软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(9):2018−2025 [doi: 10.13328/j.cnki.jos.004649]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

有序判别典型相关分析*

周航星, 陈松灿

(南京航空航天大学 计算机科学与技术学院,江苏 南京 210016)

通讯作者: 陈松灿, E-mail: s.chen@nuaa.edu.cn, http://parnec.nuaa.edu.cn

摘 要: 多视图学习方法通过视图间互补信息的融合,达到增强单一视图方法的鲁棒性并提升学习性能的目的.典型相关分析(canonical correlation analysis,简称 CCA)是一种重要的多视图信息融合技术.其研究的是针对同一组目标两组不同观测数据间的相关性,目标是得到一组相关性最大的投影向量.但当面对标号有序的分类任务时,CCA 因没有利用类信息和类间有序信息,造成了对分类性能的制约.为此,通过将有序类信息嵌入 CCA 进行扩展,发展出有序判别典型相关分析(ordinal discriminative canonical correlation analysis,简称 OR-DisCCA).实验结果表明,OR-DisCCA 的性能比相关方法更优.

关键词: 典型相关分析;有序回归;分类;信息融合;判别分析

中图法分类号: TP181

中文引用格式: 周航星,陈松灿.有序判别典型相关分析.软件学报,2014,25(9):2018−2025. http://www.jos.org.cn/1000-9825/4649.htm

英文引用格式: Zhou HX, Chen SC. Ordinal discriminative canonical correlation analysis. Ruan Jian Xue Bao/Journal of Software, 2014,25(9):2018−2025 (in Chinese). http://www.jos.org.cn/1000-9825/4649.htm

Ordinal Discriminative Canonical Correlation Analysis

ZHOU Hang-Xing, CHEN Song-Can

(College of Computer Science & Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China)

Corresponding author: CHEN Song-Can, E-mail: s.chen@nuaa.edu.cn, http://parnec.nuaa.edu.cn

Abstract: Multi-View learning is a method to improve the robustness and learning performance of single-view learning by fusing the complementary information. Canonical correlation analysis (CCA) which is used to analyze correlation between two datasets of the same objects is an important method for multi-view feature fusion. CCA aims to seek a pair of projections associated with the two sets of data such that they are maximally correlated. However, CCA results in constraint of the classification performance due to not utilizing the class information or ordinal information of different classes for some applications in which the data labels are ordinal. In order to compensate such a shortcoming, ordinal discriminative canonical correlation analysis (OR-DisCCA) is proposed in this paper by incorporating the class information and ordinal information for extending the traditional CCA. The experimental results show that OR-DisCCA outperforms existing related methods.

Key words: canonical correlation analysis; ordinal regression; classification; information fusion; discrimination analysis

多视图学习是机器学习和模式识别领域中处理同一对象多种表示(如同一图片同时由颜色和纹理构成)数据的一类重要学习方法,近年来已吸引了众多研究者的关注.典型相关分析(canonical correlation analysis,简称 CCA)[1]是一种处理多视图数据的经典方法,其用于分析同一组目标两组不同观测(视图)数据间的相关性,获得一组相关性最大的投影向量,从而达到综合视图信息并提升学习性能的目的.CCA、主成分分析(principal component analysis,简称 PCA)和线性判别分析(linear discriminant analysis,简称 LDA)是在多元统计中起着重要

* 基金项目: 国家自然科学基金(61170151); 江苏省自然科学基金(BK2011728); 江苏省“青蓝”工程创新团队项目
收稿时间: 2014-04-10; 定稿时间: 2014-05-14

© 中国科学院软件研究所 http://www.jos.org.cn

作用的特征分析方法,它们及其后续变形已广泛应用于图像分析[2]、模式识别[3]、多媒体分析[4]和信号处理[5]等众多领域.

Sun 等人[6]利用 CCA 提取出多视图特征后,通过并行和串行两种信息融合方式进行后续分类.由于该方法不仅对信息进行特征级融合,而且消除了信息冗余.因此,其后借助向量对的组合能够获得比使用单个向量更好的分类性能.然而该工作仅将 CCA 用于独立于分类的特征提取,尽管保证了所提取的两组特征间的相关性最大,但因未利用数据的类信息,导致获得的投影向量未必具有好的判别性能,因而也难以保证有好的分类效果.自然地,在相关分析中充分利用数据的类信息获取有利于分类效果的判别特征,成为研究的关注点之一[7].Arandjelovic[8]先提取两组数据中最相似的变量,提出扩展的 CCA(E-CCA);再通过归一化的类间散度矩阵及平均类内散度矩阵对输入向量进行变换,将 E-CCA 推广到判别框架中,提出 DE-CCA.Ji 等人[9]通过将视觉检索中 3 个等级信息(非常相关/相关/不相关)作为类别信息嵌入 CCA,用相关性约束矩阵替换原有的互协方差矩阵,提出 Rank-CCA,并且该方法最大程度地利用了有标号及无标号数据进行特征提取.Sun 等人[10]通过最小化类间相关距离和同时最大化类内相关距离提出了 DCCA,由于加入了类信息,获得了比 CCA 更有判别性的相关特征.然而当数据标号有序时,DCCA 虽然考虑了样本的类信息,但是与 CCA 一样,没有考虑类间的有序信息,难以保证好的有序分类性能.

有序回归是机器学习领域另外一个研究热点,其目标是自动判别那些具有固定离散等级样本的等级.例如,在一个 5 星电影打分评级系统中,给定一部电影特定的评分(1~5),分值越高,表明电影综合质量越高.这些具体分值可对应到相应的有序标号,使标号不仅表明样本的等级,而且不同标号的样本通常互相关,如标号为 3 的样本等级要比标号为 2 的高而比标号为 4 的低.一部电影具有很多属性(演员演技、服装设计、音乐编排等),由这些属性最终决定了该电影的评级(标号),也就意味着标号提供了有序信息.不同于普通的回归问题,有序回归的输出属于有限离散,更重要的是,对多类分类问题,有序回归的目标不仅希望能得到尽可能高的分类精度,同时还希望预测标号与真实标号间的差别尽可能小.因此,考虑有序信息能使分类时得出的绝对平均误差会比单纯的分类方法小.目前,有序回归已在信息检索、协同过滤或用户推荐系统、医疗和心理学等领域得到了应用[11].

对于有序回归,优化分类精度并不意味着对绝对误差的优化,反之亦然,这就要在两个目标优化时寻找平衡点.求解有序回归时,最通常的想法是将有序回归转换为传统的回归问题.Kramer 等人[12]通过将有序信息映射到对应数值上,得出距离敏感、预测有序标号的分类器.然而多数情况下,样本间真实的度量距离是未知的,导致并没有一种通用的方法可用来寻找合适的投影函数.更常见的做法是,将有序回归问题转换为多类分类问题或在传统分类问题中增添相应的约束.Herbrich 等人[13]对有序回归作了理论研究并将结构风险最小原理应用于有序回归.Hsu 等人[14]与 Smola 等人[15]都将有序回归转化为标准机器学习问题,分别提出了 SVC1V1,SVC1VA 以及支持向量回归(SVR).Crammer 和 Singer[16]通过寻找有序回归的方向及阈值,归纳出具有多阈值的感知机算法.在此基础上,Shashua 和 Levin[17]提出两个最大间隔原则用来处理方向及阈值,但该方法在某些情况下会导致获得的阈值排序错误.Chu 和 Keerthi[18]通过在约束中添加与阈值相关的约束对该方法进行改进,提出两个支持向量回归.基于从原始样本及代价矩阵中提取的样本,Li 等人[19]提出一个从有序回归到二分类问题的约简(reduce)框架,该框架对于现有几个有序回归算法提供了一个独特的视角.在此基础上,Lin 等人[20]通过增强二分类框架来解决有序回归问题,提出了 REDSVM.Cardoso 和 Costa[21]使用数据复制方法将有序回归问题转换为标准二分类问题,并将该方法应用于支持向量机及神经网络.Sun 等人[11]在有序回归框架中引入线性判别分析,提出基于有序回归的判别方法.通过保持原有线性判别分析对于最小类间距离的要求,同时添加约束保证各类样本投影后均值有序,提出基于有序回归的判别方法.进而运用核方法将其推广到非线性情况.

据我们了解,仅有 Kawashima[22]的工作致力于采用 CCA 进行多视图的有序回归学习,其算法的实现不是将有序信息直接嵌入 CCA,而是先用 CCA 进行信息融合,然后在融合空间中再借助现成的有序 LDA 设计有序回归算法.该方法与基于 CCA 的方法[9,10]类似,将信息融合与判别提取独立进行,如图 1 所示.受我们先前工作的启发[23],本文旨在通过对融合后的信息进行判别信息提取,将类信息以及有序信息都加入到 CCA 的目标之中,发展出真正意义上的有序判别典型相关分析(ordinal discriminative canonical correlation analysis,简称

© 中国科学院软件研究所 http://www.jos.org.cn

OR-DisCCA).在保持(广义)相关性的前提下,使最后获得的投影不仅让不同类的数据投影后尽可能地分开,还能使投影后的数据按照一定的顺序排列.这样既提升分类精度,又降低平均绝对误差(mean absolute error,简称MAE).实验结果表明,OR-DisCCA 的性能要比 CCA,DCCA 以及文献[22]的更优.另外,虽然 OR-DisCCA 只是对基本模型 CCA 的改造,但所用方法可方便地推广到其他改进形式的 CCA.

Fig.1 Feature extraction and information fusion

图 1 特征提取以及信息融合过程

本文第 1 节介绍典型相关分析以及有序判别回归.第 2 节具体描述本文提出的 OR-DisCCA.第 3 节是实验结果及与其他算法的对比.最后对本文的工作进行总结和展望.

1 相关背景

1.1 典型相关分析

本节将对 CCA 进行简单的回顾.

设 $\{x_i\}_{i=1}^n \in R^p$ 与 $\{y_i\}_{i=1}^n \in R^q$ 分别是同一目标两组中心化了的两视图观测数据,即 $\sum_{i=1}^n x_i = 0$ 和 $\sum_{i=1}^n y_i = 0$.

记 $X=[x_1\ldots x_n]\in R^{p\times n}$,$Y=[y_1\ldots y_n]\in R^{q\times n}$.CCA 旨在寻找一对投影向量 $w_x\in R^p$ 与 $w_y\in R^q$,使得 $w_x^T x$ 与 $w_y^T y$ 间的相关性最大.等价于使下式最大:

$$\max_{w_x\neq 0, w_y\neq 0} \frac{w_x^T XY^T w_y}{\sqrt{w_x^T XX^T w_x}\sqrt{w_y^T YY^T w_y}} \tag{1}$$

由于优化问题式(1)中目标函数不随 w_x 与 w_y 的大小变化而变化,则公式(1)可改写为

$$\min_{w_x, w_y} \| X^T w_x - Y^T w_y \|_2^2$$
$$\text{s.t. } w_x^T XX^T w_x = 1, w_y^T YY^T w_y = 1.$$

1.2 有序判别分析

设 $\{x_i, l_i\}_{i=1}^N \in R^l \times R$ 是一组训练样本,其中,$x_i\in R^l$ 表示输入,$l_i\in\{1,2,\ldots,K\}$表示样本对应的标号,则类内与类间散度矩阵分别为:

$$S_w = \frac{1}{N}\sum_{k=1}^K \sum_{x\in X_k} (x-m_k)(x-m_k)^T,$$
$$S_b = \frac{1}{N}\sum_{k=1}^K N_k (m_k - m)(m_k - m)^T,$$

其中,$m_k = \frac{1}{N_k}\sum_{x\in X_k} x$ 表示第 k 类的样本均值,$m = \frac{1}{N}\sum_{i=1}^N x_i$ 表示总样本均值.传统 LDA 通过最小化类内散度同时最大化类间散度得出如下优化目标:

$$\min_w \frac{w^T S_w w}{w^T S_b w}.$$

在 LDA 基础上,Sun 等人[17]考虑在投影过程中同时保留有序信息,使得等级高的样本在投影后的均值要比等级低的大,由此建立如下优化目标:

© 中国科学院软件研究所 http://www.jos.org.cn

$$\min_{w,\rho}\ w^T S_w w - C\rho,$$
$$\text{s.t.}\ \ w^T(m_{k+1}-m_k) \geqslant \rho, k=1,2,...,K-1,$$

其中,C 是惩罚系数.若ρ>0,则各类样本投影后的均值能够按照标号等级保持有序.

2 有序判别典型相关分析

本文将类信息与有序信息同时加入到传统CCA框架中,提出有序判别线性相关分析.为此,需达成如下3个目标:

(1) 保持样本投影后相关性最大.类似CCA,相关性最大可通过如下方式获得:

$$\min_{w_x,w_y}\ \| X^T w_x - Y^T w_y \|_2^2 \tag{2}$$

注意到,公式(2)并未包含传统 CCA 中防止产生退化解的约束,而本文会通过其他约束保证这一点.获得投影向量后,传统CCA有如下两种常用的串行和并行特征融合方式用于后续分类[6]:

$$w_x^T x + w_y^T y = \begin{pmatrix} w_x \\ w_y \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix},$$
$$\begin{pmatrix} w_x^T x \\ w_y^T y \end{pmatrix} = \begin{pmatrix} w_x & 0 \\ 0 & w_y \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix}.$$

根据经验[23],使用以上两种不同融合方法,最后得出的结果不同.而实验发现:使用并行融合,最后能够获得更好的分类性能.所以本文使用了并行融合方法,这种特征融合方法也被称为特征融合策略2(FFS-II).

(2) 与已提出的先提取特征再进行融合的方法不同,OR-DisCCA 期望融合后的信息具有判别性.

(3) 利用类信息保证融合特征的类内散度尽可能小、类间散度尽可能大.利用有序信息对融合特征投影后各类均值进行约束,使相邻两类样本投影后的均值按照标号等级保持有序,可通过如下方式达到:

$$\begin{cases} \min\limits_{w,\rho}\ (w_x^T\ \ w_y^T)\begin{pmatrix} S_w^x & 0 \\ 0 & S_w^y \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix} - C\rho \\ \text{s.t.}\ \ w^T(m_{k+1}-m_k) \geqslant \rho, k=1,2,...,K-1 \end{cases} \tag{3}$$

其中,$w=\begin{bmatrix} w_x \\ w_y \end{bmatrix}, m_k=\begin{bmatrix} m_{x,k} \\ m_{y,k} \end{bmatrix}$,$m_{x,k}$与 $m_{y,k}$ 为 X 与 Y 第 k 类的样本均值.结合以上各式,最终形成如下优化目标:

$$\begin{cases} \min\limits_{w,\rho}\ \| X^T w_x - Y^T w_y \|_2^2 + \lambda_1 w^T S_w w - C\rho \\ \text{s.t.}\ \ w^T(m_{k+1}-m_k) \geqslant \rho, k=1,2,...,K-1 \end{cases} \tag{4}$$

其中,$S_w=\begin{pmatrix} S_w^x & 0 \\ 0 & S_w^y \end{pmatrix}$,$\lambda_1$ 是参数.将平方项展开,公式(4)可改写为

$$\begin{cases} \min\limits_{w,\rho}\ w^T M w - C\rho \\ \text{s.t.}\ \ w^T(m_{k+1}-m_k) \geqslant \rho, k=1,2,...,K-1 \end{cases} \tag{5}$$

其中,$M=ZZ^T+\lambda S_w, Z=\begin{bmatrix} X \\ -Y \end{bmatrix}$.为了求解公式(5),定义拉格朗日函数:

$$L(w,\rho,\alpha) = w^T M w - C\rho - \sum_{k=1}^{K-1}\alpha_k\{w^T(m_{k+1}-m_k)-\rho\} \tag{6}$$

其中,拉格朗日乘子α_k≥0.对 L 求偏导,得出:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \frac{1}{2}M^{-1}\sum_{k=1}^{K-1}\alpha_k(m_{k+1}-m_k) \tag{7}$$

© 中国科学院软件研究所 http://www.jos.org.cn

$$\frac{\partial L}{\partial \rho}=0 \Rightarrow \sum_{k=1}^{K-1}\alpha_k=C \tag{8}$$

根据公式(7)以及公式(8),优化问题(5)可改写为:

$$\begin{cases}\min\limits_{\alpha} \ \sum\limits_{k=1}^{K-1}\alpha_k(m_{k+1}-m_k)^T M^{-1}\sum\limits_{k=1}^{K-1}\alpha_k(m_{k+1}-m_k) \\ \text{s.t.} \ \ \alpha_k \geqslant 0, k=1,2,...,K-1, \sum\limits_{k=1}^{K-1}\alpha_k=C\end{cases} \tag{9}$$

优化问题(9)是一个标准凸二次规划问题(quadratic program,简称 QP),许多现成软件包都可用来求解 QP 问题.

用上述方法可求出第 1 组 $w_1=[w_x^1,w_y^1]$,由于 OR-DisCCA 额外包含有序约束,不像 DCCA 可通过求解广义特征方程的方法一次获得多个投影向量,故需顺序地获取多个投影.大致步骤如下:若已求得前 d–1 个 w,即 $w_1,\ldots,w_{d-1}$,则第 d 个 w_d 要与之前求得的各 w 尽可能不相关,这可通过如下方法求解 w_d,其中,$d\leqslant\min(p,q)$:

$$\begin{cases}\min\limits_{w_d,\rho} \ w_d^T M w_d - C\rho + \lambda_2\sum\limits_{i=1}^{d-1}(w_i^T w_d)^2 \\ \text{s.t.} \ \ w_d^T(m_{k+1}-m_k)\geqslant \rho, k=1,2,...,K-1\end{cases} \tag{10}$$

其中,λ_2 是参数,为使投影方向间尽可能不相关,该参数可设为较大的值.现将平方项展开,公式(10)可改写为

$$\begin{cases}\min\limits_{w_d,\rho} \ w_d^T M_d w_d - C\rho \\ \text{s.t.} \ \ w_d^T(m_{k+1}-m_k)\geqslant \rho, k=1,2,...,K-1\end{cases} \tag{11}$$

其中,$M_d=M+\lambda_2\sum\limits_{i=1}^{d-1}w_i w_i^T$.公式(11)可运用之前所提方法求解.

3 实验结果与分析

由于现有许多研究都表明多视图的统计意义比单视图要好[24,25],所以本文进行对比的算法都是多视图算法.本节通过在多特征手写体数据集以及 8 个标准数据集上与 CCA,DCCA 及文献[22]中的方法进行比较.使用五折(5-fold)交叉验证来选择参数λ_1,λ_2,C 取为 10.最后,使用平均分类正确率(mean classification accuracy,简称 MCA)以及平均绝对误差(mean absolute error,简称 MAE)来度量性能.

设 N 个测试样本真实的标号为$\{y_1,y_2,\ldots,y_N\}$,预测的标号为 $\{\hat{y}_1,\hat{y}_2,...,\hat{y}_N\}$,那么 MCA 与 MAE 的定义分别为

$$MCA:\frac{N_1}{N},$$

$$MAE:\frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i-y_i|,$$

其中,N_1 为预测标号与真实标号间误差不超过 e 的样本个数.

3.1 多特征手写体数据集

多特征手写数据集(http://www.ics.uci.edu/~mlearn/MLSummary.html)是 UCI 数据集中的一个多视图数据集,包括 0~9 共 10 个手写数字的特征.数据集共有 2 000 个样本,每类 200 个,每个样本都由 6 个特征组成,分别是轮廓相关特征(mfeat_fac,216 维)、傅立叶系数(mfeat_fou,76 维)、K-L 展开系数(mfeat_kar,64 维)、像素平均(mfeat_pix,240 维)、Zernike 矩特征(mfeat_zer,47 维)、形态学特征(mfeat_mor,6 维).

任选两个特征作为两个视图,共有 15 种组合方式.实验中,每类随机选取 100 个样本作为训练集,剩余样本作为测试集,该过程独立地重复 20 次.最后,采用最近邻法进行分类,以 20 次平均得出的结果作为最终的性能衡量指标.其中,计算 MCA 时的参数 e 设置为 1.由于原数据特征的维度较高,先采用 PCA 对除形态学以外的特征进行降维,保持数据 95%的能量.另外,实验后发现:实验结果较为稳定,重复多次实验得出结果的方差较小,所以

© 中国科学院软件研究所 http://www.jos.org.cn

表中并未列出方差.实验结果见表 1.

Table 1 Test results on on multiple feature database

表 1 多特征手写数据集上的实验结果

特征组合方式			平均分类正确率(MCA)				平均绝对误差(MAE)			
			CCA	DCCA	OR-DisCCA	Ref.[22]	CCA	DCCA	OR-DisCCA	Ref.[22]
1	fac	fou	0.966 8	0.958 6	**0.974 7**	0.836 5	0.141 5	0.176 4	**0.099 2**	1.190 3
2	fac	kar	0.958 8	0.965 2	**0.970 8**	0.790 0	0.183 1	0.165 1	**0.129 7**	1.501 0
3	fac	mor	0.885 0	0.945 9	**0.958 7**	0.692 4	0.333 5	0.199 7	**0.145 4**	2.824 1
4	fac	pix	0.957 6	0.948 8	0.956 6	0.791 4	0.213 8	0.245 6	**0.175 2**	1.558 5
5	fac	zer	0.961 5	**0.966 2**	0.955 4	0.794 0	0.161 5	**0.151 9**	0.181 1	1.500 4
6	fou	kar	0.941 5	0.928 1	0.941 7	0.790 4	0.245 1	0.305 9	**0.213 9**	1.505 5
7	fou	mor	0.805 4	0.838 8	**0.841 4**	0.606 6	0.664 5	**0.536 2**	0.546 8	3.851 5
8	fou	pix	0.887 5	0.912 6	**0.933 3**	0.837 6	0.433 2	0.382 5	**0.250 1**	1.189 4
9	fou	zer	**0.850 3**	0.844 6	0.839 4	0.692 8	**0.480 1**	0.523 6	0.533 2	2.457 8
10	kar	mor	0.858 7	0.908 0	**0.940 2**	0.665 3	0.474 7	0.304 3	**0.204 7**	3.299 9
11	kar	pix	0.967 4	0.930 5	0.966 5	0.795 3	0.146 9	0.306 0	**0.136 4**	1.490 6
12	kar	zer	0.922 3	**0.957 8**	0.956 4	0.775 6	0.274 1	0.178 7	0.175 4	1.615 1
13	mor	pix	0.814 0	0.902 4	**0.930 5**	0.616 5	0.631 0	0.323 7	**0.253 6**	3.689 3
14	mor	zer	0.791 9	**0.811 8**	0.788 8	0.614 8	0.673 9	**0.590 7**	0.651 5	3,820 4
15	pix	zer	0.887 1	**0.952 2**	0.932 9	0.741 3	0.400 2	**0.202 7**	0.277 5	1.897 7

3.2 标准数据集

在 8 个标准数据集(www.liacc.up.pt/~ltorgo/Regression/DataSets.html)上进行对比实验.对每个数据集,为方便实验,依照原始标号的不同赋予其从 1~l 的新标号.实验中,前 4 个较小的数据集与后 4 个较大的数据集分别随机地以 50%和 30%的比例将每一类划分成训练集和测试集两部分,该过程独立地重复 20 次.最后,采用最近邻法进行分类,以 20 次平均得出的结果作为最终的性能衡量指标.其中,为了实验更能反映出各个方法性能,计算 MCA 值时,所需参数 e 的大小也随着各个数据集的标号不同而变化.参数设置及实验结果见表 2.

Table 2 Test results on eight data sets

表 2 标准数据集上的实验结果

数据集	参数		平均分类正确率(MCA)				平均绝对误差(MAE)			
	l	e	CCA	DCCA	OR-DisCCA	Ref.[22]	CCA	DCCA	OR-DisCCA	Ref.[22]
Pyrim	10	1	0.496± 0.0008	0.5530± 0.0006	**0.6061± 0.005**	0.4167± 0.005	1.8530± 0.052	1.7045± 0.049	**1.4500± 0.047**	2.1227± 0.199
CPU	14	1	0.894± 0.00002	0.890± 0.00002	**0.9015± 0.0009**	0.878± 0.006	0.6134± 0.004	0.6297± 0.002	**0.5767± 0.0011**	0.8173± 0.06
Boston	46	5	0.8458± 0.0007	0.8460± 0.0005	**0.8475± 0.001**	0.6569± 0.004	3.0163± 0.071	3.0952± 0.053	**2.9792± 0.019**	5.7577± 0.863
Abalone	28	3	**0.9302± 0.00001**	0.8340± 0.00002	0.8302± 0.0004	0.6647± 0.004	2.0199± 0.001	2.0009± 0.002	2.0500± 0.0086	5.2389± 1.218
Bank	30	3	0.6628± 0.0001	0.6409± 0.00004	**0.6910± 0.0002**	0.5378± 0.002	3.4089± 0.011	3.6799± 0.003	**3.0883± 0.024**	5.3128± 1.137
Computer	30	3	0.8874± 0.00004	0.9336± 0.00004	**0.9413± 0.006**	0.5795± 0.003	1.6302± 0.004	1.281± 0.0008	**1.2066± 0.0004**	3.6141± 0.220
California	50	5	0.5922± 0.00009	**0.6817± 0.00002**	0.6539± 0.0006	0.3625± 0.005	6.2524± 0.031	**5.0896± 0.002**	5.617± 0.035	10.49± 3.615
Census	117	5	**0.728± 0.00006**	0.728± 0.00009	0.7166± 0.0007	0.238± 0.008	5.5258± 0.004	**5.5411± 0.003**	5.7522± 0.0058	27.41± 66.9

3.3 实验结果分析

由表 1 可见,在 15 种特征组合中:

(1) 从 MCA 指标来看:OR-DisCCA 有 8 种组合性能优于其他方法,3 种与 CCA 持平;DCCA 有 4 种优于其他方法;CCA 仅有 1 种优于其他方法,3 种与 OR-DisCCA 持平;

(2) 从 MAE 指标来看:OR-DisCCA 有 9 种组合性能优于其他方法,1 种与 CCA 持平;DCCA 有 4 种优于其他方法;CCA 仅有 1 种优于其他方法,1 种与 OR-DisCCA 持平.

© 中国科学院软件研究所 http://www.jos.org.cn

由表 2 可见,在 8 个标准数据集上:

(1) 从 MCA 指标来看:OR-DisCCA 有 5 种性能优于其他方法,DCCA 仅有 1 种优于其他方法,CCA 有 2 种优于其他方法;

(2) 从 MAE 指标来看:OR-DisCCA 有 5 种性能优于其他方法,1 种与 DCCA 持平,DCCA 有 2 种优于其他方法.

上述实验结果表明:在 MFD 数据集上,大多数特征组合方式无论是在 MCA 还是 MAE,OR-DisCCA 的性能都优于其他方法,其中有 8 种显著更优.在 8 个标准数据集上,OR-DisCCA 整体上也体现出了更好的性能.此外,文献[22]中所提方法即使与最基本的 CCA 相比,其性能也无优势,这归咎于其将信息融合与判别信息提取的两个过程单独处理.这也再次表明了 OR-DisCCA 通过添加有序约束和对融合后信息的判别性特征提取,使数据投影后尽可能地保持了有序结构的同时,也能获得更好的有序回归特性.然而,我们的方法在某些数据集上也没有获得好的效果,比如在 Abalone 上,这可能是由于这些数据集本身的稀疏性或样本分布的非均匀性,导致对类均值估计的有偏.同时,由于某些样本点可能与其他同类样本点间距离间隔过大,使估计所得的类均值无法真实反映相关类的中心趋势,导致在投影空间中难以保持有序.

4 结论与展望

本文在 CCA 基础上,对融合后的信息进行判别特征提取及有序信息嵌入,提出了 OR-DisCCA.尽管 CCA 在处理多视图问题上简单有效,但并未利用类信息.LDA 利用了数据总体信息,但并未利用类间有序信息.本文对 CCA 进行了拓展,提出的 OR-DisCCA 不仅保持了 CCA 能达成视图数据间尽量相关的要求,同时有效地利用类信息和有序信息,使获得的组合特征更适合有序分类.在标准数据集上的实验结果,验证了本方法的有效性.下一步的工作一是运用核方法将其推广到非线性情况,二是推广至更现实的视图间样本的半配对场景,三是在最后分类过程中使用其他分类器替代最近邻分类器.

References:

[1] Hotelling H. Relations between two sets of variates. Biometrika, 1936,28(3/4):321−377. [doi: 10.1007/978-1-4612-4380-9_13]

[2] Hardoon DR, Szedmak S, Shawe TJ. Canonical correlation analysis: An overview with application to learning methods. Neural Computation, 2004,16(12):2639−2664. [doi: 10.1162/0899766042321814]

[3] Yan Y, Xiu P, Zhi X. A new fuzzy approach for handling class labels in canonical correlation analysis. Neurocomputing, 2008, 71(7-9):1735−1740. [doi: 10.1016/j.neucom.2007.11.008]

[4] Feng WW, Kim BU, Yu YZ. Real-Time data driven deformation using kernel canonical correlation analysis. ACM Trans. on Graphics, 2008,27(3):91. [doi: 10.1145/1360612.1360690]

[5] Li YO, Adali T, Wang W, Calhoun VD. Joint blind source separation by multiset canonical correlation analysis. IEEE Trans. on Signal Processing, 2009,57(10):3918−3929. [doi: 10.1109/TSP.2009.2021636]

[6] Sun QS, Zeng SG, Liu Y, Heng PA, Xia DS. A new method of feature fusion and its application in image recognition. Pattern Recognition, 2005,38(12):2437−2448. [doi: 10.1016/j.patcog.2004.12.013]

[7] Peng Y, Zhang DQ, Zhang JC. A new canonical correlation analysis algorithm with local discrimination. Neural Processing Letters, 2010,31(1):1−15. [doi: 10.1007/s11063-009-9123-3]

[8] Arandjelović Q. Discriminative extended canonical correlation analysis for pattern set matching. Machine Learning, 2014,94(3): 353−370. [doi: 10.1007/s10994-013-5380-5]

[9] Ji Z, Jing P, Su Y, Pang Y. Rank canonical correlation analysis and its application in visual search reranking. Signal Processing, 2013,93(8):2352−2360. [doi: 10.1016/j.sigpro.2012.05.006]

[10] Sun TK, Chen SC, Yang JY, Shi PF. A novel method of combined feature extraction for recognition. In: Giannotti F, ed. Proc. of the 8th IEEE Int'l Conf. on Data Mining. Los Alamitos: IEEE Computer Society, 2008. 1043−1048. [doi: 10.1109/ICDM.2008.28]

[11] Sun BY, Li JY, Wu DD, Zhang XM, Li WB. Kernel discriminant learning for ordinal regression. Knowledge and Data Engineering, 2010,22(6):906−910. [doi: 10.1109/TKDE.2009.170]

© 中国科学院软件研究所 http://www.jos.org.cn

[12] Kramer S, Widmer G, Pfahringer B, Groeve MD. Prediction of ordinal classes using regression trees. Fundamenta Informaticae, 2001,47(1):1−13.

[13] Herbrich R, Graepel T, Obermayer T. Large margin rank boundaries for ordinal regression. In: Solla SA, ed. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1999. 115−132.

[14] Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE Trans. on Neural Networks, 2002,13(2): 415−425. [doi: 10.1109/72.991427]

[15] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing, 2004,14(3):199−222. [doi: 10.1023/B: STCO.0000035301.49549.88]

[16] Crammer K, Singer Y. Pranking with ranking. In: Dietterich TG, ed. Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2001. 641−647.

[17] Shashua A, Levin A. Ranking with large margin principle: Two approaches. In: Becker S, ed. Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2002. 937−944.

[18] Chu W, Keerthi SS. New approaches to support vector ordinal regression. In: Raedt LD, ed. Proc. of the 22nd Int'l Conf. on Machine Learning. New York: Association for Computing Machinery Press, 2005. 145−152. [doi: 10.1145/1102351.1102370]

[19] Li L, Lin HT. Ordinal regression by extended binary classification. In: Becker S, ed. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2002. 865−872.

[20] Lin HT, Li L. Reduction of cost-sensitive ordinal ranking to weighted binary classification. Neural Computation, 2012,24(5): 1329−1367. [doi: 10.1162/NECO_a_00265]

[21] Cardoso JS, Da Costa JFP. Learning to classify ordinal data: The data replication method. Journal of Machine Learning Research, 2007,8(6):1393−1429.

[22] Kawashima T, Ogawa T, Haseyama M. A rating prediction method for e-commerce application using ordinal regression based on LDA with multi-modal features. In: Haraikawa T, ed. Proc. of the 2nd IEEE Global Conf. on Consumer Electronics. New York: IEEE Consumer Electronics Society, 2013. 260−261. [doi: 10.1109/GCCE.2013.6664818]

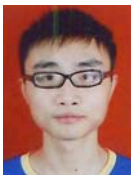
[23] Zhou XD, Chen XH, Chen SC. Combined-Feature-Discriminability enhanced canonical correlation analysis. Pattern Recognition and Artificial Intelligence, 2012,25(2):285−291 (in Chinese with English abstract).

[24] Lu JW, Tan YP. Fusing shape and texture information for facial age estimation. In: Zoubir A, ed. Proc. of the 2011 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. New York: IEEE Signal Processing Society, 2011. 1477−1480. [doi: 10.1109/ICASSP.2011.5946772]

[25] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23(6):681−685. [doi: 10.1109/34.927467]

附中文参考文献:

[23] 周旭东,陈晓红,陈松灿.增强组合特征判别性的典型相关分析.模式识别与人工智能,2012,25(2):285−291.



周航星(1991－),男,江西鹰潭人,硕士,主要研究领域为机器学习,模式识别.
E-mail: zhx1991@gmail.com

陈松灿(1962－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别,机器学习,数据挖掘,图像处理.
E-mail: s.chen@nuaa.edu.cn

© 中国科学院软件研究所 http://www.jos.org.cn