Canonical Random Correlation Analysis

Jianchun Zhang
Department of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics
jczhang@nuaa.edu.cn

Daoqiang Zhang
Department of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics
dqzhang@nuaa.edu.cn

ABSTRACT

Canonical correlation analysis (CCA) is one of the most well-known methods to extract features from multi-view data and has attracted much attention in recent years. However, classical CCA is unsupervised and does not take class label information into account. In this paper, we introduce the within-class cross correlation into CCA and propose a new method called canonical Random Correlation Analysis (RCA). In RCA, besides considering the correlation between two views from the same sample, the cross correlations between two views respectively from different within-class samples are also used to achieve good performance. Two approaches for randomly generating cross correlation samples are developed.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: Design Methodology; H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Algorithms, Experimentation

Keywords

Canonical Correlation Analysis, Discriminant, Multi-View, Dimensionality Reduction

1. INTRODUCTION

In many applications data can be described by multiple sets of features. One of advantages of using multi-view data in classification is that different and complementary information contained in respective group of features can be used. Canonical correlation analysis (CCA) [2], is the most well-known two view-based method. CCA seeks to find two sets of directions, one for each set. Related features are extracted through maximizing the correlation between the two set of canonical variables.

CCA is inherently an unsupervised method and the label information can not be utilized in CCA, which limits its classification performance in practice. In this paper, we attempt to incorporate label information into CCA and propose a simple feature extraction method based on CCA, named canonical Random Correlation Analysis (RCA). RCA is based on the observation that an example is more related to the examples belonging to the same class than those belonging to different classes. For that goal, we introduce the idea of cross correlation [1] into CCA. In this study, we extend it to estimate the correlation relationships between two sets of examples falling into multiple classes. In RCA, within-class cross correlation is used to extract discriminative information. Two approaches are developed to produce within-class cross correlations, called RCA-I and RCA-II respectively. Both approaches generate cross correlation randomly except there are some constraints imposed on RCA-II. A smaller cross correlation set are needed in RCA-II than in RCA-I.

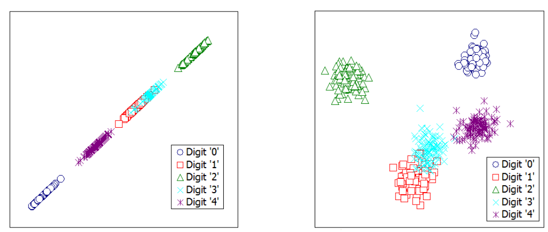

Figure 1: In RCA not only correlated features are extracted, but also discriminative information among various classes can be retained

2. CANONICAL RANDOM CORRELATION ANALYSIS

2.1 Sample Cross Correlation

Cross correlation is a standard method of measuring similarity of two time series. Borrowing idea from the standard cross correlation, not rigorously, we define sample cross correlation as

$$R = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j^T}{\sqrt{\sum_{i=1}^{n} x_i x_i^T}\sqrt{\sum_{j=1}^{n} y_j y_j^T}} \quad (1)$$

where $(x_i, y_i) \in \mathcal{S}$ is centered observations, and each sum term $x_i y_j^T$ is referred to as a cross correlation term, or correlation term for short. Now we can extend CCA to new canonical cross-correlation analysis as

$$\arg\max_{\omega_x,\omega_y} \omega_x^T\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j^T\Big)\omega_y \tag{2}$$

subject to

$$\omega_x^T\Big(\sum_{i=1}^{n} x_i x_i^T\Big)\omega_x = 1 \quad \text{and} \quad \omega_y^T\Big(\sum_{j=1}^{n} y_j y_j^T\Big)\omega_y = 1$$

2.2 RCA Algorithm

Suppose all samples fall into c classes $\{\omega_k\}_{k=1}^{c}$, and $\mathcal{X}_k$, $\mathcal{Y}_k$ are the jth subset of training data. Let $\widetilde{\mathcal{X}}_k$ and $\widetilde{\mathcal{Y}}_k$ denote the jth subset in final correlation term set. Two approaches are developed to determine $\widetilde{\mathcal{X}}_k$ and $\widetilde{\mathcal{Y}}_k$.

In RCA-I, we sample $\mathcal{X}_k$ and $\mathcal{Y}_k$ with replacement to form corresponding $\widetilde{\mathcal{X}}_k$ and $\widetilde{\mathcal{Y}}_k$. In RCA-II, only the second view $\mathcal{Y}$ is considered to be sampled and the first view is kept unchanged, i.e. $\widetilde{\mathcal{X}}_k = \mathcal{X}_k$. $\widetilde{\mathcal{X}}_k$ and $\widetilde{\mathcal{Y}}_k$ have the same size as $\mathcal{X}_k$ and $\mathcal{Y}_k$. The process can be repeated t times. As a result, t sets are generated, denoted as $\widetilde{\mathcal{X}}_k^{(l)}$, $\widetilde{\mathcal{Y}}_k^{(l)}$, $l = 1 \dots t$. Some correlation terms in the set may occur several times and the frequencies are defined as their weights. RCA (for both RCA-I and II) can be formulated as

$$\arg\max_{\omega_x,\omega_y} \omega_x^T\Big(\sum_{k=1}^{c}\sum_{l=1}^{t}\sum_{i=1}^{n_k} \widetilde{x}_i^{(l)}\widetilde{y}_i^{(l)^T}\Big)\omega_y \tag{3}$$

Let X,Y denote data matrices of $\mathcal{X}$ and $\mathcal{Y}$. A indication matrix $R_{corr} \in \mathbb{R}^{n\times n}$ is constructed to represent cross correlations, where n is the size of data set. The (i,j) entry of R_{corr} corresponds to the weight of the cross correlation term $x_i y_j^T$. Now, RCA (both I and II) can be rewritten as

$$\arg\max_{\omega_x,\omega_y} \omega_x^T X R Y^T \omega_y \tag{4}$$

where $R = R_{corr} + R_{corr}^T$ is set to guarantee symmetry of the correlation relationships, which implies that all symmetric terms are taken into account automatically to reinforce correlated relation further. The optimization problem can be solved by following eigenvalue decomposition,

$$\begin{bmatrix} & XRY^T \\ YRX^T & \end{bmatrix}\begin{bmatrix}\omega_x \\ \omega_y\end{bmatrix} = \lambda \begin{bmatrix} XX^T & \\ & YY^T \end{bmatrix}\begin{bmatrix}\omega_x \\ \omega_y\end{bmatrix} \tag{5}$$

The algorithm of RCA is summarized in Table 1.

Table 1: The Algorithm of RCA

Inputs:	Training data $\mathcal{X} = \bigcup_{k=1}^{c}\mathcal{X}_k$, $\mathcal{Y} = \bigcup_{k=1}^{c}\mathcal{Y}_k$,
	The number of correlation term set t,
	The dimension of canonical subspace d,
Begin:	Initialize $R_{corr} = (0)_{n\times n}$;
	For $l = 1$ **To** t **Do**
	Let $\widetilde{\mathcal{X}}^{(l)} = \emptyset$, $\widetilde{\mathcal{Y}}^{(l)} = \emptyset$;
	for $k = 1$ **To** c **Do**
	Construct lth set $\widetilde{\mathcal{X}}_k^{(l)}$, $\widetilde{\mathcal{Y}}_k^{(l)}$ from $\mathcal{X}_k$ and $\mathcal{Y}_k$ through the RCA-I or II;
	Set $\widetilde{\mathcal{X}}^{(l)} = \widetilde{\mathcal{X}}^{(l)} \cup \widetilde{\mathcal{X}}_k^{(l)}$, $\widetilde{\mathcal{Y}}^{(l)} = \widetilde{\mathcal{Y}}^{(l)} \cup \widetilde{\mathcal{Y}}_k^{(l)}$;
	Fill R_{corr} according to $\widetilde{\mathcal{X}}^{(l)}$ and $\widetilde{\mathcal{Y}}^{(l)}$;
	Set $R = R_{corr} + R_{corr}^T$;
	Solve Eq. (5);
Output:	$W_x = [\omega_{x_1} \cdots \omega_{x_d}]$ and $W_y = [\omega_{y_1} \cdots \omega_{y_d}]$.

Table 2: Recognition Rates on Multiple Features

Data	CCA	PLS	LPCCA	RCA-I	RCA-II
Fac Fou	0.8663	0.9353	0.9049	0.9544	**0.9560**
Fac Kar	0.9588	0.9365	0.9667	0.9750	**0.9768**
Fac Mor	0.7516	0.8661	0.7835	**0.8844**	0.8696
Fac Pix	0.9451	0.9400	0.9554	0.9732	**0.9748**
Fac Zer	0.8435	0.9476	0.8725	**0.9559**	0.9547
Fou Kar	0.8894	**0.9714**	0.9299	0.9349	0.9427
Fou Mor	0.7559	0.4343	0.7326	**0.8110**	0.8100
Fou Pix	0.8234	**0.9742**	0.8117	0.9288	0.9359
Fou Zer	0.8246	0.8062	0.8308	0.8419	**0.8435**
Kar Mor	0.7806	0.6256	0.8271	**0.8858**	0.8645
Kar Pix	0.9608	**0.9728**	0.9702	0.9478	0.9512
Kar Zer	0.8811	0.8163	**0.9454**	0.9368	0.9414
Mor Pix	0.7263	0.7053	0.7369	**0.8556**	0.8310
Mor Pix	0.7256	0.6919	0.7139	**0.7881**	0.7865
Pix Zer	0.8209	0.8261	0.8787	**0.9250**	0.9239

3. EXPERIMENTS

We evaluate the classification performances of the methods RCA-I and RCA-II on Multiple Features data set picked out from UCI repository. We compare our methods with several related algorithms, e.g. CCA, partial least square (PLS) [4], locality preserving CCA (LPCCA) [3]. The results on the data set are shown in Table 2. The first two columns in the table correspond all fifteen combinations of six views, i.e. Fac, Fou, Kar, Mor, Pix, Zer.

4. CONCLUSION

In this paper, we attempt to introduce label information into the CCA to obtain better class separation. The notion of cross correlation has been proven to be an effective method. We propose to choose a set of cross correlation terms within every class to form correlation term set on which CCA process is performed. Two simple and easy-to-do methods are developed to form the correlation term set.

5. ACKNOWLEDGMENTS

This work is partially supported by National Science Foundation of China under Grant No. 60875030 and the Innovation Foundation of NUAA.

6. REFERENCES

[1] P. Bourke. Cross correlation. http://local.wasp.uwa.edu.au/∼pbourke/miscellaneous/correlate/, 1996.

[2] D. R. Hardoon, S. Szedmák, and J. Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[3] T. Sun and S. Chen. Enhanced canonical correlation analysis with applications. *Dissertation in Nanjing University of Aeronautics and Astronautics*, 2006.

[4] J. A. Wegelin. A survey of partial least squares (PLS) methods, with emphasis on the two-block case. Technical Report 371, Department of Statistics, University of Washington, 2000.