Canonical Correlation Analysis based on Contrastive Learning

Anonymous submission

Abstract

High-dimensional data, abundant in information, plays a pivotal role across various fields. Yet, its sheer volume of redundant features can lead to the curse of dimensionality. Dimensionality reduction is a well-acknowledged remedy for this predicament. Canonical Correlation Analysis (CCA) stands out as a classic multi-view reduction technique, primarily aiming to identify low-dimensional features that maximize correlation between two views. However, when distinguishing the target dataset from others, CCA tends to emphasize shared features, often overlooking essential differential ones. This oversight compromises the efficacy of the low-dimensional features. To address such issues, we propose a novel method called contrastive Canonical Correlation Analysis (cCCA), which takes the advantage of Contrastive Learning (CL) in solving the problem. Concretely, we introduce the concepts of target datasets and background datasets similar to CL. cCCA can seek the projections which not only maximize the correlation between the two views of the target dataset and the variance within a single view of the target dataset, but also minimize the variance of the background dataset. This approach eliminates interference caused by the background portion in the target dataset, extracting concealed but significant features from the target dataset. Experiments on artificially synthesized dataset and two real datasets were conducted in this paper, and the results further validate the effectiveness of the cCCA algorithm. Datasets and code will be made available.

Keywords: Data Dimensionality Reduction, Canonical Correlation Analysis, Contrastive Learning, contrastive Principal Component Analysis.

Introduction

Research Background and Significance

High-dimensional data are prevalent in many real-world applications, such as machine learning, statistical analysis, and data mining. Typical examples of high-dimensional data include images, gene expression sequences, and time series. Compared to low-dimensional data, high-dimensional data encompasses richer information. However, it also contains a larger number of redundant features that can cause interference, which often limits the processing and analysis of high-dimensional data. To address this issue, dimensionality reduction (DR) of high-dimensional data is necessary. Dimensionality reduction methods can be broadly divided into two categories: feature selection and feature extraction. Feature selection involves choosing features from the existing high-dimensional data that best reflect its effective information. In contrast, feature extraction maps the original high-dimensional data to a low-dimensional space, generating new features based on the existing ones. These new features should stably and adequately express the effective information of the original high-dimensional data. The dimensionality reduction method studied in this paper falls under the category of feature extraction. The most representative algorithms in this category are Principal Component Analysis (PCA) [Pearson 1901] [Hotelling 1933] [Kwak 2008] [Nie, Yuan, and Huang 2014] [Liao et al. 2018] and Canonical Correlation Analysis (CCA) [Hotelling 1936] [Hardoon, Szedmak, and Shawe-Taylor 2004].

Research Status

PCA was first introduced by K.Pearson in 1901 [Pearson 1901] and refined by H.Hotelling in 1933 [Hotelling 1933]. In 1936, he extended PCA, which was initially designed for single-view data, to dual-view data, introducing the dimensionality reduction algorithm CCA[Hotelling 1936]. The new algorithm focuses on building a linear correlation between two views. It aims to find the directions that maximize the correlation between the two sets of projected representations in the low-dimensional space. Classical CCA has numerous improved algorithms. Kettenring extended CCA to multi-view data and proposed Multi-view Canonical Correlation Analysis (MCCA) [Kettenring 1971]. CCA is a linear DR algorithm and can only search for linear correlations between views, making it ineffective in non-linear situations. Kernel Canonical Correlation Analysis (KCCA) [Melzer, Reiter, and Bischof 2001], which combines kernel methods with CCA, effectively solving the problem of non-linearity. Since CCA does not utilize the class information contained in the data, it is categorized as unsupervised. To enhance the separability of low-dimensional features, Sun et al. proposed Discriminant Canonical Correlation Analysis (DCCA) [Sun et al. 2008], which maximizes intra-class correlation and minimizes inter-class correlation to extract features beneficial for classification. Furthermore, Discriminant Multiple Canonical Correlation Analysis (DMCCA) extends DCCA to multi-view data [Gao et al. 2012] [Gao et al. 2018]. Sun combined CCA with manifold learning, introducing Locality Preserving Canonical Correlation Analysis (LPCCA) [Sun and Chen 2007], which not only retains local structural information of the data but also achieves correlation between two views. Deep Canonical Correlation Analysis (Deep CCA) combines CCA with deep learning, which employs a deep neural network for non-linear transformation to achieve high linear correlation after projection [Andrew et al. 2013].

CCA and its improved algorithms are crucial in various fields including computer vision [Luo and Tjahjadi 2020], facial expression recognition [Qi et al. 2022], medical testing [Shi et al. 2021], and mechanical engineering [Ibrahim and Sidiropoulos 2020] [Nielsen 2002]. The aforementioned DR algorithms achieve desirable results in different applications. However, they can only extract features on a given dataset. In practical, it is often necessary to discern the differences between the given dataset and other datasets. For instance, when collecting gene sequence data from individuals of different races [Garte 1998] [Abid et al. 2018], which includes genetic information from individuals expressing cancer and those who don't, researchers aim to emphasize the part of the gene sequence related to cancer expression (e.g., genetic features that can identify cancer subtypes). If the mentioned dimensionality reduction algorithms are directly applied to this dataset, the extracted features might only reflect demographic information of the individual, i.e., racial information, rather than the genetic expression information related to cancer. This is because, compared to the effects on the genetic variation level (racial information), the expression effects of cancer are not pronounced, resulting in unsatisfactory outcomes. To better emphasize the concealed yet crucial features within a dataset, James et al. introduced Contrastive Learning (CL) and combined it with Latent Dirichlet Allocation (LDA) and Hidden Markov Models (HMMs), resulting in two improved algorithms[Zou et al. 2013].

Contrastive Learning (CL) refers to the target dataset for analysis as the "target dataset" (as shown in Figure 1.1(a)). The portion of the target dataset that is of interest is termed the "foreground" (e.g., the dolphins in Figure 1.1(a)), while the portion that isn't of interest but has a significant impact is termed the "background" (e.g., the ocean in Figure 1.1(a)). CL introduces another dataset with features similar to the background, termed the "background dataset" (as illustrated in Figure 1.1(b)). The purpose of CL is to leverage the background dataset to assist in dimensionality reduction, eliminate the interference caused by the background in the target dataset, and highlight the foreground of the target dataset.

In 2017, Abubakar et al. combined Contrastive Learning with PCA and proposed contrastive Principal Component Analysis (cPCA) [Abid et al. 2018]. This method aims to find a projection with the maximized variance of the projected target dataset, while the variance of the projected background dataset is minimized. This emphasizes the unique features in the target dataset. Contrastive Latent Variable Model (cLVM) combines contrastive learning with latent variable models [Severson, Ghosh, and Ng 2019]. It encompasses structures shared by both the target dataset

(a) Target Dataset (b) Background Dataset

Figure 1: Examples for target dataset and background dataset.

and the background dataset, while also containing specific structures of the target dataset. Based on cPCA, the concept of clustering is incorporated, leading to the proposal of contrasting clusters in PCA (ccPCA) [Fujiwara, Kwon, and Ma 2019]. The contrastive Multivariate Singular Spectrum Analysis (cMSSA) [Dirie, Abid, and Zou 2019] combines contrastive learning with Multivariate Singular Spectrum Analysis (MSSA). It is an extension of cPCA for time series data. The concept of contrastive learning also plays a significant role in statistical analysis [Dai et al. 2019], bioinformatics [Boileau, Hejazi, and Dudoit 2020], and data visualization [Fujiwara et al. 2020].

Related Work

Principal Component Analysis(PCA)

PCA aims to project high-dimensional data into a low-dimensional space, ensuring that the projected samples optimally describe the information of the original high-dimensional data [Zhou 2016] [Gao 2005]. Specifically, given a centralized sample matrix $X \in R^{D\times n}$, where D is the dimension of the original data and n represents the number of samples. Denote the sample covariance matrix as C . PCA aims to find a projection vector that maximizes the sample variance after projection, The corresponding objective can be described as follows:

$$\max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{C}\boldsymbol{w} \quad \text{s.t.} \quad \boldsymbol{w}^T\boldsymbol{w} = 1 \tag{1.1}$$

By the Lagrange technique, the optimization of (1.1) boils down to solving a eigenvalue problem:

$$\boldsymbol{C}\boldsymbol{w} = \lambda\boldsymbol{w} \tag{1.2}$$

Multiplying both sides of (1.2) from the left by w^T, we get:

$$\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} = \lambda\boldsymbol{w}^T\boldsymbol{w} = \lambda \tag{1.3}$$

From (1.3), to maximize the sample variance after projection, we should choose the eigenvector corresponding to the largest eigenvalue of the sample covariance matrix C as the projection vector w, which is the first principal component.

If the original high-dimensional data is projected into a multi-dimensional space, then PCA can be represented as:

$$\begin{aligned} &\max_{\boldsymbol{w}} \operatorname{tr}(W^T C W) \\ &\text{s.t.} \quad W^T W = I \end{aligned} \tag{1.4}$$

Where $W \in R^{D\times d}$ is the projection matrix, and $\boldsymbol{d}(\boldsymbol{d} \leq D)$ is the final dimension. Sort the eigenvalues of the sample covariance matrix from large to small: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\boldsymbol{d}} \geq \lambda_{\boldsymbol{d}+1} \geq \cdots \geq \lambda_D$, and take the eigenvectors corresponding to the top $\boldsymbol{d}$ eigenvalues to form the projection matrix W .

Canonical Correlation Analysis(CCA)

Given n pairs of pairwise samples $(\boldsymbol{x_1}, \boldsymbol{y_1}), \ldots, (\boldsymbol{x_n}, \boldsymbol{y_n})$ centralized by subtracting the total samples means from each sample. Let $\boldsymbol{x_1} = [\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}] \in R^{p\times n}$ and $\boldsymbol{x_2} = [\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}] \in R^{p\times n}$. CCA attempts to find a set of projections $\boldsymbol{x_1^T} W_1$ and $\boldsymbol{x_2^T} W_2$ for each view such that the correlation between $\boldsymbol{x_1^T} W_1$ and $\boldsymbol{x_2^T} W_2$ is maximized[Hotelling 1936]. The corresponding objective can be described as follows:

$$\max_{w_1, w_2} \frac{w_1^T C_{X_1X_2} w_2}{\sqrt{w_1^T C_{X_1X_1} w_1 \cdot w_2^T C_{X_2X_2} w_2}} \tag{1.5}$$

Where $C_{X_1X_1} = X_1 X_1^T$ and $C_{X_2X_2} = X_2 X_2^T$ are the covariance matrices of X_1 and X_2, respectively, and $C_{X_1X_2} = X_1 X_2^T$ is the cross-covariance matrix for X_1 and X_2.

Since (1.5) has scale invariance with respect to w_X and w_Y, CCA can also be expressed as:

$$\begin{aligned} &\max_{w_1, w_2} w_1^T C_{X_1X_2} w_2 \\ &\text{s.t.} \quad w_1^T C_{X_1X_1} w_1 = 1, \\ &\qquad\; w_2^T C_{X_2X_2} w_2 = 1. \end{aligned} \tag{1.6}$$

Using Lagrange multiplier technique, we can easily turn (1.6) into the following generalized eigenvalue problem:

$$\begin{bmatrix} 0 & C_{X_1X_2} \\ C_{X_2X_1} & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \lambda \begin{bmatrix} C_{X_1X_1} & 0 \\ 0 & C_{X_2X_2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{1.7}$$

Finally we select the eigenvector corresponding to the largest eigenvalue as the projection vectors wX and wY. Further, we can jointly get two projections Wx and Wy consisting the top $r(< \min(p, q))$ generalized eigenvectors of (1.7). In this way, a common dimensionality reduced subspace maximizing the between-view correlation is established.

Contrastive Principal Component Analysis(cPCA)

The idea behind cPCA is to utilize a background dataset as an auxiliary tool for DR. This ensures that the projected low-dimensional data reflects the differences between the target dataset and the background dataset, highlighting the distinctive information hidden in the target dataset [Abid et al. 2018]. Specifically, given a centralized target dataset $X \in \mathbb{R}^{D\times \boldsymbol{n}}$ and a background dataset $Y \in \mathbb{R}^{D\times \boldsymbol{m}}$, where D is the dimension of both the target dataset and the background dataset, $\boldsymbol{n}$ is the number of samples in the target dataset, and $\boldsymbol{m}$ is the number of samples in the background dataset. Let the sample covariance matrix of $C_X = XX^T$ and that of $C_Y = YY^T$. cPCA aims to find a projection vector $\boldsymbol{w} \in \mathbb{R}^D$ such that the target variance $\lambda_X(\boldsymbol{w}) = \boldsymbol{w}^T C_X \boldsymbol{w}$ is maximized while the background variance $\lambda_Y(\boldsymbol{w}) = \boldsymbol{w}^T C_Y \boldsymbol{w}$ is minimized, as expressed by:

$$\begin{aligned} &\max_{\boldsymbol{w}} w^T (C_X - \alpha C_Y) \boldsymbol{w} \\ &\text{s.t.} \quad \boldsymbol{w}^T \boldsymbol{w} = 1 \end{aligned} \tag{1.8}$$

Here, α is a parameter that balances the target variance against the background variance. The solution process for cPCA is similar to PCA. The eigenvector corresponding to the largest eigenvalue of $C_X - \alpha C_Y$ is taken as the projection vector $\boldsymbol{w}$. If the original high-dimensional data is projected into a multi-dimensional space, cPCA can be represented as:

$$\begin{aligned} &\max_{W} \operatorname{tr}\left(W^T (C_X - \alpha C_Y) W\right) \\ &\text{s.t.} \quad W^T W = I \end{aligned} \tag{1.9}$$

Where $W \in \mathbb{R}^{D\times \boldsymbol{s}}$ is the projection matrix, and $\boldsymbol{d}(d \leq D)$ is the dimension after reduction. The eigenvectors corresponding to the first d largest eigenvalues from the projection matrix W. Compared to PCA, cPCA replaces the singular C_x with $C_x - \alpha C_y$, eliminating the background part in the target dataset that causes interference and highlighting the crucial features. When $\alpha = 0$, cPCA degenerates into PCA.

Main Work

Motivation

CCA is an unsupervised DR algorithm that seeks a pair of projection vectors to maximize the correlation between two canonical variables in the low-dimensional space. It performs well on dual-view datasets with distinct features. However, background often plays an influential role in real-world datasets, making distinctive features less pronounced. Directly applying CCA to such datasets for low-dimensional feature extraction might lead to crucial features being overlooked, thereby affecting subsequent tasks such as classification and clustering. cPCA eliminates the negative effect of background data in the process of DR by virtue of CL. Inspired by cPCA, this paper introduces CL to the DR of dual-view datasets, which results in our algorithm: contrastive Canonical Correlation Analysis (cCCA). Figure 2 illustrates the relationships between the PCA, cPCA, CCA, and cCCA algorithms.

Modeling

cCCA integrates the concept of contrastive learning into CCA, introducing a dual-view background dataset. On the

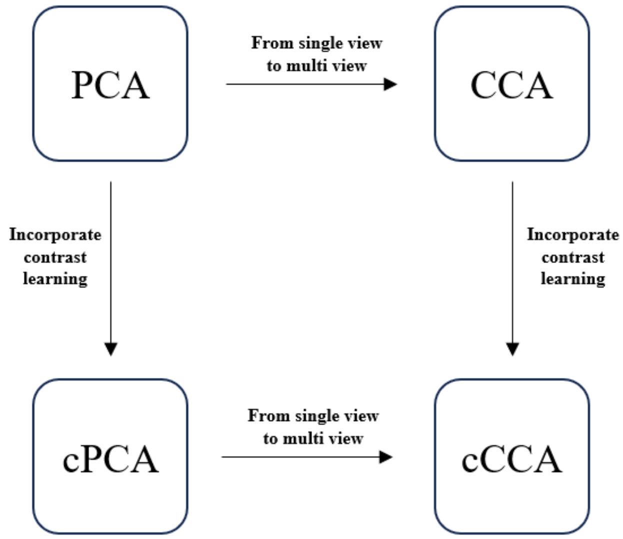

Figure 2: Schematic representation of the relationship between PCA, cPCA, CCA and cCCA.

one hand, it eliminates interference caused by the background portion of the target dataset. On the other hand, it maximizes the correlation between the two views after dimensionality reduction. Specifically, given a centralized paired target dataset $X_1 \in R^{D_1 \times \boldsymbol{n}}$ and $X_2 \in R^{D_2 \times \boldsymbol{n}}$, and a centralized paired background dataset $Y_1 \in R^{D_1 \times \boldsymbol{m}}$ and $Y_2 \in R^{D_2 \times \boldsymbol{m}}$, where X_1 and Y_1 belong to view 1, and X_2 and Y_2 belong to view 2. D_1 is the dimension of data in view 1, D_2 is the dimension of data in view 2, $\boldsymbol{n}$ is the number of samples in the target dataset, and $\boldsymbol{m}$ is the number of samples in the background dataset. Table 1 describes the relationship between the views and datasets in cCCA.

	Target Dataset	Background Dataset
View 1	X_1	Y_1
View 2	X_2	Y_2

Table 1: Relationship between views and datasets

Let $C_{X_1X_1} = X_1X_1^T$ be the covariance matrix of X_1, $C_{X_2X_2} = X_2X_2^T$ be the covariance matrix of X_2, $C_{Y_1Y_1} = Y_1Y_1^T$ be the covariance matrix of Y_1 ,$C_{Y_2Y_2} = Y_2Y_2^T$ be the covariance matrix of Y_2, and $C_{X_1X_2} = X_1X_2^T$ be the cross-covariance matrix between X_1 and X_2. cCCA aims to find a pair of projection vectors w_1 and w_2 such that the correlation between the two views of the target dataset in the low-dimensional space is maximized. Meanwhile, the variance within each view for the target dataset is maximized, and the variance for the background dataset is minimized, thus eliminating the influence of the background in the target dataset.

$$\begin{aligned}\max_{w_1,w_2} \Big(& w_1^T C_{X_1X_2} w_2 \\ & + \frac{a}{2}\left[w_1^T (C_{X_1X_1} - bC_{Y_1Y_1}) w_1\right. \\ & \left. + w_2^T (C_{X_2X_2} - cC_{Y_2Y_2}) w_2\right] \Big) \\ \text{s.t.} \quad & w_1^T C_{X_1X_1} w_1 + w_2^T C_{X_2X_2} w_2 = 1 \end{aligned} \tag{2.1}$$

Where $a (a \geq 0)$ is a weighting parameter. Moreover, $b (b \geq 0)$ and $c (c \geq 0)$ are parameters that balances the target variance and background variance within view 1 and view 2 respectively. Evidently, the first term of (2.1) aims to maximize the correlation of the low-dimensional features projected by w_1 and w_2. The second term performs cPCA separately on view 1 and view 2, aiming to eliminate the effects of backgrounds X_1 and X_2 that are similar to Y_1 and Y_2, highlighting the crucial features in the target dataset.

Further expanding the objective function in (2.1), we get:

$$\begin{aligned} & w_1^T C_{X_1X_2} w_2 \\ & + \frac{a}{2}[w_1^T (C_{X_1X_1} - bC_{Y_1Y_1}) w_1 \\ & + w_2^T (C_{X_2X_2} - cC_{Y_2Y_2}) w_2] \\ = \;& w_1^T C_{X_1X_2} w_2 \\ & - \frac{ab}{2} w_1^T C_{Y_1Y_1} w_1 - \frac{ac}{2} w_2^T C_{Y_2Y_2} w_2 \\ & + \frac{a}{2}\left(w_1^T C_{X_1X_1} w_1 + w_2^T C_{X_2X_2} w_2\right) \\ = \;& w_1^T C_{X_1X_2} w_2 - \frac{ab}{2} w_1^T C_{Y_1Y_1} w_1 \\ & - \frac{ac}{2} w_2^T C_{Y_2Y_2} w_2 + \frac{a}{2} \end{aligned}$$

Let $\alpha = ab$, $\beta = ac$, the optimization problem (2.1) is equivalent to:

$$\begin{aligned}\max_{w_1,w_2} \Big(& w_1^T C_{X_1X_2} w_2 \\ & - \frac{\alpha}{2} w_1^T C_{Y_1Y_1} w_1 - \frac{\beta}{2} w_2^T C_{Y_2Y_2} w_2 \Big) \\ \text{s.t.} \quad & w_1^T C_{X_1X_1} w_1 + w_2^T C_{X_2X_2} w_2 = 1 \end{aligned} \tag{2.2}$$

When $\alpha = 0$, $\beta = 0$, this method boils down into CCA. Optimization problem (2.2) takes into account both the correlation between views and the contrast within each view. It is suitable for dual-view datasets with complex backgrounds and can effectively eliminate interference caused by the background in the target dataset.

Optimization

In order to optimize problem (2.2), we employ the Lagrangian technique and define the following:

$$L(w_1, w_2, \lambda) = w_1^T C_{X_1X_2} w_2 - \frac{\alpha}{2} w_1^T C_{Y_1Y_1} w_1 - \frac{\beta}{2} w_2^T C_{Y_2Y_2} w_2 - \frac{\lambda}{2}\left(w_1^T C_{X_1X_1} w_1 + w_2^T C_{X_2X_2} w_2 - 1\right) \tag{2.3}$$

Differentiating (2.3) with respect to w_1 and w_2 and zeroing their derivatives, we have

$$\begin{cases} -\alpha C_{Y_1Y_1} w_1 + C_{X_1X_2} w_2 = \lambda C_{X_1X_1} w_1 \\ C_{X_1X_2}^T w_1 - \beta C_{Y_2Y_2} w_2 = \lambda C_{X_2X_2} w_2 \end{cases} \tag{2.4}$$

By utilizing algebraic operations, we can represents (2.4) as a generalized eigenvalue problem

$$\begin{bmatrix} -\alpha C_{Y_1Y_1} & C_{X_1X_2} \\ C_{X_1X_2}^T & -\beta C_{Y_2Y_2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \lambda \begin{bmatrix} C_{X_1X_1} & 0 \\ 0 & C_{X_2X_2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{2.5}$$

The eigenvectors corresponding to the largest eigenvalue of (2.6) are chosen as the projection vectors w_1 and w_2. Next, consider projecting the original high-dimensional data into a multi-dimensional space (let's denote it as d-dimensional).

Let $\binom{w_{1i}}{w_{2i}}$ be the eigenvector corresponding to the i-th eigenvalue λ_i in (2.6), then we have

$$\begin{bmatrix} -\alpha C_{Y_1Y_1} & C_{X_1X_2} \\ C_{X_1X_2}^T & -\beta C_{Y_2Y_2} \end{bmatrix} \begin{bmatrix} w_{1i} \\ w_{2i} \end{bmatrix} = \lambda_i \begin{bmatrix} C_{X_1X_1} & 0 \\ 0 & C_{X_2X_2} \end{bmatrix} \begin{bmatrix} w_{1i} \\ w_{2i} \end{bmatrix} \tag{2.6}$$

Therefore we can reformat (2.4) into

$$\begin{cases} -\alpha w_{1i}^T C_{Y_1Y_1} w_{1i} + w_{1i}^T C_{X_1X_2} w_{2i} = \lambda w_{1i}^T C_{X_1X_1} w_{1i} \\ w_{2i}^T C_{X_1X_2}^T w_{1i} - \beta w_{2i}^T C_{Y_2Y_2} w_{2i} = \lambda w_{2i}^T C_{X_2X_2} w_{2i} \end{cases} \tag{2.7}$$

Then, the objective function in (2.2) can be denoted as

$$\begin{aligned} &\sum_{i=1}^{d}\left(w_{1i}^T C_{X_1X_2} w_{2i} - \frac{\alpha}{2} w_{1i}^T C_{Y_1Y_1} w_{1i} - \frac{\beta}{2} w_{2i}^T C_{Y_2Y_2} w_{2i}\right) \\ &= \frac{1}{2}\sum_{i=1}^{d}\left(-\alpha w_{1i}^T C_{Y_1Y_1} w_{1i} + w_{1i}^T C_{X_1X_2} w_{2i}\right) \\ &\quad + \frac{1}{2}\sum_{i=1}^{d}\left(w_{2i}^T C_{X_1X_2}^T w_{1i} - \beta w_{2i}^T C_{Y_2Y_2} w_{2i}\right) \\ &= \frac{1}{2}\sum_{i=1}^{d}\left(\lambda_i w_{1i}^T C_{X_1X_1} w_{1i} + \lambda_i w_{2i}^T C_{X_2X_2} w_{2i}\right) \\ &= \frac{1}{2}\sum_{i=1}^{d}\lambda_i. \end{aligned}$$

From the function above, in order to obtain the maximum value for the objective function in the optimization problem (2.2), we should select a set of eigenvectors w_{1i} and $w_{2i} (i = 1, 2, \ldots, d)$ corresponding to the top d largest non-negative eigenvalues. Thus we obtain the projection matrices $W_1 = [w_{11}, w_{12}, \ldots, w_{1d}] \in \mathbb{R}^{D_1 \times d}$ and $W_2 = [w_{21}, w_{22}, \ldots, w_{1d}] \in \mathbb{R}^{D_2 \times d}$ for view 1 and view 2 respectively. The algorithm is summarized in Algorithm 1.

Algorithm 1: Contrastive Canonical Correlation Analysis (cCCA)

Require: View 1 target dataset $\hat{X}_1 \in \mathbb{R}^{D_1 \times n}$ and background dataset $\hat{Y}_1 \in \mathbb{R}^{D_1 \times m}$; View 2 target dataset $\hat{X}_2 \in \mathbb{R}^{D_2 \times n}$ and background dataset $\hat{Y}_2 \in \mathbb{R}^{D_2 \times m}$; Parameters α, β; Dimension d of the low-dimensional space after dimensionality reduction

1: **if** $W_1 \in \mathbb{R}^{D_1 \times d}$ and $W_2 \in \mathbb{R}^{D_2 \times d}$ **then**

2: Center the datasets $\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2$ to get X_1, Y_1, X_2, Y_2

3: Compute the covariance matrices and cross-covariance matrices: $C_{X_1X_1} = X_1X_1^T, C_{X_2X_2} = X_2X_2^T, \ldots$

4: Solve equation (2.5) to get the eigenvectors w_{1i} and w_{2i} corresponding to the top d largest eigenvalues (where $i = 1, 2, \ldots, d$)

5: Construct the projection matrices $W_1 = [w_{11}, w_{12}, \ldots, w_{1d}]$ and $W_2 = [w_{21}, w_{22}, \ldots, w_{2d}]$

6: **end if**=0

Experiments and Analysis

Conclusion

References

Abid, A.; Zhang, M. J.; Bagaria, V. K.; et al. 2018. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature communications*, 9(1): 1–7.

Andrew, G.; Arora, R.; Bilmes, J.; et al. 2013. Deep canonical correlation analysis. In *International conference on machine learning*, 1247–1255. PMLR.

Boileau, P.; Hejazi, N.; and Dudoit, S. 2020. Exploring high-dimensional biological data with sparse contrastive principal component analysis. *Bioinformatics*, 36(11): 3422–3430.

Dai, W.; Ng, K.; Severson, K.; et al. 2019. Generative oversampling with a contrastive variational autoencoder. 101–109.

Dirie, A. H.; Abid, A.; and Zou, J. 2019. Contrastive multivariate singular spectrum analysis. 1122–112.

Fujiwara, T.; Kwon, O. H.; and Ma, K. L. 2019. Supporting analysis of dimensionality reduction results with contrastive learning. *IEEE transactions on visualization and computer graphics*, 26(1): 45–55.

Fujiwara, T.; Zhao, J.; Chen, F.; et al. 2020. A visual analytics framework for contrastive network analysis. 48–59.

Gao, H. 2005. *Applied Multivariate Statistical Analysis*. Peking University Press.

Gao, L.; Qi, L.; Chen, E.; et al. 2012. Discriminative multiple canonical correlation analysis for multi-feature information fusion. In *2012 IEEE International Symposium on Multimedia*, 36–43. IEEE.

Gao, L.; Qi, L.; Chen, E.; et al. 2018. Discriminative Multiple Canonical Correlation Analysis for Information Fusion. *IEEE Transactions on Image Processing*, 27(4): 1951–1965.

Garte, S. 1998. The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1. *Carcinogenesis*, 19(8): 1329–1332.

Hardoon, D. R.; Szedmak, S.; and Shawe-Taylor, J. 2004. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12): 2639–2664.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6): 417.

Hotelling, H. 1936. Relations Between Two Sets of Variates. *Biometrika*, 28(3/4): 321–377.

Ibrahim, M. S.; and Sidiropoulos, N. D. 2020. Reliable detection of unknown cell-edge users via canonical correlation analysis. *IEEE Transactions on Wireless Communications*, 19(6): 4170–4182.

Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58(3): 433–451.

Kwak, N. 2008. Principal component analysis based on L1-norm maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(9): 1672–1680.

Liao, S.; Li, J.; Liu, Y.; et al. 2018. Robust formulation for pca: Avoiding mean calculation with l2, p-norm maximization. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Luo, J.; and Tjahjadi, T. 2020. Multi-set canonical correlation analysis for 3d abnormal gait behaviour recognition based on virtual sample generation. *IEEE Access*, 8: 32485–32501.

Melzer, T.; Reiter, M.; and Bischof, H. 2001. Kernel canonical correlation analysis. In *Proceedings of the International Conference on Artificial Neural Networks*, 353–360.

Nie, F.; Yuan, J.; and Huang, H. 2014. Optimal mean robust principal component analysis. In *International conference on machine learning*, 1062–1070. PMLR.

Nielsen, A. A. 2002. Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. *IEEE transactions on image processing*, 11(3): 293–305.

Pearson, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11): 559–572.

Qi, L.; Binu, D.; Rajakumar, B. R.; et al. 2022. 2-D canonical correlation analysis based image super-resolution scheme for facial emotion recognition. *Multimedia Tools and Applications*, 1–24.

Severson, K. A.; Ghosh, S.; and Ng, K. 2019. Unsupervised learning with contrastive latent variable models. 33(01): 4862–4869.

Shi, J.; Chen, C.; Liu, H.; et al. 2021. Automated Atrial Fibrillation Detection Based on Feature Fusion Using Discriminant Canonical Correlation Analysis. *Computational and mathematical methods in medicine*, 2021: 6691177.

Sun, T.; and Chen, S. 2007. Locality preserving CCA with applications to data visualization and pose estimation. *Image and Vision Computing*, 25(5): 531–543.

Sun, T.; Chen, S.; Yang, J.; et al. 2008. A novel method of combined feature extraction for recognition. In *2008 Eighth IEEE International Conference on Data Mining*, 1043–1048. IEEE.

Zhou, Z. 2016. *Machine Learning*. Tsinghua University Press.

Zou, J.; Hsu, D.; Parkes, D.; et al. 2013. Contrastive learning using spectral methods. 2238–2246.