

融合异构特征的子空间迁移学习算法

张景祥^{1,2} 王士同¹ 邓赵红¹ 蒋亦樟¹ 李奕¹

摘要 特征迁移重在领域共有特征间学习, 然而其忽略领域特有特征的判别信息, 使算法的适应性受到一定的局限. 针对此问题, 提出了一种融合异构特征子空间的迁移学习(The subspace transfer learning algorithm integrating with heterogeneous features, STL-IHF) 算法. 该算法将数据的特征空间看成共享和特有特征子空间的组合, 同时基于经验风险最小框架将共享特征和特有特征共同嵌入到支持向量机(Support vector machine, SVM) 的训练过程中. 其在共享特征子空间上实现知识迁移的同时兼顾了领域特有的异构信息, 增强了算法的适应性. 模拟和真实数据集上的实验结果表明了所提方法的有效性.

关键词 特征空间, 异构特征, 迁移学习, 分类, 支持向量机

引用格式 张景祥, 王士同, 邓赵红, 蒋亦樟, 李奕. 融合异构特征子空间的迁移学习算法. 自动化学报, 2014, 40(2): 236–246

DOI 10.3724/SP.J.1004.2014.00236

A Subspace Transfer Learning Algorithm Integrating Heterogeneous Features

ZHANG Jing-Xiang^{1,2} WANG Shi-Tong¹ DENG Zhao-Hong¹ JIANG Yi-Zhang¹ LI Yi¹

Abstract The traditional transfer feature algorithms usually focus on learning by using the common features between the source domain and the target domain but ignore the discriminant information of the specific features of each domain, which makes the existing algorithms lack the adaptability to some extent. In order to circumvent this issue, in this paper a novel subspace transfer learning algorithm integrating heterogeneous features (STL-IHF) is proposed based on the empirical risk minimum framework. The proposed method is based on the support vector machine (SVM)-like framework with the feature space of each domain as a combination of the common features and the specified features. The proposed algorithm can not only realize the transfer learning from the common features but also effectively leverage the specified features of each domain, which makes it have much better adaptability in learning. Experimental results on simulation and real data set show the power of the proposed algorithm.

Key words Feature subspace, heterogeneous features, transfer learning, classification, support vector machine (SVM)

Citation Zhang Jing-Xiang, Wang Shi-Tong, Deng Zhao-Hong, Jiang Yi-Zhang, LI Yi. A subspace transfer learning algorithm integrating heterogeneous features. *Acta Automatica Sinica*, 2014, 40(2): 236–246

迁移学习作为一种新的学习框架, 其放松了训练数据与测试数据独立且服从相同的分布 (Identically and independently distributed, IID) 的束缚, 其强调在相似但存在差别的不同领域、任务和分布之间进行知识迁移, 从而实现在目标领域得到的分类器具有更好的泛化性能. 迁移学习由于其良好的灵活性和广泛的适应性等特点, 已经成为机器学习领域的研究热点^[1–3].

近几年, 学者们从不同角度提出很多迁移学习的方法. 从领域数据使用的角度, 迁移学习大致可

分为两类: 实例加权迁移学习方法和共同特征迁移学习方法. 基于实例迁移学习方法是通过对领域数据进行评估, 并从源领域中选出部分合适的实例迁移到目标任务的学习中^[4–6]. 代表性的工作有: 2004 年, Wu 等^[7] 提出的 LP-SVM 方法, 其通过加强辅助数据的训练, 提高学习机的分类性能; 2007 年, Dai 等^[8] 基于领域间样本的相似性, 提出 TrAdaBoost 算法; 2009 年, Quanz 等^[9] 引入约束源领域和目标领域间的样本分布均值差的策略, 提出了 LMPROJ 算法; 在最近兴起的多视角集成迁移学习方面, 2011 年, Xu 等^[10] 提出了多视角 Adaboost 迁移学习算法和 Chen 等^[11] 提出了多源学习的策略; 2012 年, Xu 等^[12] 进一步在迁移学习过程中同时融入多源和多视角的机制, 又提出了多源多视角 Adaboost 迁移学习算法, 有效地避免了负迁移的影响, 稳定地提高了迁移学习的学习效果; 2012 年, Jiang 等^[13] 针对目标领域信息缺失的问题, 提出了基于源领域知识模糊系统的迁移算法, 有效提高了迁移学习的性能. 基于

收稿日期 2012-12-31 录用日期 2013-04-02
Manuscript received December 31, 2012; accepted April 2, 2013
国家自然科学基金 (61170122, 61202311, 61272210), 江苏省自然科学基金 (BK2012552) 资助
Supported by National Natural Science Foundation of China (61170122, 61202311, 61272210), and Natural Science Foundation of Jiangsu Province (BK2012552)

本文责任编辑 刘成林
Recommended by Associate Editor LIU Cheng-Lin
1. 江南大学数字媒体学院 无锡 214122 2. 江南大学理学院 无锡 214122

1. School of Digital Media, Jiangnan University, Wuxi 214122
2. School of Science, Jiangnan University, Wuxi 214122

特征迁移学习方法是从特征结构上来建立领域数据间的联系, 通过特征的变换重构来发现潜在的共同特征空间, 利用特征层面上的交集来进行迁移学习^[14-15]. 代表性的工作有: 2008 年, Argyriou 等^[16] 提出了基于正则化的谱框架来学习多任务结构; 2008 年, Pan 等^[17-18] 通过在低维的映射空间中增强源领域数据和目标领域数据的均值中心一致性, 来减小两个领域的差异, 提出了基于流形结构的 MMDE 算法; 2011 年, Tu 等^[19] 提出了利用域适应算法来实现基于用户迁移的降维方法 (Transferable discriminative dimensionality reduction, TDDR), 在一定程度上提高了学习的性能; Gao 等^[20-21] 将提出的潜变量的核空间模型应用于行人检测问题, 表明基于特征的迁移策略能有效提高学习的性能.

上述方法虽然取得一定的效果, 但也面临着一些不足之处: 1) 基于实例的迁移学习算法性能往往受到样本选择“偏置性”影响较大或目标领域数据量较小的时候分布估计容易产生偏差, 因而该类算法的性能受到一定限制; 2) 基于特征的迁移学习, 拆分了源领域间和目标领域数据原本的特征空间, 仅选择部分共同特征进行迁移学习, 忽略了领域间特有特征的判别信息.

针对上述问题, 本文提出了一种融合异构特征的子空间迁移学习 (Subspace transfer learning algorithm integrating with heterogeneous features, STL-IHF) 算法. 该方法把不同领域的数据从纵向看作是共享特征和特有特征的组合, 以融入数据分布的嵌入期望距离 (Embedded mean distribution discrepancy, EMD) 度量共享特征空间的分布差异^[22]. 其保持训练样本共享特征在再生核 Hilbert 空间 (Reproduced kernel Hilbert space, RKHS) 中流形结构特征的一致性的同时, 基于结构风险最小化模型, 把领域数据的共享特征和特有特征同时嵌入到支持向量机的训练中, 从而能更加有效地实现了知识的迁移学习. 较之于现有相关方法, 本文方法的特色之处在于:

1) STL-IHF 方法将源领域和目标领域数据集看成共享特征和特有特征的组合, 继承已有特征迁移算法优势的同时融合数据本身的特有异构判别信息;

2) STL-IHF 方法继承了 SVM 基于结构风险最小化框架的优点, 通过数据分布的 EMD 距离, 减小不同分布之间的差异;

3) 新方法利用 Woodbury 矩阵转换引理, 解决了矩阵求逆的问题, 实现在再生核 Hilbert 空间上完全核化.

1 融合异构特征的子空间迁移学习机(STL-IHF)

1.1 相关概念及问题描述

为了便于讨论相关的迁移学习算法, 首先对一些相关概念和符号进行统一说明.

定义 1 (领域 (Domain)). 领域常用 $D = \{(\mathbf{x}_n, y_n) \in X \times Y: 1 \leq n \leq N\}$ 表示. 其中, X 表示领域样本的特征空间, Y 表示样本的标签空间, (\mathbf{x}_i, y_i) 表示领域 D 的一个样本.

定义 $P(D) = P(X, Y)$ 表示领域 D 中数据的分布. 有两个领域 D_s 和 D_t , 如果 $P(D_s) \neq P(D_t)$, 则 D_s 和 D_t 是两个不同的领域.

由定义 1 知, 迁移学习中的源领域 (Source domain) 可表示为 $SD = \{(\mathbf{x}_{s_1}, y_{s_1}), \dots, (\mathbf{x}_{s_l}, y_{s_l})\} \in (X_s, Y_s)_l$, 目标源领域 (Target domain) 可表示为 $TD = \{(\mathbf{x}_{t_1}, y_{t_1}), \dots, (\mathbf{x}_{t_n}, y_{t_n})\} \in (X_t, Y_t)_n$.

在迁移学习中, 源领域数据和目标领域数据通常来源于同一宏观领域, 因此两类数据存在一定的相似性. 但从微观角度分析, 这两类数据的分布仍存在一定的差异性, 特别表现在特征空间的构造上, 这两类数据集既兼具了宏观的相似特征空间即共享特征空间, 又存在微观上的表现差异性的特有特征. 具体的表现满足下述特性: 领域间在实现部分特征数据共享的同时, 每个数据集仍保有自己的特有特性. 因而, 将领域从纵向可分为共享特征和特有特征, 即源领域数据集 X_s 纵向表示为 $X_s = [\hat{X}, \tilde{X}_s]$ 和目标领域数据集纵向表示为 $X_t = [\hat{X}, \tilde{X}_t]$, 其中, \tilde{X}_s 和 \tilde{X}_t 分别表示为源领域和目标领域特有特征, \hat{X} 表示领域间共享特征.

现实应用中这样的问题是存在的, 例如: 从不同角度观察同一个物体, 虽然采样所得到的样本属性集合存在差异性, 但因采样对象是一致的, 因此, 在各角度下的样本属性集合存在交集. 本文定义此类交集为共享特征空间.

为度量领域间共享特征之间的分布距离, 并能反映数据集的分布结构特征, 为此, 引入定义一种再生核 Hilbert 空间距离度量.

定义 2 (嵌入期望分布距离度量 (Embedded mean distribution discrepancy, EMD))^[9].

设 X_1 和 X_2 是两个数据集, 数据容量分别是 n 和 m , 若在 RKHS 中存在投影函数 $f(\mathbf{x}) = \langle \mathbf{w}, \psi(\mathbf{x}) \rangle$, \mathbf{w} 为投影法向量, 则两个数据集在 RKHS 嵌入期望分布距离度量定义为 $EMD^2 = \|\frac{1}{n} \sum_{i=1}^n \mathbf{w}^T \psi(\mathbf{x}_i) - \frac{1}{m} \sum_{j=1}^m \mathbf{w}^T \psi(\mathbf{z}_j)\|^2 = \mathbf{w}^T (\frac{1}{n} \sum_{i=1}^n \psi(\mathbf{x}_i) - \frac{1}{m} \sum_{i=1}^m \psi(\mathbf{z}_j)) (\frac{1}{n} \sum_{i=1}^n \psi(\mathbf{x}_i) - \frac{1}{m} \sum_{i=1}^m \psi(\mathbf{z}_j))^T \mathbf{w}$, 其中, $\mathbf{x}_i \in X_1, \mathbf{z}_j \in X_2$.

本文研究的对应场景简述如下:

1) 仅考虑一个源领域和一个目标领域.

2) 领域间数据分布虽不相同但相似, 源领域中有大量数据且带有标签信息; 目标领域有少量带有标签的数据. 学习的任务是为目标领域训练一个具有较好分类性能的分类器.

1.2 STL-IHF 原理

基于特征迁移学习方法主要考虑源领域和目标领域的相似性, 通过特征变换寻找领域间的交集, 它利用特征结构信息对迁移过程进行约束, 并通过引入新的优化方式将它们有机结合起来, 平衡控制从源数据到目标数据的知识迁移. 基于这种认识, 特征迁移学习算法的决策优化函数可以表示如下:

$$f = \min \eta \sum_{i=1}^n V(\mathbf{x}_i, y_i, f) + \frac{1}{2} \|f\|_K^2 + \lambda d_{f,K}(P_s, P_t) \quad (1)$$

其中, $V(\mathbf{x}_i, y_i, f)$ 为结构风险函数, 正则化参数 η 是对错分样本的惩罚程度. P_s 表示源领域的分布, P_t 表示目标领域的分布, $d_{f,K}(P_s, P_t)$ 是源领域和目标领域间共有特征间距离的一种度量. 平衡参数 λ 的作用是对风险结构和领域间分布距离的协调.

已有的特征迁移方法大都通过构造距离函数 $d_{f,K}(P_s, P_t)$, 使得源领域和目标领域共有特征间的某些分布相似或距离相近, 进而顺利实现知识的迁移. 但文献 [22] 指出最好性能的分类器应有效地反映出两个不同数据集的分布结构特征, 包括同构和异构信息. 因此, 仅依赖共享特征而忽略数据特有的异构判别信息, 在一定程度上制约了学习算法泛化能力的进一步提高.

针对此问题, 将源领域数据表示为 $X_s = [\hat{X}, \tilde{X}_s]$ 和目标领域数据 $X_t = [\hat{X}, \tilde{X}_t]$, 其中 \hat{X} 为共享特征数据. 本文提出了一种融合异构特征的子空间迁移学习策略, 对于领域间共享特征通过减小它们在再生核 Hilbert 空间中结构的距离, 实现知识迁移学习的目标; 同时还要充分挖掘源领域和目标领域特有特征的异构信息, 以达到对迁移学习模型进行优化的目的. 我们把这种新策略称之为融合异构特征的子空间迁移学习 (The subspace transfer learning algorithm integrating with heterogeneous features, STL-IHF). STL-IHF 算法的原理图如图 1 所示.

由式 (1) 和图 1, 对 STL-IHF 方法的优化目标函数可表示为如下的一般化形式:

$$J = \min_{\theta} \eta \sum_{i=1}^{N_s} V(\tilde{\mathbf{x}}_i, y_i, f_s) + \frac{1}{2} \|f_s\|_K^2 +$$

$$\eta \sum_{i=1}^{N_t} V(\tilde{\mathbf{x}}_i, y_i, f_t) + \frac{1}{2} \|f_t\|_K^2 + \frac{1}{2} \|\hat{f}\|_K^2 + \lambda d_{f,K}(P_s, P_t) \quad (2)$$

其中, θ 为待优化参数, 对式 (2) 进一步分析如下:

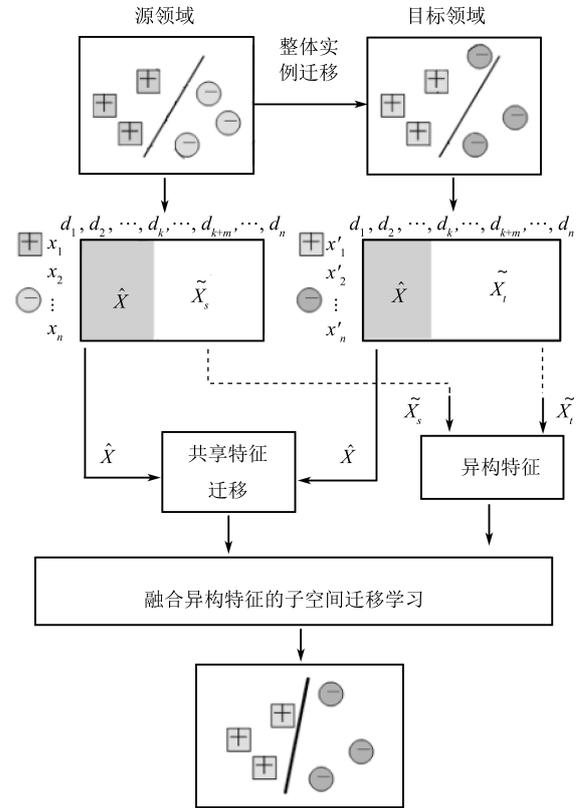


图 1 STL-IHF 方法的原理图

Fig. 1 Schematic of STL-IHF method

1) 令 $J_s = \eta \sum_{i=1}^{N_s} V(\tilde{\mathbf{x}}_i, y_i, f_s) + \frac{1}{2} \|f_s\|_K^2$, $J_t = \eta \sum_{i=1}^{N_t} V(\tilde{\mathbf{x}}_i, y_i, f_t) + \frac{1}{2} \|f_t\|_K^2$, 由式 (1) 可知, J_s 和 J_t 项分别表示源领域和目标领域特有特征的结构风险函数, 其中 $\|f_s\|_K^2$, $\|f_t\|_K^2$ 分别表示源领域和目标领域特有特征对应的泛化能力项.

2) 令 $J_C = \frac{1}{2} \|\hat{f}\|_K^2 + \lambda d_{f,K}(P_s, P_t)$, 则 J_C 包含了源领域与目标领域共享特征的结构风险和实现领域间迁移学习的距离. 其中 $\|\hat{f}\|_K^2$ 表示源领域和目标领域共享特征对应的泛化能力项.

综上所述, 本文提出的融合异构特征的子空间迁移学习方法 (STL-IHF) 在已有特征迁移框架的基础上, 强调融合异构特征的判别信息, 并通过优化使得共享特征和异构特征之间达到较好的平衡, 从而更好地实现迁移学习.

1.3 STL-IHF 的目标函数

根据式 (2) 的迁移学习策略, 本文引入 EMD 距离来表征源领域和目标领域共享特征在再生核

Hilbert 空间中的 $d_{f,K}(P_s, P_t)$. 以支持向量机 (Support vector machine, SVM) 分类器为基础, 给出如下的融合异构特征的子空间迁移学习目标函数:

$$\min_{\mathbf{w}, \mathbf{w}_s, b_s, \mathbf{w}_t, b_t, \xi_i^s, \xi_j^t} \frac{1}{2} \|\mathbf{w}_s\|^2 + \eta_1 \sum_{i=1}^{N_s} \xi_i^s + \frac{1}{2} \|\mathbf{w}_t\|^2 + \eta_2 \sum_{j=1}^{N_t} \xi_j^t + \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \mathbf{w}^T - \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}^i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}^j) \right\| \mathbf{w} \quad (3a)$$

$$\text{s.t. } y_i(\mathbf{w}_s^T \phi(\mathbf{x}_s^i) + \mathbf{w}^T \phi(\mathbf{x}^i) + b_s) \geq 1 - \xi_i^s \quad (3b)$$

$$\xi_i^s \geq 0, \quad \xi_i^s \geq 0, \quad i = 1, 2, \dots, N_s \quad (3c)$$

$$y_j(\mathbf{w}_t^T \phi(\mathbf{x}_t^j) + \mathbf{w}^T \phi(\mathbf{x}^j) + b_t) \geq 1 - \xi_j^t \quad (3d)$$

$$\xi_j^t \geq 0, \quad \xi_j^t \geq 0, \quad j = 1, 2, \dots, N_t \quad (3e)$$

对于上述目标函数, 说明如下:

1) \mathbf{x}_s 表示源领域特有特征数据; \mathbf{x}_t 表示目标领域特有特征数据, \mathbf{x} 表示两个领域共有特征数据; $\mathbf{w}_s, \mathbf{w}_t$ 分别表示源领域特有特征数据 \tilde{X}_s 和目标领域的特有特征数据 \tilde{X}_t 在 RKHS 内待求解的分类器的法向量, \mathbf{w} 表示源领域和目标领域的共享特征 \hat{X} 在 RKHS 内待求解的分类器的法向量, b_s, b_t 为待求解的分类器偏置项; 参数 η_1 和 η_2 是正则化参数; ξ_i^s 和 ξ_j^t 是松弛变量; λ 是结构风险共享特征分布差异间的平衡参数; N_s 和 N_t 为源领域和目标域样本个数.

2) 目标函数中, 除了领域间共享特征、特有特征结构风险项和最大间隔项之外, 还在共享特征空间的法向量中引入了 EMD 距离项, 由文献 [22–23] 的分析可知, 在迁移学习过程中, 要充分考虑到数据的流形结构特征, 所以, RKHS 中引入 EMD 距离能更加有效地减小共享特征间的分布距离, 更好地实现知识迁移.

3) 约束条件式 (3a) 和式 (3b) 亦从领域的特有特征和共享特征分别进行投影, 要求待求的分类器对源领域和目标领域数据要尽量正确分类.

式 (3) 的优化问题可以借助拉格朗日函数方法求解, 但文献 [24] 指出使用拉格朗日函数方法求解 \mathbf{w} 的过程中, 由于数据矩阵行列式为零或非方阵, 因而有时无法求出 \mathbf{w}^{-1} . 本文引入一种逆矩阵转换引理.

引理 1 [25]. 设矩阵 M , 存在两个矩阵 U, V , 使得矩阵 M 可以表示为 $M = E - UV$, 则矩阵 M 的逆矩阵 $M^{-1} = E + U(E - VU)^{-1}V$.

引理 1 给出了一种逆矩阵的变换求解方法, 使得在矩阵 M 求逆的过程中将 U 和 V 互换, 若列向量 $\mathbf{U}_{n \times 1}$ 和行向量 $\mathbf{V}_{1 \times n}$ 相乘, 则得到一个 $n \times n$

的矩阵; 若 $\mathbf{V}_{1 \times n}$ 和 $\mathbf{U}_{n \times 1}$ 交换后相乘, 则得到的是一个实数. 在 STL-IHF 算法中利用引理 1 逆矩阵转换的技巧可将 \mathbf{w} 的逆矩阵问题转化为在再生核 Hilbert 空间中核化问题来求解, 进而避免了求逆和伪逆的问题. 基于引理 1 和拉格朗日函数方法可以将式 (3) 的优化问题转化为利用如下的对偶问题对其求解.

定理 1. 对于式 (3) 所示优化问题, 其对偶问题可表示为如下的二次规划优化形式:

$$\min_{\boldsymbol{\tau}} \frac{1}{2} \boldsymbol{\tau}^T \bar{K} \boldsymbol{\tau} - \mathbf{1}^T \boldsymbol{\tau} \quad (4a)$$

$$\text{s.t. } \sum_{i=1}^{N_s+N_t} \tau_i l_i = 0 \quad (4b)$$

$$0 < \tau_i < \eta, \quad i = 1, \dots, N_s, N_s+1, \dots, N_s+N_t \quad (4c)$$

其中, $\boldsymbol{\tau} = (\alpha_1, \dots, \alpha_{N_s}, \beta_{N_s+1}, \dots, \beta_{N_s+N_t})$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{N_s})$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{N_t})$ 为拉格朗日算子. $\eta = \eta_1 + \eta_2$. $\mathbf{l} = (y_1, \dots, y_{N_s}, y_{N_s+1}, \dots, y_{N_s+N_t})$ 为标签. $\bar{K} = \begin{bmatrix} K_{S,S} & K_{S,T} & K_{T,S} & K_{T,T} \end{bmatrix} \in \mathbf{R}^{(N_s+N_t) \times (N_s+N_t)}$ 为核矩阵, $K_{S,S} = y_i y_j [\phi^T(\mathbf{x}_s^i) \phi(\mathbf{x}_s^j) + \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)]$, $K_{T,T} = y_i y_j [\phi^T(\mathbf{x}_t^i) \phi(\mathbf{x}_t^j) + \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)]$, $K_{S,T} = y_i y_j \phi^T(\mathbf{x}_s^i) H \phi(\mathbf{x}_t^j)$, $K_{T,S} = y_i y_j \phi^T(\mathbf{x}_t^i) H \phi(\mathbf{x}_s^j)$, $H = E - 2CC^T + CC^T CC^T + 2CC^T CC^T CC^T$, $C = \lambda^{\frac{1}{2}} (\frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}^i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}^j))$, $\mathbf{1}$ 表示元素为全 1 的向量.

证明. 构造拉格朗日函数:

$$\begin{aligned} L(\mathbf{w}_s, \mathbf{w}_t, \mathbf{w}, b_s, b_t, \xi_i, \xi_j) &= \frac{1}{2} \mathbf{w}_s^2 + \frac{1}{2} \mathbf{w}_t^2 + \frac{1}{2} \mathbf{w}^2 + \\ &\lambda \mathbf{w}^T \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}_j) \right\| \mathbf{w} + \\ &\eta_1 \sum_{i=1}^{N_s} \xi_i^s + \eta_2 \sum_{j=1}^{N_t} \xi_j^t - \sum_{i=1}^{N_s} \alpha_i^* \xi_i^s - \sum_{j=1}^{N_t} \beta_j^* \xi_j^t - \\ &\sum_{i=1}^{N_s} \alpha_i (y_i (\mathbf{w}_s^T \phi(\mathbf{x}_s^i) + \mathbf{w}^T \phi(\mathbf{x}^i) + b_s) - 1 + \xi_i^s) - \\ &\sum_{j=1}^{N_t} \beta_j (y_j (\mathbf{w}_t^T \phi(\mathbf{x}_t^j) + \mathbf{w}^T \phi(\mathbf{x}^j) + b_t) - 1 + \xi_j^t) \end{aligned} \quad (5)$$

这里拉格朗日乘子 $\alpha_i > 0$, $\beta_j > 0$, $\alpha_i^* > 0$ 和 $\beta_j^* > 0$; 式 (5) 分别对 $\mathbf{w}_s, \mathbf{w}_t, \mathbf{w}, b_s, b_t, \xi_i^s, \xi_j^t$ 求导数, 由 KKT 条件可得:

$$\frac{\partial L}{\partial \mathbf{w}_s} = 0 \Rightarrow \mathbf{w}_s = \sum_{i=1}^{N_s} \alpha_i y_i \phi(\mathbf{x}_s^i) \quad (6)$$

$$\frac{\partial L}{\partial \mathbf{w}_t} = 0 \Rightarrow \mathbf{w}_t = \sum_{j=1}^{N_t} \beta_j y_j \phi(\mathbf{x}_t^j) \quad (7)$$

$$\frac{\partial L}{\partial b_s} = 0 \Rightarrow \sum_{i=1}^{N_s} \alpha_i y_i = 0 \quad (8)$$

$$\frac{\partial L}{\partial b_t} = 0 \Rightarrow \sum_{j=1}^{N_t} \beta_j y_j = 0 \quad (9)$$

$$\frac{\partial L}{\partial \xi_i^s} = \eta_1 - \alpha_i - \alpha_i^* = 0 \Rightarrow 0 < \alpha_i < \eta_1 \quad (10)$$

$$\frac{\partial L}{\partial \xi_i^t} = \eta_2 - \beta_j - \beta_j^* = 0 \Rightarrow 0 < \beta_j < \eta_2 \quad (11)$$

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} + \lambda \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}_j) \right) \times \\ \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}_j) \right)^T \mathbf{w} = \\ \sum_{i=1}^{N_s} \alpha_i y_i \phi(\mathbf{x}^i) + \sum_{j=1}^{N_t} \beta_j y_j \phi(\mathbf{x}^j) \end{aligned} \quad (12)$$

由式 (12), 解得:

$$\begin{aligned} \mathbf{w} = \left[E + \lambda \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}_j) \right) \times \right. \\ \left. \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\mathbf{x}_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(\mathbf{x}_j) \right)^T \right]^{-1} \\ \left(\sum_{i=1}^{N_s} \alpha_i y_i \phi(\mathbf{x}^i) + \sum_{j=1}^{N_t} \beta_j y_j \phi(\mathbf{x}^j) \right) \end{aligned} \quad (13)$$

由引理 1, 化简 \mathbf{w} . 将式 (6)~(11) 和式 (13) 代入式 (5), 化简可得:

$$\begin{aligned} L = -\frac{1}{2} (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T) \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \\ \sum_{i=1}^{N_s} \alpha_i + \sum_{i=1}^{N_t} \beta_j \end{aligned} \quad (14)$$

式 (14) 进一步可表示为式 (4), 从而定理得证. \square

对于式 (4) 的二次规划, 通常需核矩阵正定半正定, 为此需要其满足 Mercer 核条件, 这样才能保证式 (4) 的二次规划为凸二次规划, 进而满足了 KKT 条件, 使得式 (4) 的二次规划的解为全局最小解. 为了验证这一条件, 给出如下定理:

引理 2^[24]. 设 X 是 \mathbf{R}^n 上的一个紧集, 若 $K(x_i, x_j)$ 是 $X \times X$ 上的连续对称函数且关于任

意 $x_i \in X$ 的 Gram 矩阵半正定, 则 $K(x_i, x_j)$ 是 Mercer 核.

定理 2. 形如式 (4) 中 $\bar{K} = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \in \mathbf{R}^{(N_s+N_t) \times (N_s+N_t)}$ 所表达的核函数是 Mercer 核.

证明. 由定理 1 可知, \bar{K} 中的 $K_{S,S} = y_i y_j [\phi^T(\mathbf{x}_s^i) \phi(\mathbf{x}_s^j) + \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)]_{N_s \times N_s}$, 由于 $\phi^T(\mathbf{x}_s^i) \phi(\mathbf{x}_s^j)$ 为实对称阵, 且 $H = E - 2CC^T + CC^T CC^T + 2CC^T CC^T CC^T$, 即 $H = (E - CC^T)^2 + 2CC^T CC^T CC^T$. 展开化简后, $\phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)$ 亦为实对称阵, 因此 $K_{S,S}$ 矩阵是一个对称矩阵. 其次, 对任意的 α_i , 则:

$$\begin{aligned} \sum_{i,j} \alpha_i \alpha_j (\phi^T(\mathbf{x}_s^i) \phi(\mathbf{x}_s^j)) = \\ \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}_s^i) \right)^T \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}_s^i) \right) \geq 0 \end{aligned} \quad (15)$$

$$\begin{aligned} \sum_{i,j} \alpha_i \alpha_j \phi(\mathbf{x}^i)^T H \phi(\mathbf{x}^j) = \\ \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}^i) (E - CC^T) \right)^T \times \\ \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}^i) (E - CC^T) \right) + \\ 2 \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}^i) CC^T C \right)^T \times \\ \left(\sum_{i=1}^{N_s} \alpha_i \phi(\mathbf{x}^i) CC^T C \right) \geq 0 \end{aligned} \quad (16)$$

由式 (15) 和式 (16) 可知, $K_{S,S} = y_i y_j [\phi^T(\mathbf{x}_s^i) \phi(\mathbf{x}_s^j) + \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)]_{N_s \times N_s} \geq 0$, 因而, $K_{S,S}$ 为半正定矩阵且 $K_{S,S}$ 为 $X \times X$ 上连续对称函数; 同理, $K_{T,T} = y_i y_j [\phi^T(\mathbf{x}_t^i) \phi(\mathbf{x}_t^j) + \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)]$, $K_{S,T} = y_i y_j \phi^T(\mathbf{x}^i) H \phi(\mathbf{x}^j)$, $K_{T,S} = y_i y_j \phi^T(\mathbf{x}^j) H \phi(\mathbf{x}^i)$, $K_{S,T} \geq 0$, $K_{T,S} \geq 0$ 和 $K_{T,T} \geq 0$ 均为半正定矩阵, 且 $X \times X$ 上均为连续对称函数, 由引理 2 可知, 核 $\bar{K} = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \in \mathbf{R}^{(N_s+N_t) \times (N_s+N_t)}$ 为 Mercer 核矩阵.

1.4 STL-IHF 算法

根据前节推导和分析, 这里给出如下的 STL-IHF 算法.

输入. 带标记的源领域数据集和带标记的目标领域数据集;

预处理. 通过特征选择方法, 分别提取源领域特

有特征数据 \mathbf{x}_s , 目标领域特有特征数据 \mathbf{x}_t 和共享特征数据 \mathbf{x} , 将数据标准化

步骤 1. 设置核窗宽参数, 构造最优化问题式 (4);

步骤 2. 根据式 (4) 求解拉格朗日系数 α 和 β , 从而根据式 (6)、式 (7) 和式 (13) 得到对应的 \mathbf{w}_s , \mathbf{w}_t , \mathbf{w} , 然后, 通过核化求出 b_s 和 b_t ;

步骤 3. 由 \mathbf{w}_s , \mathbf{w}_t , \mathbf{w} , b_s 和 b_t , 构造决策函数 $f_s(\mathbf{x})$, $f_t(\mathbf{x})$;

输出. 源领域和目标领域分类决策函数.

2 实验分析

为了验证本文所提方法 STL-IHF 的有效性, 本节将在不同类型数据集上对其进行性能评估, 所采用的数据集主要包括如下两大类: 1) 模拟数据集; 2) 真实数据集, 包括 Yale 和 ORL 人脸分类数据集. 实验中将 STL-IHF 与相关的方法进行了性能比较, 用于比较的方法有: SVM, LSSVM, LP-SVM 和 TrAdaBoost. 比较算法的参数均根据具体应用通过 5 重交叉验证获得.

本文算法均采用高斯核函数 (RBF): $K(\mathbf{x}, \mathbf{y}) = \exp(-\frac{1}{\gamma}\|\mathbf{x} - \mathbf{y}\|^2)$, 其中窗宽参数 γ 值通过网格搜索选取, $\gamma \in \{\sigma^2/64, \sigma^2/32, \sigma^2/16, \sigma^2/8, \sigma^2/4, \sigma^2/2, \sigma^2, 2\sigma^2, 4\sigma^2, 8\sigma^2, 16\sigma^2, 32\sigma^2, 64\sigma^2\}$, σ 为训练样本平均范数的平方根. 对于 STL-IHF 算法中平衡参数 λ 和正则化参数取 $\eta_1 = \eta_2 = \eta$, 均在区间 $\{10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}\}$ 中搜索选取.

实验环境: Windows XP, CPU Intel T4300 2.1 GHz, RAM 3 G, Matlab R2010a 等.

2.1 模拟数据集

2.1.1 模拟数据集预处理

本部分实验生成了 2 个模拟数据集, 现分别介绍如下:

1) 数据集 1 (DS1): 源领域和目标领域都是 4

维数据点, 领域间共享特征 \hat{X} 是 2 维数据, 由均匀分布和单峰数据构成, 如图 2(a) 所示; 领域间共享特征的相似程度通过添加高斯噪声控制, 噪声的标准差表示为 $noise_sigma$; 特有特征 \tilde{X}_s 和 \tilde{X}_t 是 2 维团状高斯数据集, 如图 2(b) 所示; DS1 数据集每类样本的个数为 1000 个; 期望和标准差见表 1.

2) 数据集 2 (DS2): 源领域和目标领域亦是 4 维数据点, 领域间的 \hat{X} 是 2 维模拟香蕉数据, 目标领域共享特征数据由源领域数据 \hat{X} 旋转 10° 后获得, 如图 2(c) 所示, 领域特征空间相似程度由旋转角度 θ 控制、特有特征数据不同参数的 2 维高斯分布的数据构成, 如图 2(d) 所示; DS2 数据集每类样本个数都是 150 个; 期望和标准差见表 1.

3) 模拟数据集在实验开始前均进行了标准化预处理, 其中 D_s^+ , D_s^- 和 D_t^+ , D_t^- 分别代表源领域和目标领域中正负类样本. 样本的各参数如表 1 所示.

4) 为了比较说明模拟数据两个领域数据之间的关系, 分别画出 DS1 和 DS2 在 (x_1, x_2) 和 (x_3, x_4) 视角上的投影, 如图 2 所示.

实验步骤: 目标领域数据每次随机选取 20% 作为有标签数据, 然后利用源领域数据和目标领域有标签数据进行训练, 并利用 5 重交叉验证获取最优参数, 再利用目标领域余下数据来测试受训所得模型. 上述过程重复 10 次, 所得平均测试精度及标准差作为算法的性能指标, 并与其他算法进行比较.

从图 2(a) 和图 2(c) 上可以看出两个领域的特征子空间上存在分布相近的数据, 为特征空间迁移学习提供了有力的知识参照. 另外, 从图 2(b) 和图 2(d) 上可以看出两个领域数据集的特有特征在各自特征方向上的投影都保持了独特的异构信息, 通过融入该异构信息可进一步增强分类模型的泛化性能.

为了说明领域间相似特征部分数据对迁移学习的影响, 领域间的相似程度可以由两个参数控制, 因此, 本部分实验中噪声参数分别选择 $noise_sigma = 0.5, 2, 10$; 对于数据集 2 领域间共享部分的旋转角度 $\theta = 10^\circ, 20^\circ, 30^\circ$, 比较在不同相似程度下各算法的分类性能.

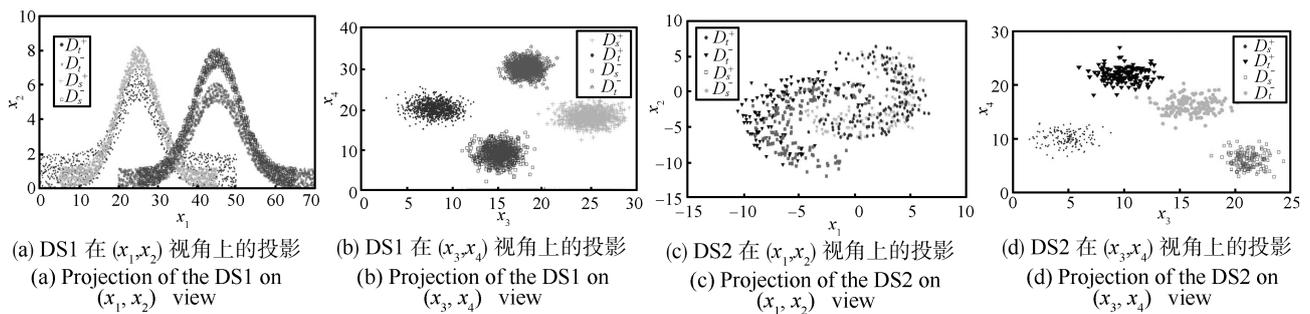


图 2 实验数据集不同特征的示意图

Fig. 2 Illustration of data sets adopted in following experiments

表 1 模拟数据集的构造
Table 1 Structure of synthetic datasets

| 数据集 | 参数 | 正类 | 负类 | 正类 | 负类 |
|-------|----|----------------------|---------------------|---------------------|---------------------|
| | 数量 | 1000 | 1000 | 1000 | 1000 |
| 数据集 1 | 维数 | 4 | 4 | 4 | 4 |
| DS1 | 期望 | [25 3.5 25 18] | [45 3.5 15 9] | [25 3 8 20] | [45 3 18 30] |
| | 方差 | [10.53 5.08 2.3 2.2] | [10.53 5.1 1.4 3.3] | [8.43 6 2.14 1.34] | [8.43 5.5 1 2] |
| | 数量 | 150 | 150 | 150 | 150 |
| 数据集 2 | 维数 | 4 | 4 | 4 | 4 |
| DS2 | 期望 | [1.91 -1.83 5 10] | [-5.91 -2.51 21 6] | [2.42 -1.06 10 22] | [-4.69 -4.37 16 16] |
| | 方差 | [2.9 3.36 1.8 1.5] | [2.9 3.48 1.2 2.2] | [2.45 3.73 2.4 1.6] | [2.44 3.82 2.5 2.2] |

表 2 模拟数据集分类精度和标准差 (%)
Table 2 Percentage overall accuracy and standard deviation exhibited on different domain instances (%)

| 数据集 | 参数 | SVM | LSSVM | TrAdaBoost | LP-SVM | STL-IHF |
|-------|------------------------|--------------|--------------|--------------|--------------|--------------|
| 数据集 1 | $noise_{\sigma} = 0.5$ | 85.76 ± 3.72 | 85.09 ± 1.75 | 91.24 ± 0.62 | 94.88 ± 0.77 | 98.35 ± 0.64 |
| DS1 | $noise_{\sigma} = 2$ | 83.34 ± 1.41 | 83.75 ± 1.82 | 85.61 ± 1.27 | 90.28 ± 1.73 | 90.61 ± 0.51 |
| | $noise_{\sigma} = 10$ | 78.90 ± 2.58 | 80.44 ± 2.86 | 82.26 ± 1.75 | 83.43 ± 1.39 | 86.93 ± 1.03 |
| 数据集 2 | $\theta = 10^{\circ}$ | 84.12 ± 1.44 | 84.36 ± 0.46 | 89.14 ± 1.19 | 91.97 ± 0.49 | 96.64 ± 0.39 |
| DS2 | $\theta = 20^{\circ}$ | 83.51 ± 0.88 | 82.10 ± 0.67 | 87.47 ± 0.78 | 89.70 ± 1.10 | 92.28 ± 0.59 |
| | $\theta = 30^{\circ}$ | 79.72 ± 1.93 | 78.82 ± 2.41 | 84.82 ± 1.69 | 86.13 ± 2.05 | 86.88 ± 1.83 |

2.1.2 模拟数据集实验结果及分析

利用本文提出的 STL-IHF 方法分别与传统的 SVM 和 LSSVM、实例迁移学习算法 TrAdaBoost 和 LP-SVM 方法进行比较. 实验结果如表 2 所示.

1) 传统的 SVM 和 LSSVM 方法在所有数据集上的分类效果在一定程度上均逊于其他 3 种迁移学习方法. 这也验证了如果仅考虑数据集的最大间隔, 而忽视数据空间的结构信息尤其是保持局部流形的特征信息, 不能保证较优的分类性能. 而在考虑了数据空间的本质流形结构的情况下, STL-IHF 方法在数据集 1 和数据集 2 的模式分类性能优于 TrAdaBoost 和 LP-SVM 等方法. 这进一步说明, 在充分考虑领域数据空间的局部几何特征结构和特有判别信息的情况下, STL-IHF 方法能取得较好的泛化性能且算法稳定性较高.

2) 由表 2 可以看出, 在噪音参数 $noise_{\sigma}$ 和旋转角度 θ 的数值相对较小时, 三种具有迁移学习能力方法的分类性能明显优于传统的 SVM 和 LSSVM 方法, 随着 $noise_{\sigma}$ 和 θ 的逐渐增加, TrAdaBoost, LP-SVM 和 STL-IHF 方法分类精度虽然都呈现下降趋势, 但 STL-IHF 方法的分类性

能还是明显优越于其他 4 种方法, 这是因为 TrAdaBoost 算法需要利用与源领域相似的数据, 但是随着 $noise_{\sigma}$ 和 θ 值增大, 领域间相似的程度逐渐减小, 甚至不相似, 适用于目标领域的源领域的数据很少, 所以, TrAdaBoost 算法性能下降明显; LP-SVM 算法是借助领域间的辅助数据进行迁移学习, 减小训练数据和测试数据之间的分布距离来提高分类性能, 但随着噪声的增大使得可利用的辅助数据减少, 因而, LP-SVM 方法的分类性能减小. STL-IHF 算法是在共享特征的基础上又考虑了数据的特有判别信息, 虽然随着噪声增大共享特征间的相似性减弱, 但借助 EMD 和特有特征的判别信息, 使得 STL-IHF 仍能保持较好的迁移学习性能. 实验证明随着噪音参数 $noise_{\sigma}$ 和旋转角度 θ 值的变化, STL-IHF 算法具有更好的鲁棒性.

2.2 人脸数据集实验与分析

为了进一步评价本文所提方法在真实数据集下的分类性能, 选取 Yale 人脸数据和 ORL 人脸数据进行了实验, 因为人脸图像数据集呈现出明显的非线性流形结构, 被很多迁移学习方法用来作为测试数据集^[26-27].

2.2.1 实验设置

1) 数据构造: 为了多角度说明本文所提方法的有效性, 本节实验分别构造了同数据源和交叉数据源两类数据集进行实验.

a) 同数据源是指源领域和目标领域数据均来自同一数据库.

i) Yale 人脸数据集源领域包括 15 个类别的人脸数据, 同一类由不同光照条件和脸部表情的 11 张图片构成 (见图 3(a)); 将源领域图像旋转 10° 后获得目标领域图像, 如图 3(b); 在目标领域数据集中随机抽取每个子集的部分图像 (本节实验分别取 2, 4 和 5 个图像) 和源领域数据共同作为训练集, 其余目标领域数据作为测试子集.



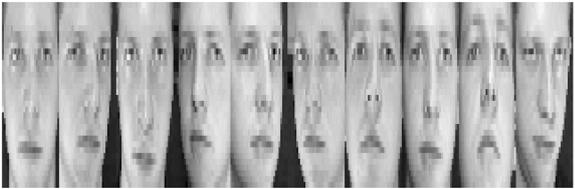
(a) 源领域 Yale 数据集中某一类对象的人脸图像

(a) Yale face images for one object



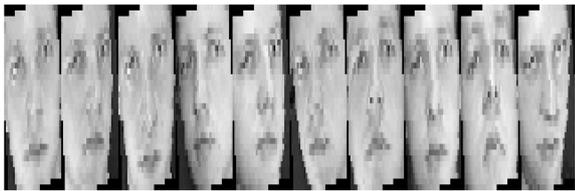
(b) 旋转 10° 后得到的目标领域 Yale 数据集中某一类对象的人脸图像

(b) Yale face images of an object with rotation angle 10°



(c) 源领域 ORL 数据集中某一类对象的人脸图像

(c) ORL face images for one object



(d) 旋转 10° 后得到的目标领域 ORL 数据集中某一类对象的人脸图像

(d) ORL face images of an object with rotation angle 10°

Fig. 3 Cropped face image in the face image datasets

ii) ORL 人脸数据集源领域包含 40 个类别的人脸数据, 同一类中有 10 种不同表情的图片构成 (见

图 3(c)); 将源领域图像旋转 10° 后获得目标领域图像, 如图 3(d). 在目标领域数据集中随机抽取每个数据子集部分图像 (本节实验分别取 2, 4 和 5 个图像) 和源领域共同作为训练集, 其余目标领域数据作为对应的测试数据集.

b) 交叉数据源是指源领域和目标领域数据来自不同的数据库.

作为对比实验, 源领域选取 Yale 数据库中的某一类图像作为源领域的正类样本, 其余类作为负类; 将源领域正类旋转 10° 后作为目标领域的正类样本, 将 ORL 数据库中的各类作为目标领域的负类样本. 在目标领域数据集中随机抽取每个数据子集部分图像 (本节实验分别取 3, 5 和 7 个图像) 和源领域共同作为训练集, 其余目标领域数据作为对应的测试数据集.

这里需要说明一点: 实验前把训练集中的每个人脸数据图像缩放成像素大小, 则每个图像可由一个 1024 维向量表示 (见图 3(a) 和图 3(c)). 实验步骤同上.

2) 数据提取: 因为人脸图像数据被表达为一个向量, 为了提取源领域和目标领域特有和共享特征信息, 本文采用了小波变换的分析方法, 通过伸缩和平移等功能对图像进行多尺度细化分析和处理. 文献 [28] 指出在采用特征脸进行人脸识别时, 双正交小波和三次 B 样条小波对图像进行小波变换具有较好的识别率和较快的识别速度. 因此, 在本部分实验中使用三次 B 样条小波基提取领域的共享特征和特有特征.

首先, 选用三次 B 样条小波基对源领域和目标领域人脸图像数据集共同进行特征分解, 在相同参数条件下得到体现人脸特征的小波低频细节, 再经过水平和垂直投影运算, 确定出人脸识别所需要的人脸关键特征作为共享特征空间 \hat{X} . 然后, 用三次 B 样条小波基分别对源领域和目标领域单独进行特征提取高频细节作为特有特征 \tilde{X}_s 和 \tilde{X}_t .

2.2.2 实验结果及分析

表 3 和表 4 中记录了各算法在同源 Yale 数据集和同源 ORL 数据集的实验结果. 表 5 记录了各算法在交叉数据源上的实验结果. 从实验结果可以给出如下观察:

1) LS-SVM 方法采用最小二乘线性系统作为损失函数, 代替传统的支持向量机采用的二次规划方法, 简化了计算的复杂性, 因此运算速度明显快于 SVM; 但 SVM 和 LS-SVM 方法仅考虑最大风险原则, 未能考虑到领域间的迁移, 故 SVM 和 LSSVM 的分类性能均低于其他三种迁移学习方法.

2) 从表 3 和表 4 的实验结果可以看出, 对于同数据源的人脸数据集, 虽然光照角度和表情有一定的

表 3 Yale 数据集分类精度和标准差 (%)

Table 3 Accuracy and standard deviation on Yale datasets with different ratio (%)

| 数据集 | 目标域数据训练/测试 | SVM | LSSVM | TrAdaBoost | LP-SVM | STL-IHF |
|------|------------|--------------|--------------|--------------|--------------|--------------|
| Yale | 2/9 | 66.98 ± 4.09 | 67.14 ± 5.27 | 74.83 ± 4.41 | 81.32 ± 3.55 | 85.86 ± 2.67 |
| | 4/7 | 69.51 ± 3.68 | 72.28 ± 3.91 | 82.08 ± 3.02 | 84.61 ± 1.80 | 89.61 ± 1.90 |
| | 5/6 | 71.42 ± 4.82 | 74.00 ± 2.96 | 89.46 ± 3.09 | 90.81 ± 2.26 | 93.37 ± 1.37 |

表 4 ORL 数据集分类精度和标准差 (%)

Table 4 Accuracy and standard deviation on ORL datasets with different ratio (%)

| 数据集 | 目标域数据训练/测试 | SVM | LSSVM | TrAdaBoost | LP-SVM | STL-IHF |
|-----|------------|--------------|--------------|--------------|--------------|--------------|
| ORL | 2/8 | 77.34 ± 2.79 | 80.00 ± 3.42 | 87.35 ± 2.96 | 87.51 ± 2.49 | 94.28 ± 1.64 |
| | 4/6 | 78.88 ± 0.92 | 82.15 ± 1.04 | 92.32 ± 1.89 | 91.03 ± 0.99 | 95.84 ± 0.74 |
| | 5/5 | 82.00 ± 0.99 | 86.55 ± 1.07 | 93.08 ± 1.14 | 92.27 ± 0.69 | 96.99 ± 1.29 |

表 5 交叉数据源分类精度和标准差 (%)

Table 5 Accuracy and standard deviation on across dataset with different ratio (%)

| 数据集 | 目标域数据训练/测试 | SVM | LSSVM | TrAdaBoost | LP-SVM | STL-IHF |
|--------------|------------|--------------|--------------|--------------|--------------|--------------|
| Yale vs. ORL | 3/8 | 78.04 ± 2.54 | 81.31 ± 3.06 | 85.09 ± 2.68 | 86.02 ± 2.24 | 91.84 ± 1.16 |
| | 5/6 | 81.64 ± 1.85 | 84.80 ± 0.91 | 92.79 ± 0.72 | 92.91 ± 0.94 | 93.79 ± 0.93 |
| | 7/4 | 83.71 ± 1.16 | 87.14 ± 1.01 | 94.14 ± 0.92 | 93.02 ± 1.15 | 96.73 ± 0.62 |

变化,但是对于同一类人脸数据的主要信息在某些属性上还是具有很大相似性,提取这部分相似信息对提高领域间的学习能力必然起到积极的作用;STL-IHF 算法在训练之前就把数据属性分为共享和特有特征,在共享特征之间迁移学习,这样就更加具有针对性、适应性和灵活性,在人脸数据集上获得了比 TrAdaBoost 和 LP-SVM 更好的分类性能. STL-IHF 方法的识别性能高于其他几种方法,也说明了在考虑了领域间共有特征的基础上,融合异构特征信息的迁移学习机更加符合实际场景,因而取得了相对较好的识别性能.

3) 从表 3 和表 4 的实验结果可以看出,随着目标领域数据作为训练数据数目的增加, SVM 和 LS-SVM 分类精度增加不明显,而其他三种方法的分类性能都在一定程度上有所增加,但 STL-IHF 方法仍高于其他两种迁移学习方法,这说明 STL-IHF 方法中引入局部结构约束对提高人脸识别效果起到一定的作用.特别是目标领域的子数据集训练数据个数增加到 5 个时,目标领域信息为 STL-IHF 算法提供了更多的信息支撑,充分考虑了数据共有特征和特有特征的 STL-IHF 迁移学习方法具备较强的识别能力,STL-IHF 方法较其他算法稳定性也较好.

4) 从表 5 的实验结果可以看出,对于交叉数据

源的数据集而言,随着训练数据数目的增加,各种方法的分类性能都有所提高.与表 4 相比较,TrAdaBoost 和 LP-SVM 方法在交叉数据源下实验的分类性能略有增加,而基于共享特征并融合异构信息的 STL-IHF 方法性能在源领域和目标领域差别较大时分类性能略有下降,说明领域间共享特征对 STL-IHF 方法的性能影响比异构信息更明显.但相比于其他方法,STL-IHF 方法的性能和算法稳定性仍相对较高.

3 结论

本文针对已有特征迁移学习算法忽略领域间特有异构信息的问题,提出了一种融合异构特征的子空间迁移学习算法.相对于已有的方法,本文提出的 STL-IHF 方法在共享特征上实现知识迁移的同时兼顾了领域间特有的异构信息,并以嵌入期望分布距离度量距离确保共享特征在核空间的中心位置的尽量靠近,强化了源领域和目标领域间的知识迁移.实验分析验证了所提方法在分类问题上的良好适应性.虽然本文方法展现出了较好的有效性和稳定性,其依然面临一些需要进一步探讨的问题.例如:如何进一步设计更好的兼顾异构特征的迁移学习策略将

是一个非常有意义的工作, 我们将在未来的工作中做更深入的研究.

References

- 1 Pan J L, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, **22**(10): 1345–1359
- 2 Duan L X, Tsang I W, Xu D. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(3): 465–479
- 3 Tu W T, Sun S L. A subject transfer framework for EEG classification. *Neurocomputing*, 2012, **82**: 109–116
- 4 Daume III H, Marcu D. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 2006, **26**(1): 101–126
- 5 Biekel S, Bruckner M, Scheffer T. Discriminative learning for differing training and test distributions. In: Proceedings of the 24th International Conference on Machine Learning. New York, USA: ACM, 2007. 81–88
- 6 Bickel S, Sawade C, Scheffer T. Transfer learning by distribution matching for targeted advertising. In: Proceedings of the 21st Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2009. 145–152
- 7 Wu P C, Dietterich T G. Improving SVM accuracy by training on auxiliary data sources. In: Proceedings of the 21st International Conference on Machine Learning (ICML). New York, USA: ACM, 2004. 110–117
- 8 Dai W Y, Yang Q, Xue G R, Yu Y. Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning (ICML). New York, USA: ACM, 2007. 193–200
- 9 Quanz B, Huan J. Large margin transductive transfer learning. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM). New York, USA: ACM, 2009. 1327–1336
- 10 Xu Z J, Sun S L. Multi-view transfer learning with Adaboost. In: Proceedings of the 23rd Conference on Tools with Artificial Intelligence. Boca Raton, FL: IEEE, 2011. 399–402
- 11 Chen M M, Weinberger K Q, Blitzer J. Co-training for domain adaptation. In: Proceedings of the 25th Conference on Neural Information Processing Systems (NIPS). New York: USA: Curran Associates, Inc., 2011. 1231–1240
- 12 Xu Z J, Sun S L. Multi-source transfer learning with multi-view adaboost. *Neural Information Processing*, 2012, **7665**: 332–339
- 13 Jiang Yi-Zhang, Deng Zhao-Hong, Wang Shi-Tong. Mamdani-Larsen type transfer learning fuzzy system. *Acta Automatica Sinica*, 2012, **38**(9): 1393–1409
(蒋亦樟, 邓赵红, 王士同. ML 型迁移学习模糊系统. 自动化学报, 2012, **38**(9): 1393–1409)
- 14 Zhu Mei-Qiang, Cheng Yu-Hu, Li Ming, Wang Xue-Song, Feng Huan-Ting. A hybrid transfer algorithm for reinforcement learning based on spectral method. *Acta Automatica Sinica*, 2012, **38**(11): 1765–1776
(朱美强, 程玉虎, 李明, 王雪松, 冯涣婷. 一类基于谱方法的强化学习混合迁移算法. 自动化学报, 2012, **38**(11): 1765–1776)
- 15 Jiang W H, Chung F L. Transfer spectral clustering. In: Proceedings of the 2012 European Conference on Machine Learning and Principles of Knowledge Discovery in Databases (ECML PKDD). Berlin, Heidelberg: Springer-Verlag, 2012. 789–803
- 16 Argyriou A, Michelli C A, Pontil M, Ying Y M. A spectral regularization frame work for multi-task structure learning. In: Proceedings of Advances in Neural Information Processing Systems (NIPS 2008). Cambridge, MA: MIT Press, 2007. 25–32
- 17 Pan S J L, Kwok J T, Yang Q. Transfer learning via dimensionality reduction. In: Proceedings of the 23rd International Conference on Artificial Intelligence. California, USA: AAAI Press, 2008. 677–682
- 18 Pan S J L, Ni X C, Sun J T, Yang Q, Chen Z. Cross-domain sentiment classification via spectral feature alignment. In: Proceedings of the 19th International Conference on World Wide Web (WWW'10). New York, USA: ACM, 2010. 751–760
- 19 Tu W T, Sun S L. Transferable discriminative dimensionality reduction. In: Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence (CTAI). Boca Raton, FL: IEEE, 2011. 865–868
- 20 Gao X B, Wang X M, Li X L, Tao D C. Transfer latent variable model based on divergence analysis. *Pattern Recognition*, 2011, **44**(10–11): 2358–2366
- 21 Gao X B, Wang Z, Yan P K, Li X L. Transfer learning for pedestrian detection. *Neurocomputing*, 2013, **100**: 51–57
- 22 Gretton A, Fukumizu K, Harchaoui Z, Sriperumbudur B K. A fast, consistent kernel two-sample test. In: Proceedings of Advances in Neural Information Processing Systems (NIPS 2010). Red Hook, NY: MIT Press, 2010. 673–681
- 23 Gao Jun, Wang Shi-Tong, Deng Zhao-Hong. Global and local preserving based semi-supervised support vector machine. *Acta Electronica Sinica*, 2010, **38**(7): 1626–1633
(皋军, 王士同, 邓赵红. 基于全局和局部保持的半监督支持向量机. 电子学报, 2010, **38**(7): 1626–1633)
- 24 Deng Nai-Yang, Tian Ying-Jie. *New Method in Data Mining: Support Vector Machine*. Beijing: Science Press, 2004. 5–16
(邓乃阳, 田英杰. 数据挖掘中的新方法: 支持向量机. 北京: 科学出版社, 2004. 5–16)
- 25 Yuan Ya-Xiang. *Optimization Theory and Methods*. Beijing: Science Press, 1997. 176–189
(袁亚湘. 最优化理论与方法. 北京: 科学出版社, 1997. 176–189)
- 26 Tao Jian-Wen, Wang Shi-Tong. Kernel support vector machine for domain adaptation. *Acta Automatica Sinica*, 2012, **38**(5): 797–881
(陶剑文, 王士同. 领域适应核支持向量机. 自动化学报, 2012, **38**(5): 797–881)

- 27 Cai D, He X F, Han J W, Zhang H J. Orthogonal Laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 2006, **15**(11): 3608–3614
- 28 Zhang J Y, Zhang B, Jiang X Z. Analyses of feature extraction methods based on wavelet transform. *Signal Processing*, 2000, **16**(2): 156–162



张景祥 江南大学数字媒体学院博士研究生. 主要研究方向为模式识别, 智能计算及应用. 本文通信作者.

E-mail: zjx145@163.com

(**ZHANG Jing-Xiang** Ph.D. candidate at the School of Digital Media, Jiangnan University. His research interest covers pattern recognition, intel-

ligent computation and their application. Corresponding author of this paper.)



王士同 江南大学数字媒体学院教授. 主要研究方向为人工智能, 模式识别和生物信息.

E-mail: wxwangst@yahoo.com.cn

(**WANG Shi-Tong** Professor at the School of Digital Media, Jiangnan University. His research interest covers artificial intelligence, pattern recognition,

and bioinformatics.)



邓赵红 博士, 江南大学数字媒体学院副教授. 主要研究方向为模糊建模和智能计算.

E-mail: dzh666828@yahoo.com.cn

(**DENG Zhao-Hong** Ph.D., associate professor at the School of Digital Media, Jiangnan University. His research interest covers fuzzy modeling and intelligent computation.)



蒋亦樟 江南大学数字媒体学院博士研究生. 主要研究方向为模式识别, 智能计算及应用. E-mail: jyz0512@163.com

(**JIANG Yi-Zhang** Ph.D. candidate at the School of Digital Media, Jiangnan University. His research interest covers pattern recognition, intelligent computation and their applica-

tion)



李奕 江南大学数字媒体学院博士研究生. 主要研究方向为模式识别, 智能计算及应用. E-mail: liyi@163.com

(**LI Yi** Ph.D. candidate at the School of Digital Media, Jiangnan University. Her research interest covers pattern recognition, intelligent computation and their application.)