canonical correlation coefficient of the two views subject to the constraints that the maximum correlation is contained between each pair. Since the projection pairs are a result of the eigenvalue decomposition, so by default they will be orthogonal in nature.

But if the same maximization method is applied on a view set with a small number of sata samples, then the canonical covariates obtained from the eigen decomposition tends to violate the conjugated orthogonality property. This problem is known as the Small Sample Size (SSS) problem. The reason of this violation of orthogonality in a small sample data adheres to the fact that since the formulation of the covariance matrix either between the same view or different, captures the information of patterns or relation between those two data and if in a similar manner covariance matrix is computed between such views having less number of data samples, then the variance or the relation between the various features of those dataset cannot be truly captured by the covariance matrix. In simple terms, the covariance matrix cannot represent all the variance between the features. So if eigen decomposition is done on such such a covariance matrix, then the eigenvectors obtained do not corresponds to the true direction of maximum variance and are not necessarily independent of each other, resulting in non-orthogonal projection vectors which are affected by various data transformations and noises. Therefore in SSS problem cases, the projection vectors or canonical covariates obtained are not optimal.

In contrast to classical CCA, OCCA(Orthogonal Canonical Correlation Analysis) always produces orthogonal projection pairs by imposing an additional orthogonality constraint on the maximization function of the canonical correlation coefficient along with the prior ones.

This variant of classical CCA was first proposed by Xiao-Bo Shen, Quan-Sen Sun, Yun-Hao Yuan in 2013[1]. Similar to classic CCA, OCCA is also a linear multivariate statistical analysis and dimensionality reduction technique with the same maximization condition as the CCA and imposed an additional constraint to ensure that all the new generated canonical covariates are perpendicular to the older ones and thus ensuring conjugated orthogonality property.

Instead of postprocessing and orthogonalizing the canonical covariates obtained from the CCA, the OCCA tries to maintain this pairwise orthogonality constraint by performing twin eigen decomposition of the optimization condition instead of a single step of normal eigenvalue decomposition. Similar to the classical CCA where the maximum number of projection pairs or bases that could be generated are less than or equal to the minimum of the rank of the two views. Orthogonality constraint is implemented by checking the orthogonality of a current projection pair to the older ones. This means that for the first canonical projection pair there will be no orthogonality check and for the *ith* the orthogonality will be checked with each of the *jth* pair, where *j* varies from *{0,1,2,.....,i-1}.*

The whole process of Orthogonal Canonical Correlation could be summarized belw in the form of 2 models:

* Model 1: Finding the first canonical projection pair from the classical CCA approach by just imposing maximum correlation constraint on the canonical correlation coefficient maximization function.
* Model 2: Iteratively finding the remaining canonical projection pair by ensuring orthogonality check with all the previously generated canonical vectors.

Notations:

Before moving ahead with the mathematical formulation of the, below are the notations which will be carried on throughout this report:

|  |  |
| --- | --- |
| **X , Y** | 2 different views of the same information whose multivariate analysis has to be made both containing *n* number of samples in each of them. |
| ***p*** | Number of features in X view such that X is *n x p* |
| ***q*** | Number of features in Y view such that Y is *n x q* |
| 𝛂 | *p* dimensional projection direction/canonical variate of X view |
| 𝛃 | *q* dimensional projection direction/canonical variate of X view |
| ***Sxx*** | Covariance matrix of X and X |
| ***Syy*** | Covariance matrix of Y and Y |
| ***Sxy*** | Covariance matrix of X and Y |
| 𝓟 | Canonical Correlation Coefficient |
| ***Wx, Wy*** | Orthogonal canonical projection matrix of X and Y |

Chapter 2

Mathematical Formulation

2.1 Formulation

The OCCA is aimed at finding a set of projection vectors or directions for each of the views such that the correlation between projections of each of the views on these vectors maximizes.

Let X and Y be the two different views of the same information each with p and q features and having n samples in each. Let ⍺ and 𝛽 be the respective projection vectors or canonical variate pairs for each of the view such that projection of these views Z1 = X.⍺ and Z2 = X.𝛽 on a common latent space have maximum correlation between them. The set of such projection vectors ⍺ and 𝛽 for each of the views X and Y is known as canonical correlation projection matrix Wx and Wy each of which are represented as:

Wx = { , ,, . . . . . , , ,} (2.1)

Wy = { , , , . . . . . , , , } (2.2)

where , *k* is the number of canonical covariate pairs that are needed to be extracted.

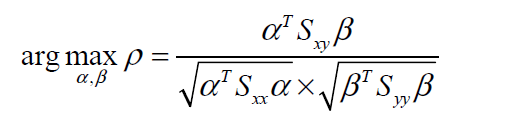
Let, Sxx, Syy and Sxy be the respective covariance matrices given by:

= /N = /N = /N

(2.3) (2.4) (2.5)

Each of the dimensions (p x p), (q x q) and (p x q) respectively.

The following maximization problem defines the OCCA objective functions:

 (2.6)

The set of ⍺ and 𝛽 that maximizes the above equation corresponds to Wx and Wy. The maximization of above equation in two steps