ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.11, November 2008, pp.2822−2832
DOI: 10.3724/SP.J.1001.2008.02822

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

半监督典型相关分析算法*

彭 岩, 张道强+

(南京航空航天大学 计算机科学与工程系,江苏 南京 210016)

Semi-Supervised Canonical Correlation Analysis Algorithm

PENG Yan, ZHANG Dao-Qiang+

(Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

+ Corresponding author: E-mail: dqzhang@nuaa.edu.cn

Peng Y, Zhang DQ. Semi-Supervised canonical correlation analysis algorithm. *Journal of Software*, 2008, 19(11):2822−2832. http://www.jos.org.cn/1000-9825/19/2822.htm

Abstract: In this paper, a semi-supervised canonical correlation analysis algorithm called Semi-CCA is developed, which uses supervision information in the form of pair-wise constraints in canonical correlation analysis (CCA). In this setting, besides abundant unlabeled data examples, the domain knowledge in the form of pair-wise constraints which specify whether a pair of data examples belongs to the same class (must-link constraints) or not (cannot-link constraints) is also available. Meanwhile, the relative importance of must-link constraints and cannot-link constraints is validated. Experimental results on the artificial dataset, multiple feature database and facial database including Yale and AR show that the proposed Semi-CCA can effectively enhance the classifier performance by using only a small amount of supervision information.

Key words: canonical correlation analysis; semi-supervised learning; pair-wise constraints; dimensionality reduction; classification

摘 要: 在典型相关分析算法(canonical correlation analysis,简称 CCA)的基础上,通过引入以成对约束形式给出的监督信息,提出了一种半监督的典型相关分析算法(Semi-CCA).在此算法中,除了考虑大量的无标号样本以外,还考虑成对约束信息,即已知两样本属于同一类(正约束)或不属于同一类(负约束),同时验证了两者的相对重要性.在人工数据集、多特征手写体数据集和人脸数据集(Yale 和 AR)上的实验结果表明,Semi-CCA 能够有效地利用少量的监督信息来提高分类性能.

关键词: 典型相关分析;半监督学习;成对约束;降维;分类
中图法分类号: TP181 **文献标识码**: A

典型相关分析(canonical correlation analysis,简称 CCA)[1]与主成分分析(principal component analysis,简称 PCA)[2]类似,在模式识别中的降维和数据可视化的应用中,都是经实验验证过的有效方法.PCA 是一种单模态分析方法,所谓的单模态识别是指利用从单一信息渠道获得的观察样本进行识别的技术.而 CCA 更侧重于多模态

* Supported by the National Natural Science Foundation of China under Grant Nos.60505004, 60875030 (国家自然科学基金); the Natural Science Foundation of Jiangsu Province of China under Grant No.BK2006521 (江苏省自然科学基金)

Received 2008-03-01; Accepted 2008-08-26

的识别(multimodal recognition),即利用互补原理,最大化不同模态数据之间的相关性,减少数据之间的不确定性,从而达到增强识别能力的目的.这里的多模态既可以是多种信息来源(如声音和图像),也可以是从同一来源的信息中抽取的不同特征.在过去的几年中,CCA 及其变型被成功地应用于人脸表情识别[3]、图像处理[4,5]、图像分析[6−8]、机器人位置估计[9]、数据回归分析[10]、基因组数据分析与基于内容的图像检索、文本挖掘[11,12]等.给定两个数据集 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$,CCA 的目标是求得一对基向量,使得两数据集之间的相关最大.值得注意的是,CCA 是一种无监督方法,它只关注成对样本之间的相关性,并将相关作为不同空间中样本之间的相似性度量.CCA 的缺点是未对样本的类信息加以利用,类信息的作用无法得到实现.而在很多实际应用中,我们在得到数据的同时,还能得到一些与数据相关的先验知识,如类标号信息、成对约束信息[13]等.如何在 CCA 中有效利用这些监督信息是本文关注的焦点.为此,我们引入半监督学习技术.

半监督学习是近年来机器学习领域的一个研究热点,已经出现了很多半监督学习算法[14],其中有代表性的方法有:半监督 EM 算法[15]、协同训练(co-training)算法[16]、直推式支持向量机[17]等.在很多实际应用中,随着数据采集技术和存储技术的发展,获取大量的无标号样本已变得非常容易,而获取有标号样本通常需要付出很大的代价.因而,相对于大量的无标号样本,有标号的样本通常会很少.传统的无监督学习只能利用无标号样本学习,监督学习则只利用少量的有标号样本学习,而半监督学习的优越性体现在能够同时利用大量的无标号样本和少量的有标号样本进行学习.目前的半监督学习一般分为三大类:半监督回归[16,18]、半监督聚类[19,20]和半监督分类[21,22].

近年来,半监督学习技术在特征提取或降维方法中也得到了应用[13,23,24].Cai 等人提出了一种半监督判别分析方法[23],Sugiyama 提出了半监督局部线性判别分析方法[24],这两种方法都是只用大量的无标号样本和少量的有标号样本进行降维,而没有采用其他约束信息.Zhang 等人提出了能够同时利用无标号样本和样本之间的成对约束信息的半监督降维方法[13].Zhou 等人借助 CCA 实现了只利用少量有标号样本的半监督学习[25].本文也在 CCA 的应用中加入了监督信息,提出一种半监督典型相关分析(Semi-CCA)算法.该方法中利用的监督信息为样本间的成对约束信息,即已知两个样本属于同一类(称为正约束,must-link)或者不属于同一类(称为负约束,cannot-link).值得指出的是,监督信息也包括类标号或其他先验信息,但在许多实际应用中,成对约束信息比类标号更容易获得,也更加实际.另外,样本之间的成对约束可以从类别标号中直接获得,反之则不可以.

本文第 1 节对 CCA 作简单介绍.第 2 节具体介绍半监督 CCA 算法(Semi-CCA).第 3 节通过实验对所提出算法进行性能测试,并对实验结果进行分析.最后对本文的工作进行总结,并展望进一步的工作.

1 典型相关分析(CCA)

给定一批成对的观察样本集 $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^n \in \boldsymbol{R}^p \times \boldsymbol{R}^q$, 其中 $\{\boldsymbol{x}_i\}_{i=1}^n$ 和 $\{\boldsymbol{y}_i\}_{i=1}^n$ 分别由不同信息渠道获得,记 $\boldsymbol{X}=[\boldsymbol{x}_1,...,\boldsymbol{x}_n]\in \boldsymbol{R}^{p\times n}$ 和 $\boldsymbol{Y}=[\boldsymbol{y}_1,...,\boldsymbol{y}_n]\in \boldsymbol{R}^{q\times n}$, 记 $(\boldsymbol{x},\boldsymbol{y})$ 为样本集 $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^n$ 中任意一对样本,并设样本已经中心化,即 $\overline{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i=0, \overline{\boldsymbol{y}}=\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i=0$, 则 CCA 的目标是分别为样本集 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 寻找两组基向量 $\boldsymbol{w}_x \in \boldsymbol{R}^p$ 和 $\boldsymbol{w}_y \in \boldsymbol{R}^q$, 使得随机变量 $\boldsymbol{x}=\boldsymbol{w}_x^T\boldsymbol{x}$ 和 $\boldsymbol{y}=\boldsymbol{w}_y^T\boldsymbol{y}$ 之间的相关达到最大,即求如下相关系数的最大值问题:

$$\rho=\frac{\boldsymbol{W}_x^T C_{xy}\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^T C_{xx}\boldsymbol{w}_x\cdot \boldsymbol{W}_y^T C_{yy}\boldsymbol{w}_y}} \tag{1}$$

其中,$C_{xx}=E[\boldsymbol{x}\boldsymbol{x}^T]=\boldsymbol{X}\boldsymbol{X}^T\in \boldsymbol{R}^{p\times p}$ 和 $C_{yy}=E[\boldsymbol{y}\boldsymbol{y}^T]=\boldsymbol{Y}\boldsymbol{Y}^T\in \boldsymbol{R}^{q\times q}$ 表示集合内协方差(within-set covariance)矩阵,$C_{xy}=E[\boldsymbol{x}\boldsymbol{y}^T]=\boldsymbol{X}\boldsymbol{Y}^T\in \boldsymbol{R}^{p\times q}$表示集合间协方差(between-set covariance)矩阵,且有 $C_{xy}=E[\boldsymbol{y}\boldsymbol{x}^T]=C_{xy}^T$. 有关 CCA 算法的求解步骤参见附录 1.

2 半监督典型相关分析(Semi-CCA)

Semi-CCA 加入了样本间成对约束信息,设有 n 对样本 $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^n \in \boldsymbol{R}^p \times \boldsymbol{R}^q$, Semi-CCA 的目标是寻找一组

投影向量 $\boldsymbol{w}_x \in \boldsymbol{R}^p$ 和 $\boldsymbol{w}_y \in \boldsymbol{R}^q$, 使得抽取的同类样本特征之间的相关最大化,同时使得不同类样本特征之间的相关最小化.具体可表述为如下优化问题:

$$\rho = \frac{\boldsymbol{w}_x^T \tilde{C}_{xy} \boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^T C_{xx} \boldsymbol{w}_x \cdot \boldsymbol{w}_y^T C_{yy} \boldsymbol{w}_y}} \tag{2}$$

其中,

$$\begin{aligned} \tilde{C}_{xy} &= \boldsymbol{X}\boldsymbol{Y}^T + \sum_{(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \boldsymbol{M}} (\boldsymbol{x}_i \boldsymbol{y}_j^T + \boldsymbol{x}_j \boldsymbol{y}_i^T) - \sum_{(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \boldsymbol{C}} (\boldsymbol{x}_i \boldsymbol{y}_j^T + \boldsymbol{x}_j \boldsymbol{y}_i^T) \\ &= \boldsymbol{X}\boldsymbol{E}\boldsymbol{Y}^T + \boldsymbol{X}\boldsymbol{M}\boldsymbol{Y}^T - \boldsymbol{X}\boldsymbol{C}\boldsymbol{Y}^T \\ &= \boldsymbol{X}(\boldsymbol{E} + \boldsymbol{M} - \boldsymbol{C})\boldsymbol{Y}^T \\ &= \boldsymbol{X}\boldsymbol{S}\boldsymbol{Y}^T \end{aligned} \tag{3}$$

$\boldsymbol{E}$ 是单位矩阵,$\boldsymbol{M}$ 是表示所有正约束的一个集合,$\boldsymbol{C}$ 是表示所有负约束的一个集合.若将 $\boldsymbol{M}$ 和 $\boldsymbol{C}$ 设为矩阵,则 $\boldsymbol{M} \in \boldsymbol{R}^{n \times n}$,$\boldsymbol{C} \in \boldsymbol{R}^{n \times n}$, 设初值时,$\boldsymbol{M}$ 和 $\boldsymbol{C}$ 都为零矩阵.由成对约束信息可知,当两个样本 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 属于同一类时,相应的 $\boldsymbol{M}_{ij}$ 和 $\boldsymbol{x}_{ji}$ 为 1;不属于同一类时,相应的 $\boldsymbol{C}_{ij}$ 和 $\boldsymbol{x}_{ji}$ 为–1.

Semi-CCA 算法的求解描述见附录 2,求得特征向量 $\boldsymbol{w}_x$,$\boldsymbol{w}_y$ 后,对任意一对样本$(\boldsymbol{x},\boldsymbol{y})$,即可用如下方式进行特征组合:

$$\boldsymbol{W}_x^T \boldsymbol{x} + \boldsymbol{W}_y^T \boldsymbol{y} \tag{4}$$

$$\begin{pmatrix} \boldsymbol{W}_x^T \boldsymbol{x} \\ \boldsymbol{W}_y^T \boldsymbol{y} \end{pmatrix} \tag{5}$$

其中,$\boldsymbol{W}_x = [\boldsymbol{w}_{x1},...,\boldsymbol{w}_{xd}] \in \boldsymbol{R}^{p \times d}$,$\boldsymbol{W}_y = [\boldsymbol{w}_{y1},...,\boldsymbol{w}_{yd}] \in \boldsymbol{R}^{q \times d}$,$d \le \min(p,q)$. 基于式(4)和式(5)的特征组合方式分别简称为“并行组合”与“串行组合”方式.利用组合的特征,可采用任何分类器(本文中采用最近邻法)进行分类.

Semi-CCA 算法的伪码描述如下:

输入:训练集数据 $\boldsymbol{X}$,$\boldsymbol{Y}$;训练集大小:n;类标号:label;约束个数:constrains;阈值:Threshold.

输出:利用 Semi-CCA 算法提取的特征向量 $\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$.

```
While (算法终止条件不满足) do
    设矩阵 E 为单位矩阵;
    初始化矩阵 M 和 C 为零矩阵,大小均为 n×n;
    For constrains=1,2,...,n do
        If (label(x_i)==label(x_j)) do
            M(i, j) = 1, M(j, i) = 1;
        Else
            C(i, j) = 1, C(j, i) = 1;
        End If
    End For
    计算相加矩阵, S = E + M + C;
    计算协方差矩阵, C_xy = XSY^T, C_xx = XX^T, C_yy = YY^T;
    计算矩阵, H = C_xx^(-1/2) · C_xy · C_yy^(-1/2);
    求 H 的 SVD(singular value decomposition)分解 U,D,V;
    选择满足条件 Threshold 的 [U_1,...,U_d] 和 [V_1,...,V_d];
    得到 w_x = C_xx^(-1/2)[U_1,...,U_d], w_y = C_yy^(-1/2)[V_1,...,V_d].
End While
```

3 实验结果与分析

在本文中,我们进行一系列实验来检验 Semi-CCA 的性能.首先通过一个简单的 Toy problem 直观地考察 Semi-CCA 抽取的特征对分类效果的影响,然后分别在多特征手写体数据集和 Yale(http://cvc.yale.edu/projects/yalefaces/yalefaces.html),AR(http://cobweb.ecn.purdue.edu/~aleix_face_DB.html)人脸数据集上检验用 Semi-CCA 降维后对识别能力的影响.

3.1 Toy problem实验

考虑包含 150 对二维样本的两类问题,记 $\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2]$ 和 $\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2]$ 表示两个数据集,其中 $\boldsymbol{X}_i,\boldsymbol{Y}_i$ 分别表示第 $i(i=1,2)$类样本.$\boldsymbol{X}_i$ 均满足高斯分布 $N(\boldsymbol{u}_i,\boldsymbol{\Sigma}_i)$, 其中,$\boldsymbol{u}_1=[10.18,0.66]^T,\boldsymbol{\Sigma}_1=\begin{bmatrix}15 & 3.75\\ 3.75 & 15\end{bmatrix},\boldsymbol{u}_2=[5,-5]^T,\boldsymbol{\Sigma}_2=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$. $\boldsymbol{Y}$ 数据集的样本 $\boldsymbol{y}_i$ 通过如下变换得到:$\boldsymbol{y}_i=\boldsymbol{W}^T\boldsymbol{x}_i+\boldsymbol{\varepsilon}_i,i=1,...,150$.其中,$\boldsymbol{W}=\begin{bmatrix}0.6 & -\sqrt{1/2}\\ 0.8 & \sqrt{1/2}\end{bmatrix}$,$\boldsymbol{\varepsilon}_i$ 为添加的高斯噪声,其分布为 $N(\boldsymbol{u}_\varepsilon,\boldsymbol{\Sigma}_\varepsilon)$,这里 $\boldsymbol{u}_\varepsilon=[1,1]^T,\boldsymbol{\Sigma}_\varepsilon=\begin{bmatrix}0.01 & 0\\ 0 & 0.01\end{bmatrix}$.可见,$\boldsymbol{x}_i$ 和 $\boldsymbol{y}_i(i=1,...,150)$之间满足一定程度的线性相关关系.

总样本个数为 300 个.图 1 显示了数据的分布情况.图 2(a)、图 2(b)分别给出了用 CCA 和 Semi-CCA 提取的第 1 对特征 $(\boldsymbol{w}_{x1}^T\boldsymbol{x},\boldsymbol{w}_{y1}^T\boldsymbol{y})$ 的分布.可以看出:(1) CCA 揭示了特征之间的线性关系,然而两类之间存在严重的重叠(如图 2(a)所示),这将可能造成错分;(2) 在 Semi-CCA 实验中,所有可能的约束个数为 $N(N-1)/2$,N 为样本数目,即 44 850 个,文中取正约束和负约束的比例相同,均为 0.5%,两类样本就可以被较好地分开(如图 2(b)所示).实验结果预示着用 Semi-CCA 提取的特征更有利于作模式识别.

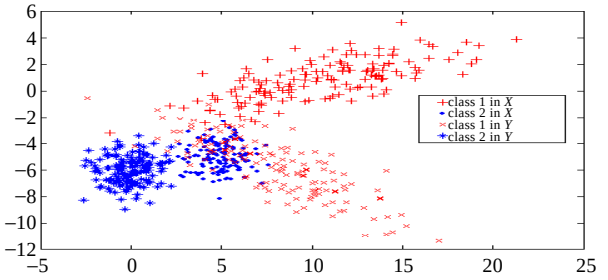


Fig.1 Data distribution of two classes. Signal +, • denote the features according to samples of the first and second class

图 1 两类数据分布.符号 +,• 分别表示第 1 类和第 2 类样本对应的特征

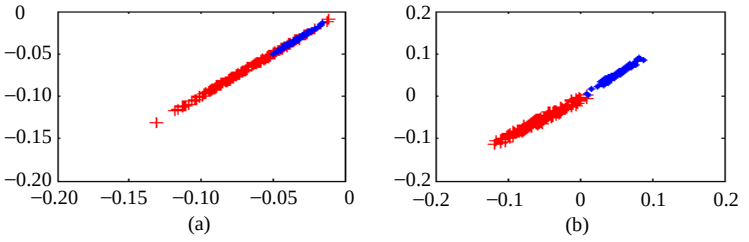


Fig.2 Distribution of the first pair of features extracted by CCA and Semi-CCA

图 2 算法 CCA 和 Semi-CCA 抽取的第 1 对特征的分布情况

3.2 手写体识别实验

本实验采用多特征手写体数据集来检验 Semi-CCA 提取特征的性能.该数据集属于 UCI

(http://www.ics.uci.edu/~mlearn/MLSummary.html)机器学习知识库的一个组成部分,包含 0~9 共 10 个数字的 6 个特征数据集,每类 200 个样本,共 2 000 个样本.从二值化手写体数字图像中抽取的 6 个特征包括傅里叶系数、轮廓相关特征、Karhunen-Loève 展开系数、像素平均、Zernike 矩和形态学特征,对应的数据集名称和样本维数分别为(mfeat_fou,76),(mfeat_fac,216),(mfeat_kar,64),(mfeat_pix,240),(mfeat_zer,47)和(mfeat_mor,6).

实验中,选任意两个数据集分别作为 $\boldsymbol{X}$ 集和 $\boldsymbol{Y}$ 集,共有 $C_6^2=15$ 种数据组合方式.对每种组合,在每类中随机抽取 100 对样本作训练,其余 100 对样本作测试,这样,训练集和测试集的样本对数均为 1 000 对,所有可能的约束个数为 $N(N-1)/2$,N 为样本数目,即 499 500 个.本文中,约束比例从 0.2%取到 2%.取任意一种约束对应的随机实验独立进行 10 次,记录其平均识别率.

3.2.1 正约束 $\boldsymbol{M}$ 和负约束 $\boldsymbol{C}$ 的比例相同

图 3 和图 4 列出了当 $\boldsymbol{C}$ 和 $\boldsymbol{M}$ 的约束比例相同,均为从 0.2%到 2%时,手写体数据集中两种不同组合(图 3 为 mfeat_fou 和 mfeat_kar 组合,图 4 为 mfeat_pix 和 mfeat_zer 组合)对应的 CCA 和 Semi-CCA 方法的识别结果.图中 pc 表示并行组合,sc 表示串行组合.从图中可以看出,CCA 与约束个数无关,而 Semi-CCA 的识别率随着约束所占比例的增加而增大;另外,在此 Semi-CCA 实验中,采用特征的串行组合方式能够获得较高的识别率.

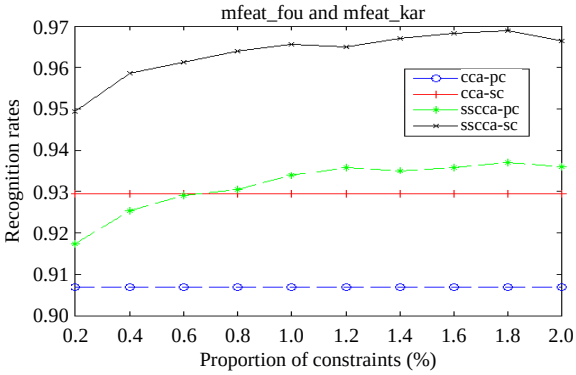


Fig.3

图 3

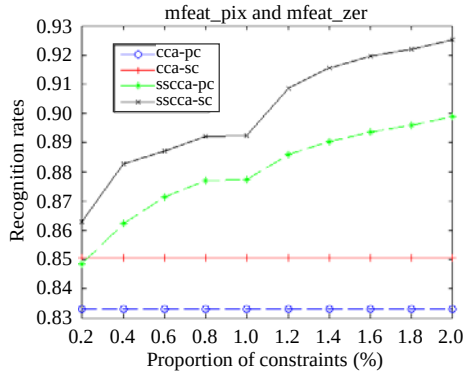


Fig.4

图 4

3.2.2 固定负约束 $\boldsymbol{C}$(或正约束 $\boldsymbol{M}$)的比例,变化正约束 $\boldsymbol{M}$(或负约束 $\boldsymbol{C}$)

为了检验正约束 $\boldsymbol{M}$ 和负约束 $\boldsymbol{C}$ 对识别效果影响的相对重要性,本次实验仍选择 mfeat_fou 和 mfeat_kar 组合(如图 5 所示)以及 mfeat_pix 和 mfeat_zer 组合(如图 6 所示).图 5(a)和图 6(a)是固定 $\boldsymbol{C}$ 的约束比例为 1%,变化 $\boldsymbol{M}$ 的约束比例从 0.2%到 2%时在两组合上的识别率,图 5(b)和图 6(b)是固定 $\boldsymbol{M}$ 的约束比例为 1%,变化 $\boldsymbol{C}$ 的约束比例从 0.2%到 2%时在两组合上的识别率.当固定 $\boldsymbol{C}$ 的约束比例为 1%,而变化 $\boldsymbol{M}$ 的约束比例从 0.2%到 2%时,Semi-CCA 的识别率随着 $\boldsymbol{M}$ 的约束比例的增加而增大(如图 5(a)和图 6(a)所示),但当固定 $\boldsymbol{M}$ 的约束比例,而变化 $\boldsymbol{C}$ 的约束比例从 0.2%到 0.5%时,Semi-CCA 的识别率虽然在 CCA 之上,但随着 $\boldsymbol{C}$ 的约束比例的增加,识别率不是保持水平,没有明显的上升趋势(如图 5(b)所示),就是有所下降(如图 6(b)所示).这说明样本的正约束对识别率的影响要远远大于负约束对识别率的影响,负约束比例的增加对识别率的影响不大,甚至可能会导致识别率的下降.实验结果表明,在采用 Semi-CCA 算法时,只需少量的先验信息(文中最多取到 2%),就可以取得较好的识别效果,这在实际应用中也是可取的.

表 1 是在 $\boldsymbol{C}$ 和 $\boldsymbol{M}$ 的约束比例相同,在文中所取的比例范围(0.2%~2%)内,随机实验进行 10 次,取平均识别率时,对所有结果数据的一个汇总.表中 PR1 和 PR2 两列分别表示利用特征并行和串行组合时的识别率.由实验结果可知,在实验中所取的约束范围内,基于 15 种数据集组合方式的实验中,除了 mfeat_kar 和 mfeat_pix 组合以外,其余 14 种组合方式下 Semi-CCA 的识别率都远远优于 CCA.

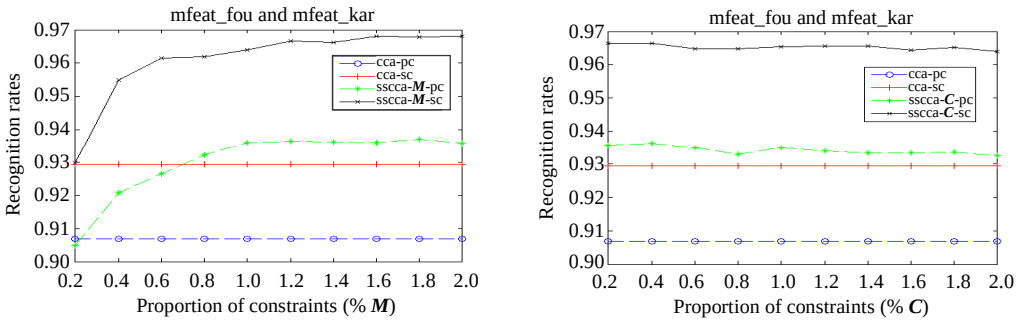


Fig.5

图 5

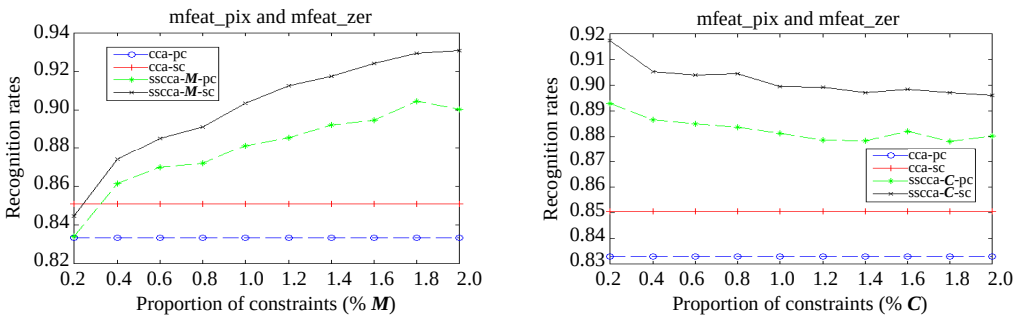


Fig.6

图 6

Table 1 Recognition rates in multiple feature database

表 1 多特征手写体数据集上的识别率

Combination of datasets			Recognition rates of CCA and Semi-CCA			
			CCA		Semi-CCA	
#	***X***	***Y***	PR1	PR2	PR1	PR2
1	mfeat_fac	mfeat_fou	0.879 2	0.889 9	**0.933 9**	**0.964 7**
2	mfeat_fac	mfeat_kar	0.964 8	0.964 9	**0.971 4**	**0.974 8**
3	mfeat_fac	mfeat_mor	0.760 5	0.769 1	**0.847 6**	**0.879 3**
4	mfeat_fac	mfeat_pix	0.952 4	0.951 4	**0.963 9**	**0.966 5**
5	mfeat_fac	mfeat_zer	0.859 7	0.869 8	**0.919 5**	**0.937 4**
6	mfeat_fou	mfeat_kar	0.906 9	0.929 4	**0.931 6**	**0.963 4**
7	mfeat_fou	mfeat_mor	0.760 2	0.769 3	**0.800 8**	**0.814 9**
8	mfeat_fou	mfeat_pix	0.837 7	0.851 9	**0.905 5**	**0.938 2**
9	mfeat_fou	mfeat_zer	0.833 2	0.842 4	**0.838 7**	**0.850 0**
10	mfeat_kar	mfeat_mor	0.787 1	0.817 4	**0.842 5**	**0.889 3**
11	mfeat_kar	mfeat_pix	**0.967 5**	**0.967 3**	0.943 2	0.941 9
12	mfeat_kar	mfeat_zer	0.914 4	0.926 3	**0.934 4**	**0.950 4**
13	mfeat_mor	mfeat_pix	0.729 5	0.760 1	**0.806 0**	**0.844 8**
14	mfeat_mor	mfeat_zer	0.741 0	0.758 0	**0.778 9**	**0.804 3**
15	mfeat_pix	mfeat_zer	0.833 1	0.850 6	**0.880 2**	**0.900 8**

3.3 人脸识别实验

为了进一步检验 Semi-CCA 的性能,我们在 Yale 和 AR 两个人脸数据集上进行了人脸识别实验,并与 CCA[26]作了对比.在实验中,将图像进行两次 Daubechies 正交小波变换,选取低频分量组成 CCA 和 Semi-CCA 实验的另一个数据集.完成特征抽取后,用最近邻法分类.

Yale 人脸数据集包含 15 个人的 165 幅灰度图像,每人 11 幅,均包括左/右/正面光照、戴/不戴眼镜、正常脸、愉快、悲伤、困乏、惊讶、眨眼等变化.本文中,通过手工校准的方式将 Yale 人脸数据集的原始图片剪裁成 100×100 像素大小,作为 $\boldsymbol{X}$ 数据集,$\boldsymbol{X}$ 对应的小波变换作为 $\boldsymbol{Y}$ 数据集,每个人选取 6 张图像进行训练,其余 5 张作测试,这样,每次实验共有 90 个训练样本和 75 个测试样本.所有可能的约束个数为 $N(N-1)/2$,N 为样本数目,即 4 465 个.

AR 人脸数据集由 126 个人(男 70 人,女 56 人)的 4 000 余幅彩色图像构成,每人 26 张图像,分为两组,每组 13 张,拍摄时间间隔两周,分别反映了人脸的表情、光照和遮挡(墨镜或围巾)的变化.本文中,我们随机选取 50 人(其中男 30 人,女 20 人),选择其中每人的 14 张无遮挡图像进行人脸识别实验,这样生成一个共有 700 张图像的人脸库.经过剪裁、尺度拉伸、人工校准等方式,将每张图像处理成 165×120 大小,作为 $\boldsymbol{X}$ 数据集,$\boldsymbol{X}$ 对应的小波变换作为 $\boldsymbol{Y}$ 数据集.对于每个人,随机选取 7 张进行训练,另 7 张进行测试.这样,每次实验共有 350 个训练样本和 350 个测试样本.所有可能的约束个数为 $N(N-1)/2$,N 为样本数目,即 61 075 个.本文中,两个数据集中 $\boldsymbol{C}$ 和 $\boldsymbol{M}$ 的约束比例均从 0.5%取到 5%.取任意一种约束对应的随机实验独立进行 10 次,记录平均识别率.

3.3.1 正约束 $\boldsymbol{M}$ 和负约束 $\boldsymbol{C}$ 的比例相同

图 7 和图 8 显示了当 $\boldsymbol{C}$ 和 $\boldsymbol{M}$ 的约束比例相同,均为从 0.5%到 5%时,CCA 和 Semi-CCA 分别在 Yale 和 AR 数据集上的识别率.从图中可以看出,CCA 与约束个数无关,且 CCA 算法的串行组合方式和并行组合方式获得的识别率相同,而 Semi-CCA 的识别率随着约束所占比例的增加,虽然识别率有局部下降的情况出现,但总体趋势是上升的.另外,在此 Semi-CCA 实验中,采用特征的并行组合方式能够获得较高的识别率.

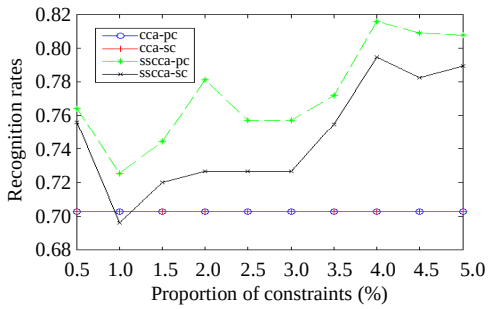


Fig.7

图 7

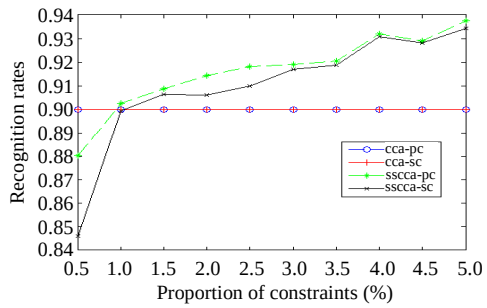


Fig.8

图 8

3.3.2 固定负约束 $\boldsymbol{C}$(或正约束 $\boldsymbol{M}$)的比例,变化正约束 $\boldsymbol{M}$(或负约束 $\boldsymbol{C}$)

图 9(a)和图 10(a)是固定 $\boldsymbol{C}$ 的约束比例为 2.5%,变化 $\boldsymbol{M}$ 的约束比例从 0.5%到 5%时分别在 Yale 和 AR 数据集上的识别率,图 9(b)和图 10(b)是固定 $\boldsymbol{M}$ 的约束比例为 2.5%,变化 $\boldsymbol{C}$ 的约束比例从 0.5%到 5%时分别在 Yale 和 AR 数据集上的识别率.从中可以看出,在图 9(a)和图 10(a)中,随着约束比例的增加,识别率是上升的,且 Yale 数据集上采用特征的并行组合方式能够获得较高的识别率,在 AR 数据集中比较特征的并行组合与串行组合方式对识别率造成的影响,发现二者并没有明显的区别;而在图 9(b)和图 10(b)中,一开始在 $\boldsymbol{C}$ 所占的比例很小时,有很高的识别率,在 Yale 上可达到 78%,在 AR 上几乎达到 93.5%,随着约束比例的增加,识别率虽然有局部增加的情况,但总体趋势是下降的.这说明,样本的正约束对识别率的影响要远大于负约束对识别率的影响;相反,随着负约束比例的增加,可能会导致识别率的下降,这不是我们所希望的.所以,在实际应用中应该控制 $\boldsymbol{C}$ 的约束比例.另外,实验中约束个数最多只取到 5%,可见,只加入少量的约束就可以使识别率有很大的提高,这在实际应用中也是可以接受的.

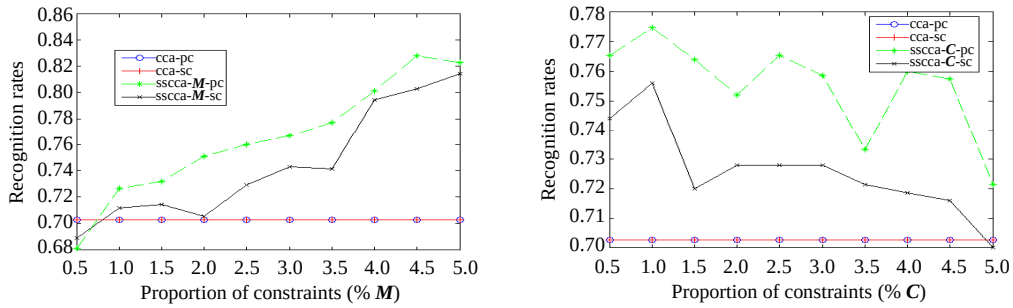


Fig.9

图 9

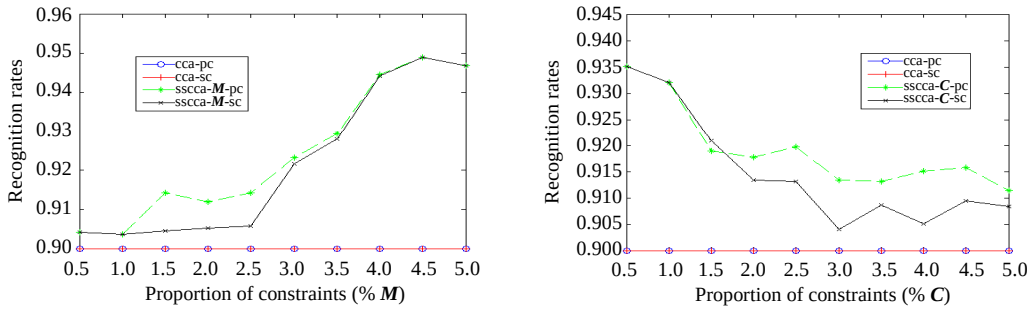


Fig.10

图 10

表 2 是将 ***C*** 和 ***M*** 的约束比例都固定为 5%,随机实验进行 10 次,取平均识别率时,对所有结果数据的一个汇总.表中 PR1 和 PR2 两列分别表示利用特征并行和串行组合时的识别率.由实验结果可知,在实验中所取的较小的约束范围内,Semi-CCA 的识别率要优于 CCA.

Table 2 Recognition rates in face database

表 2 人脸数据集上的识别率

Dataset	Recognition rates of CCA and Semi-CCA			
	CCA		Semi-CCA	
	PR1	PR2	PR1	PR2
Yale	0.702 7	0.702 7	**0.808 9**	**0.789 3**
AR	0.900 0	0.900 0	**0.937 4**	**0.934 3**

4 总结与展望

在典型相关分析(CCA)的基础上,本文结合半监督思想,提出了一种新的半监督多模态识别方法——半监督 CCA(Semi-CCA).Semi-CCA 不仅利用了大量的无标号样本,而且还考虑了样本间的成对约束信息,即已知两个样本属于同一类或不属于同一类为监督信息.用 Semi-CCA 进行特征抽取,能够有效利用少量的监督信息提高该方法在模式识别应用中的识别率.目前的多模态识别算法一般是基于两个模态,在未来的研究中,可以将半监督的思想方法与 MCCA(multiset CCA)[27]处理多数据集的能力结合起来,实现多模态半监督学习.

致谢 在此,我们向对本文工作予以支持和建议的老师和同学表示感谢,并向对本文工作不足之处提出评审意见的老师表示衷心的感谢.

References:

[1] Borga M, Knutsson H. Canonical correlation analysis in early vision Processing. In: Proc. of the 9th European Symp. on Artificial Neural Networks. 2001. 309−314.

[2] Gao HB, Hong WX, Cui JX, Xu YH. Optimization of principal component analysis in feature extraction. In: Proc. of the IEEE Int'l Conf. on Mechatronics and Automation. 2007. 3128−3132.

[3] Zheng WM, Zhou XY, Zou CR, Zhao L. Facial expression recognition using kernel canonical correlation analysis (KCCA). IEEE Trans. on Neural Networks, 2006,17(1):233−238.

[4] Loog M, B. van Ginneken B, Duin RPW. Dimensionality reduction by canonical contextual correlation projections. In: Proc. of the European Conf. on Computer Vision. 2004. 562−573.

[5] Hel-Or Y. The canonical correlations of color images and their use for demosaicing. Technical Report, HPL-2003-164(R1), HP Labs., 2004.

[6] Friman O, Carlsson J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magnetic Resonance in Medicine, 2001,45(2):323−330.

[7] Knutsson H, Borga M, Landelius T. Learning multidimensional signal processing. In: Proc. of the 14th Int'l Conf. on Pattern Recognition. 1998. 1416−1420.

[8] Nielsen AA. Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. IEEE Trans. on Image Processing, 2002,11(3):293−305.

[9] Vlassis N, Motomura Y, Kröse B. Supervised linear feature extraction for mobile robot localization. In: Proc. of the 2000 IEEE Int'l Conf. on Robotics and Automation. 2000. 2979−2984.

[10] Abraham B, Merola G. Dimensionality reductions approach to multivariate prediction. Computational Statistics and Data Analysis, 2005,48(1):5−16.

[11] Li YY, Shawe-Taylor J. Using KCCA for Japanese-English cross-language information retrieval and document classification. Journal of Intelligent Information Systems, 2006,27(2):117−133.

[12] Vinokourov A, Shawe-Taylor J, Cristianini N. Inferring a semantic representation of text via cross-language correlation analysis. In: Neural Information Processing Systems. 2002. 1473−1480.

[13] Zhang DQ, Zhou ZH, Chen SC. Semi-Supervised dimensionality reduction. In: Proc. of the 7th SIAM Int'l Conf. on Data Mining. 2007. 629−634.

[14] Zhu XJ. Semi-Supervised learning literature survey. Technical Report, 1530, Madison: Department of Computer Sciences, University of Wisconsin, 2005.

[15] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 2000,39(1):1−3.

[16] Zhou ZH, Li M. Semi-Supervised regression with co-training. In: Proc. of the 19th Int'l Joint Conf. on Artificial Intelligence. 2005. 908−913.

[17] El-Yaniv R, Pechyony D, Vapnik V. Large margin vs. large volume in transductive learning. Machine Learning, 2008,72(3): 173−188.

[18] Brefeld U, Gärtner T, Scheffer T, Wrobel S. Efficient co-regularized least squares regression. In: Proc. of the Int'l Conf. on Machine Learning. 2006. 137−144.

[19] Basu S, Bilenko M, Mooney RJ. A probabilistic framework for semi-supervised clustering. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2004. 59−68.

[20] Kulis B, Basu S, Dhillon I, Mooney RJ. Semi-Supervised graph clustering: A kernel approach. In: Proc. of the Int'l Conf. on Machine Learning. 2005. 532−537.

[21] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Peter B, Yishay M, eds. Proc. of the 11th Annual Conf. on Computational Learning Theory. Madison: ACM Press, 1998. 92−100.

[22] Zhang T, Ando RK. Analysis of spectral kernel design based semi-supervised learning. In: Neural Information Processing Systems. Cambridge: MIT Press, 2006. 1601−1608.

[23] Cai D, He XF, Han JW. Semi-Supervised discriminant analysis. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Rio de Janeiro, 2007.

[24] Sugiyama M, Ide T, Nakajima S, Sese J. Semi-Supervise local fisher discriminant analysis for dimensionality reduction. In: Proc. of the Pacific-Asia Conf. on Knowledge Discovery and Data Mining. 2008. 333−344.

[25] Zhou ZH, Zhan DC, Yang Q. Semi-Supervised learning with very few labeled training examples. In: Proc. of the 22nd AAAI Conf. on Artificial Intelligence. 2007. 675−680.

[26] 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.

[27] Vía J, Santamaría I, Pérez J. A learning algorithm for adaptive canonical correlation analysis of several data sets. Neural Networks, 2007,20(1):139−152.

[28] Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis. Pattern Recognition, 2003,36(9):1961−1971.

附录 1. CCA 求解算法

1. CCA 的一般求解算法

由于式(1)中的相关系数ρ与 $\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$ 的尺度无关,故 CCA 的求解问题可表述为如下优化形式:

$$\max_{\boldsymbol{w}_x,\boldsymbol{w}_y} \boldsymbol{w}_x^T C_{xy}\boldsymbol{w}_y,\ \text{s.t.}\ \boldsymbol{w}_x^T C_{xx}\boldsymbol{w}_x=1,\ \boldsymbol{w}_y^T C_{yy}\boldsymbol{w}_y=1 \tag{6}$$

为求解,定义 Lagrange 函数:

$$L(\lambda_1,\lambda_2,\boldsymbol{w}_x,\boldsymbol{w}_y)=\boldsymbol{w}_x^T C_{xy}\boldsymbol{w}_y-\frac{\lambda_1}{2}(\boldsymbol{w}_x^T C_{xx}\boldsymbol{w}_x-1)-\frac{\lambda_2}{2}(\boldsymbol{w}_y^T C_{yy}\boldsymbol{w}_y-1) \tag{7}$$

求解上述 Lagrange 函数,令

$$\begin{cases}\partial L/\partial \boldsymbol{w}_x = C_{xy}\boldsymbol{w}_y-\lambda_1 C_{xx}\boldsymbol{w}_x=0\\ \partial L/\partial \boldsymbol{w}_y = C_{yx}\boldsymbol{w}_x-\lambda_2 C_{yy}\boldsymbol{w}_y=0\end{cases} \tag{8}$$

分别用 $\boldsymbol{w}_x^T$ 和 $\boldsymbol{w}_y^T$ 左乘以式(8)中两等式的两边,可得:

$$0=\boldsymbol{w}_x^T C_{xy}\boldsymbol{w}_y-\lambda_1\boldsymbol{w}_x^T C_{xx}\boldsymbol{w}_x-\boldsymbol{w}_y^T C_{yx}\boldsymbol{w}_x+\lambda_2\boldsymbol{w}_y^T C_{yy}\boldsymbol{w}_y=\lambda_2\boldsymbol{w}_y^T C_{yy}\boldsymbol{w}_y-\lambda_1\boldsymbol{w}_x^T C_{xx}\boldsymbol{w}_x=\lambda_2-\lambda_1.$$

记 $\lambda_1=\lambda_2=\lambda$,设 C_{yy} 可逆且$\lambda\neq 0$,由式(8)中第 2 式可得 $\boldsymbol{w}_y=\frac{1}{\lambda}C_{yy}^{-1}C_{yx}\boldsymbol{w}_x$, 代入式(8)中第 1 式,整理得:

$$C_{xy}C_{yy}^{-1}C_{yx}\boldsymbol{w}_x=\lambda^2 C_{xx}\boldsymbol{w}_x \tag{9}$$

$$C_{yx}C_{xx}^{-1}C_{xy}\boldsymbol{w}_y=\lambda^2 C_{yy}\boldsymbol{w}_y \tag{10}$$

记 $\boldsymbol{w}=[\boldsymbol{w}_x^T,\boldsymbol{w}_y^T]^T$, 所求特征值 $\lambda=[\lambda_1,\lambda_2,...,\lambda_d,\lambda_{d+1},...,\lambda_r]$, 且 λ 非零,对应于非零特征值 λ_i 的特征向量为 $\boldsymbol{w}_{xi}$ 和 $\boldsymbol{w}_{yi}, i=1,...,d$, 这里 $d\le r\le\min(p,q)$, 则可利用任一对特征向量(即基向量)$\boldsymbol{w}_{xi}$ 和 $\boldsymbol{w}_{yi}$ 进行形如 $\boldsymbol{w}_{xi}\boldsymbol{x}$ 和 $\boldsymbol{w}_{yi}\boldsymbol{y}$ 的特征抽取,所抽取的特征 $\boldsymbol{w}_{xi}\boldsymbol{x}$ 和 $\boldsymbol{w}_{yi}\boldsymbol{y}$ 可称其为典型变量(canonical variant).

2. 利用 SVD 求解 CCA 算法

求解 CCA 方程还有一种更简单的方法,即利用奇异值分解(SVD)[28]的方法求解.令

$$\boldsymbol{H}=C_{xx}^{-1/2}C_{xy}C_{yy}^{-1/2},\ \boldsymbol{u}=C_{xx}^{1/2}\boldsymbol{w}_x,\ \boldsymbol{v}=C_{yy}^{1/2}\boldsymbol{w}_y,$$

则式(9)和式(10)可整理为

$$\begin{cases}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{u}=\lambda^2\boldsymbol{u}\\ \boldsymbol{H}^T\boldsymbol{H}\boldsymbol{v}=\lambda^2\boldsymbol{v}\end{cases} \tag{11}$$

可见,式(11)实际上对应于矩阵 H 的 SVD 分解,记 $\boldsymbol{H}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T=\sum_{i=1}^{d}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^T$ 为矩阵 $\boldsymbol{H}$ 的 SVD 分解,其中对角矩阵 $\boldsymbol{D}$ 的第 i 个对角元素恰好等于λ_i,$\boldsymbol{u}_i$ 和 $\boldsymbol{v}_i$ 分别是矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的第 i 列,对应于奇异值λ_i,有 $\boldsymbol{w}_{xi}=$

$C_{xx}^{-1/2}\boldsymbol{u}_i$ $\boldsymbol{w}_{yi}=C_{yy}^{-1/2}\boldsymbol{v}_i$,由此即可一次性得到 CCA 问题的第 $i,i=1,\ldots,d$ 对基向量,且利用 SVD 分解求解具有计算稳定的特点.

附录 2. Semi-CCA 求解描述

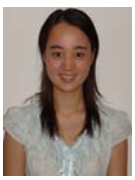
Semi-CCA 的求解可等价地表述为如下优化问题:

$$\max_{\boldsymbol{w}_x,\boldsymbol{w}_y}\boldsymbol{w}_x^T\tilde{C}_{xy}\boldsymbol{w}_y^T,\ \text{s.t.}\ \boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}_x=1,\boldsymbol{w}_y^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{w}_y^T=1 \tag{12}$$

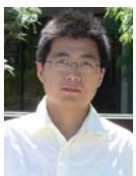
利用 Lagrange 乘子法,求解此优化问题,易得:

$$\begin{pmatrix}0 & \boldsymbol{X}\boldsymbol{S}\boldsymbol{Y}^T\\ \boldsymbol{Y}\boldsymbol{S}\boldsymbol{X}^T & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{w}_x\\ \boldsymbol{w}_y\end{pmatrix}=\lambda\begin{pmatrix}\boldsymbol{X}\boldsymbol{X}^T & 0\\ 0 & \boldsymbol{Y}\boldsymbol{Y}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{w}_x\\ \boldsymbol{w}_y\end{pmatrix} \tag{13}$$

其中,广义特征值 λ 即为目标值,式(13)最大的前 d 个广义特征值 $\lambda_i,i=1,\ldots,d$ 对应的特征向量 $\boldsymbol{w}_{xi},\boldsymbol{w}_{vi}$ 即为所求.



彭岩(1984－),女,山东枣庄人,硕士生,主要研究领域为模式识别,图像处理.



张道强(1978－),男,博士,教授,主要研究领域为模式识别,机器学习,数据挖掘.

敬 告 作 者

《软件学报》创刊以来,蒙国内外学术界厚爱,收到许多高质量的稿件,其中不少在发表后读者反映良好,认为本刊保持了较高的学术水平.但也有一些稿件因不符合本刊的要求而未能通过审稿.为了帮助广大作者尽快地把他们的优秀研究成果发表在我刊上,特此列举一些审稿过程中经常遇到的问题,请作者投稿时尽量予以避免,以利大作的发表.

1. 读书偶有所得,即匆忙成文,未曾注意该领域或该研究课题国内外近年来的发展情况,不引用和不比较最近文献中的同类结果,有的甚至完全不列参考文献.

2. 做了一个软件系统,详尽描述该系统的各个方面,如像工作报告,但采用的基本上是成熟技术,未与国内外同类系统比较,没有指出该系统在技术上哪几点比别人先进,为什么先进.一般来说,技术上没有创新的软件系统是没有发表价值的.

3. 提出一个新的算法,认为该算法优越,但既未从数学上证明比现有的其他算法好(例如降低复杂性),也没有用实验数据来进行对比,难以令人信服.

4. 提出一个大型软件系统的总体设想,但很粗糙,而且还没有(哪怕是部分的)实现,很难证明该设想是现实的、可行的、先进的.

5. 介绍一个现有的软件开发方法,或一个现有软件产品的结构(非作者本人开发,往往是引进的,或公司产品),甚至某一软件的使用方法.本刊不登载高级科普文章,不支持在论文中引进广告色彩.

6. 提出对软件开发或软件产业的某种观点,泛泛而论,技术含量少.本刊目前暂不开办软件论坛,只发表学术文章,但也欢迎材料丰富,反映现代软件理论或技术发展,并含有作者精辟见解的某一领域的综述文章.

7. 介绍作者做的把软件技术应用于某个领域的工作,但其中软件技术含量太少,甚至微不足道,大部分内容是其他专业领域的技术细节,这类文章宜改投其他专业刊物.

8. 其主要内容已经在其他正式学术刊物上或在正式出版物中发表过的文章,一稿多投的文章,经退稿后未作本质修改换名重投的文章.

本刊热情欢迎国内外科技界对《软件学报》踊跃投稿.为了和大家一起办好本刊,特提出以上各点敬告作者.并且欢迎广大作者和读者对本刊的各个方面,尤其是对论文的质量多多提出批评建议.