

基于张量表示的域适配的 迁移学习中特征表示方法

赵 鹏,王美玉,纪 霞,刘慧婷

(1. 安徽大学计算机智能与信号处理教育部重点实验室,安徽合肥 230601;2. 安徽大学计算机科学与技术学院,安徽合肥 230601)

摘 要: 本文提出一种新的基于张量表示的域适配迁移学习中的特征表示方法,即融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法. 当源域和目标域差异很大时,仅将源域对齐潜在在共享空间,会造成数据扭曲过大. 为缓解此问题,本文方法提出联合域对齐,即源域和目标域同时对齐共享子空间. 并且本文方法将适配正则化引入张量表示空间求解. 本文适配正则化包括动态分布对齐和图适配,以缩小域间分布差异和保留样本间流行一致性. 最后融合联合域对齐,动态分布对齐和图适配,通过联合优化求解获得共享子空间表示. 几个公共的跨域数据集上的大量实验结果表明了本文方法优于其它主流的迁移学习方法,验证了本文方法的有效性.

关键词: 张量表示; 迁移学习; 域适配

中图分类号: TP391.41

文献标识码: A

文章编号: 0372-2112 (2020)02-0359-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2020.02.019

A Novel Feature Representation Based on Tensor and Domain Adaption for Transfer Learning

ZHAO Peng, WANG Mei-yu, JI Xia, LIU Hui-ting

(1. Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, Anhui 230601;
2. School of Computer Science and Technology, Anhui University, Hefei, Anhui 230601, China)

Abstract: A novel feature representation based on tensor and domain adaption for transfer learning is proposed, which combines joint domain alignment and adaptation regularization. When the difference between the source domain and the target domain is very large, only aligning the source domain to the potential shared subspace will result in too much data distortion. To alleviate this problem, this paper proposes joint domain alignment, which aligns the source domain and the target domain to the potential shared subspaces simultaneously. Furthermore, the adaption regularization is introduced into the subspace learning based on tensor. In the proposed method, adaptation regularization includes dynamic distribution alignment and graph adaptation to reduce the distribution differences among different domains and preserve the manifold consistency. Finally, the joint domain alignment, dynamic distributed alignment and graph adaptation are fused, and the joint optimization is utilized to obtain the feature representation. Extensive experiments on several common cross-domain datasets show that the proposed method outperforms the state-of-the-art on the tasks of transfer learning and the effectiveness of the proposed method is verified.

Key words: tensor representation; transfer learning; domain adaption

1 引言

随着科技发展,各类数据采集工具层出不穷,数据爆炸式增长以及数据的异质性日益突出. 基于数据独

立同分布假设的传统的机器学习方法无法适应新的学习任务. 迁移学习能够较好地解决标记数据稀缺和领域间分布差异的问题,因而受到了极大的关注.

现有迁移学习特征表示方法主要分为以下三类:

收稿日期:2019-05-22;修回日期:2019-09-12;责任编辑:马兰英

基金项目:国家自然科学基金(No. 61602004);安徽省高校自然科学研究重点项目(No. KJ2018A0013, No. KJ2017A011);安徽省自然科学基金(No. 1908085MF188, No. 1908085MF182);安徽省重点研究与开发计划项目(No. 1804d08020309)

(1) 数据分布适配法 (Distribution Adaptation), 从数据分布的角度来考虑, 缩小源域和目标域分布差异, 最小化概率分布差异. 迁移成分分析 (Transfer Component Analysis, TCA)^[1] 和最大均值差异嵌入 (Maximum Mean Discrepancy Embedding MMDE)^[2] 通过适配源域和目标域的边缘概率分布来迁移学习特征表示. 联合分布适配 (Joint Distribution Adaptation, JDA)^[3] 和平衡分布适配 (Balanced Distribution Adaptation, BDA)^[4] 同时适配两个域的边缘概率分布和条件概率分布. 域正则成分分析 (Domain Regularized Component Analysis, DRCA)^[5] 通过引入核匹配准则降低边缘分布和条件分布的差异. (2) 特征选择法 (Feature Selection), 从特征选择的角度来考虑, 从源域和目标域中选择提取共享的特征, 建立统一模型. 结构对应学习 (Structural Correspondence Learning, SCL)^[6] 寻找关键的特征, 将源域和目标域对齐. Satpal 等人通过特征子集建立的条件概率模型适配源域和目标域的条件概率分布^[7]. (3) 子空间学习法 (Subspace Learning), 从特征变换的角度来考虑, 学习一个潜在子空间, 将源域和目标域变换到共享的潜在子空间. 子空间对齐法 (Subspace Alignment, SA)^[8] 直接寻求一个线性变换, 把源空间变换到目标空间中. 散射成分分析 (Scatter component analysis, SCA) 将样本转化为一组子空间, 然后将它们的差异最小化^[9]. 成对双线性判别映射 (Coupled Bilinear Discriminant Projection, CB-DP)^[10] 对齐图像样本, 并且不损失空间信息.

大多数现有的迁移学习方法是基于向量的表示方法, 因而需要对数据进行向量化, 而现有很多应用中, 向量或矩阵无法自然的表示和保存重要的结构信息. 本文提出了一种新的迁移学习中基于张量表示的域适配特征表示方法, 即融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法 (Tensor based Transfer Learning combining Joint-Domain Alignments and Adaption Regulation, TTL-JDAAR).

不同于之前的相关工作, 本文的主要贡献包括以下三方面. (1) 本文提出联合域对齐的基于张量的特征表示, 缓解仅源域对齐共享子空间造成源域空间数据扭曲过大的问题. (2) 本文在基于张量的迁移学习特征表示方法中引入了动态分布对齐, 在张量子空间下动态适配源域和目标域的边缘概率分布和条件概率分布, 以减小源域和目标域的差异. (3) 本文在基于张量的迁移学习特征表示方法中引入了图适配, 保持张量空间下特征学习中的流行一致性.

2 相关工作

2.1 张量表示

张量是数据的多路阵列表示, 是矩阵的一种扩展.

在计算机视觉中, 图像的张量表示避免图像信息的损失, 因此张量表示比矩阵表示更有效^[11]. 其中, Ben 等人^[12] 提出的高阶张量表示框架在跨视角的步态识别中取得了显著的效果.

张量分解是指一个张量可以表示为有限个秩 1 张量之和. 本文采用的张量分解方法是 Tucker 分解, 又称高阶奇异值分解^[13] (higher-order SVD). Tucker 分解将 K 阶张量 $\chi \in \mathbb{R}^{n_1 \times \dots \times n_K}$ 分解为一个核心张量 \mathcal{G} 和 K 个模式矩阵 $\mathbf{U}^{(i)}$ ($i=1, \dots, K$) 的 i -模式积, 如式(1)所示.

$$\chi = \mathcal{G} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_K \mathbf{U}^{(K)} = [\mathcal{G}; \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(K)}] \quad (1)$$

其中运算符 \times_i 表示执行沿第 i -模式的矩阵乘法, 即 i -模式积, $[\cdot; \cdot]$ 表示 K 阶张量分解中 \mathcal{G} 与 $\mathbf{U}^{(i)}$ ($i=1, \dots, K$) 的 1-至 K -模式积. $\mathbf{U}^{(i)}$ ($i=1, \dots, K$) 是一个 $n_i \times j_i$ 的半正交矩阵, 即 $\mathbf{U}^{(i)T} \mathbf{U}^{(i)} = \mathbf{I}_{j_i \times j_i}$, 且 $j_i < n_i$. 核心张量 \mathcal{G} 是一个 $j_1 \times j_2 \times \dots \times j_K$ 张量. 由公式(1)可得:

$$\mathcal{G} = \chi \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \dots \times_K \mathbf{U}^{(K)T} \quad (2)$$

2.2 张量子空间学习

Wu 等人提出张量对齐不变的子空间学习^[14] (Tensor-Aligned Invariant Subspace Learning, TAISL) 方法, 利用张量分解的方法将源域 χ_s 和目标域 χ_t 映射到同一子空间 $\mathcal{U} = \{\mathbf{U}^{(i)}\}_{i=1, \dots, K+1}$. 同时为了使源域和目标域样本线性对齐, 因而优化目标式(3).

$$\begin{aligned} \min_{\mathcal{U}, \mathcal{G}_s, \mathcal{G}_t, \mathcal{M}} & \|\chi_s; \mathcal{M}\|_F^2 + \|\chi_t; \mathcal{M}\|_F^2 + \|\chi_s; \mathcal{G}_s; \mathcal{U}\|_F^2 \\ & + \lambda \|\mathcal{M}\|_F^2 \\ \text{s. t. } & \forall i, \mathbf{U}^{(i)T} \mathbf{U}^{(i)} = \mathbf{I}, \mathbf{M}^{(i)} \mathbf{M}^{(i)T} = \mathbf{I} \end{aligned} \quad (3)$$

其中 λ 是正则化参数, 约束条件表示 $\mathbf{U}^{(i)}$ 是列正交, $\mathbf{M}^{(i)}$ 是行正交.

2.3 动态分布对齐

流形嵌入式分布对齐^[15] (Manifold Embedded Distribution Alignment, MEDA) 中提出动态分布对齐. 动态分布对齐 $\bar{D}_f(\mathbf{X}_s, \mathbf{X}_t)$ 可以定义如公式(4)所示.

$$\bar{D}_f(\mathbf{X}_s, \mathbf{X}_t) = (1 - \mu) D_f(P_s, P_t) + \mu \sum_{b=1}^B D_f^{(b)}(P_s^{(b)}, P_t^{(b)}) \quad (4)$$

其中, $b \in \{1, \dots, B\}$ 表示第 b 个图像类别, $D_f(P_s, P_t)$ 表示边缘概率分布对齐, $D_f^{(b)}(P_s^{(b)}, P_t^{(b)})$ 表示第 b 个类别的条件概率分布对齐, $\mu \in [0, 1]$ 是平衡因子, 调节边缘概率分布和条件概率分布的权重, μ 的计算如式(5)所示.

$$\mu \approx 1 - \frac{d_M}{d_M + \sum_{b=1}^B d_b} \quad (5)$$

其中, d_M 表示样本全局分布距离, 即源域 \mathbf{X}_s 和目标域 \mathbf{X}_t 的边缘概率分布距离, d_b 表示局部分布距离, 即

源域和目标域第 b 类样本的条件概率分布距离.

3 融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法

与本文最为相关的方法有 TAISL^[14] 和 MEDA^[15]. TAISL 方法在张量子空间学习中将源域进行对齐共享子空间. 不同于 TAISL, 本文提出联合域对齐, 将源域和目标域同时对齐共享子空间, 缓解仅单域对齐造成的数据扭曲过大的问题, 同时本文在张量子空间学习中引入了动态分布对齐和图适配. MEDA 方法是基于向量的特征空间, 存在无法自然表达数据重要结构信息的问题. 不同于 MEDA, 本文方法基于张量表示的特征空间, 在张量子空间学习过程中融合了联合域对齐, 动态分布对齐和图适配, 通过联合优化求解最优张量子

空间特征表示. 下面给出本文方法的主要思想和算法实现细节.

3.1 主要思想

设 S 为源域, T 为目标域. 源域的每一个样本 $\chi_s \in \mathbb{R}^{n_1 \times \dots \times n_k}$ ($n = 1, \dots, N_s$) 是 K 阶张量, 总共 N_s 个源域样本堆叠为 $(K+1)$ 阶张量 $\chi_s \in \mathbb{R}^{n_1 \times \dots \times n_k \times n_{k+1}}$, $n_{k+1} = N_s$, 且源域样本有标签, χ_s^n 的标签记为 y_s^n ($n = 1, \dots, N_s$). 同样的, N_t 个样本堆叠为 $(K+1)$ 阶张量 $\chi_t \in \mathbb{R}^{m_1 \times \dots \times m_k \times m_{k+1}}$, $m_{k+1} = N_t$. 本文研究的问题为同构迁移学习, 即 $n_i = m_i$ ($i = 1, 2, \dots, K$). 融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法 (TTL-JDAAR) 的主要思想如图 1 所示.

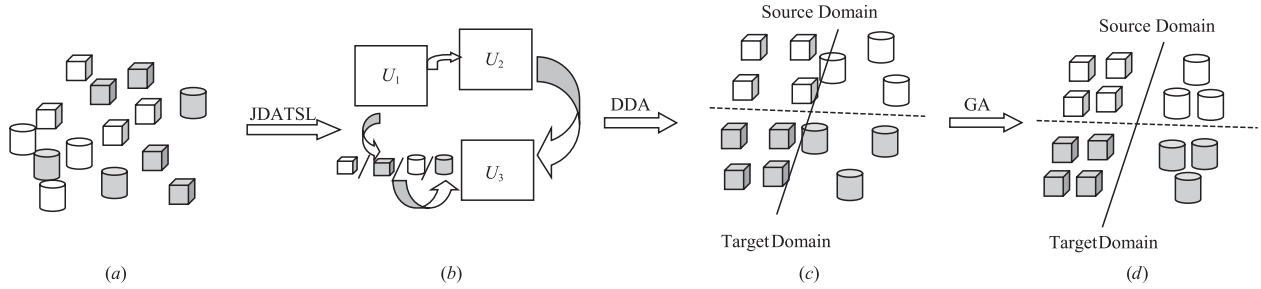


图1 本文方法的主要思想. 正方体和圆柱分别代表两个不同的样本类别, 其中白色填充的为源域样本, 灰色填充的为目标域样本. 实线为样本分类边界, 虚线为源域和目标域的分隔边界. (a) 源域和目标域的初始样本, 图例样本特征用三阶张量进行存储表示. (b) JDATSL (Joint-Domain Alignments based Tensor Subspace Learning): 联合域对齐的张量子空间学习, 源域和目标域同时对齐共享子空间. (c) DDA (Dynamic Distribution Alignment): 张量子空间下的动态分布对齐. (d) GA (Graph Adaptation): 张量子空间下的图适配

TTL-JDAAR 模型如公式 (6) 所示.

$$\min_{\mathcal{Z}} \psi(\chi_s, \chi_t) + \gamma \bar{D}_{\mathcal{Z}}(\chi_s, \chi_t) + \rho R_{\mathcal{Z}}(\chi_s, \chi_t) \quad (6)$$

其中, \mathcal{Z} 是在适配矩阵下重构的源域和目标域, 第一项 $\psi(\chi_s, \chi_t)$ 为联合域对齐的张量子空间学习. 第二项 $\bar{D}_{\mathcal{Z}}(\chi_s, \chi_t)$ 为张量子空间下的动态分布对齐. 第三项 $R_{\mathcal{Z}}(\chi_s, \chi_t)$ 为张量子空间下的图适配正则项. γ 和 ρ 分别为对应的权重参数.

3.2 联合域对齐的张量子空间学习

本文引入两个对齐矩阵集 \mathcal{M}_1 和 \mathcal{M}_2 , 分别将源域和目标域对齐共享子空间. 联合域对齐张量子空间学习的目标函数如式 (7) 所示.

$$\begin{aligned} \psi(\chi_s, \chi_t) = & \min_{U, \mathcal{G}_s, \mathcal{G}_t, \mathcal{M}_1, \mathcal{M}_2} \| [\chi_s; \mathcal{M}_1] - [\mathcal{G}_s; \mathcal{U}] \|_F^2 \\ & + \| [\chi_t; \mathcal{M}_2] - [\mathcal{G}_t; \mathcal{U}] \|_F^2 \\ \text{s. t. } & \forall i, U^{(i)T} U^{(i)} = I, M_1^{(i)} M_1^{(i)T} = I, M_2^{(i)} M_2^{(i)T} = I \end{aligned} \quad (7)$$

其中, 本文只讨论 1 至 K 阶的对齐矩阵集 $\mathcal{M}_1 = \{M_1^{(i)}\}_{i=1, \dots, K}$, $M_1^{(i)} \in \mathbb{R}^{n_i \times n_i}$, $\mathcal{M}_2 = \{M_2^{(i)}\}_{i=1, \dots, K}$, $M_2^{(i)} \in \mathbb{R}^{m_i \times m_i}$. 张量子空间 $\mathcal{U} = \{U^{(i)}\}_{i=1, \dots, K+1}$, $U^{(i)} \in \mathbb{R}^{n_i \times j_i}$. $\mathcal{G}_s \in \mathbb{R}^{j_1 \times \dots \times j_{k+1}}$ 和 $\mathcal{G}_t \in \mathbb{R}^{j_1 \times \dots \times j_{k+1}}$ 是张量子空间下表示的核心张

量, j_i ($i = 1, \dots, K+1$) 表示张量子空间第 i 维的维度. 约束条件中 U 是列正交, \mathcal{M}_1 和 \mathcal{M}_2 是行正交.

3.3 张量子空间下的动态分布对齐

本文提出在张量表示下的动态分布对齐, 在保存数据结构信息的同时, 动态适配源域和目标域边缘分布和条件分布.

将流形正则化原理扩展到张量子空间, 本文重构源域和目标域, 即 $\mathcal{Z}_s = \beta \chi_s = \beta_s [\mathcal{G}_s; \mathcal{U}]$, $\mathcal{Z}_t = \beta_t \chi_t = \beta_t [\mathcal{G}_t; \mathcal{U}]$, 张量子空间下的动态分布对齐如式 (8) 所示.

$$\begin{aligned} \bar{D}_{\mathcal{Z}}(\chi_s, \chi_t) & = (1 - \mu) \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \beta_i [\mathcal{G}_s; \mathcal{U}] - \frac{1}{N_t} \sum_{j=1}^{N_t} \beta_j [\mathcal{G}_t; \mathcal{U}] \right\|_F^2 \\ & + \mu \sum_{b=1}^B \left\| \frac{1}{N_s^{(b)}} \sum_{x_i \in X_s^{(b)}} \beta_i [\mathcal{G}_s; \mathcal{U}] - \frac{1}{N_t^{(b)}} \sum_{x_i \in X_t^{(b)}} \beta_j [\mathcal{G}_t; \mathcal{U}] \right\|_F^2 \\ & = (1 - \mu) \text{tr}(\beta^T [\mathcal{G}; \mathcal{U}] Q_0 [\mathcal{G}; \mathcal{U}]^T \beta) \\ & + \mu \sum_{b=1}^B \text{tr}(\beta^T [\mathcal{G}; \mathcal{U}] Q_b [\mathcal{G}; \mathcal{U}]^T \beta) \\ & = \text{tr}(\beta^T [\mathcal{G}; \mathcal{U}] Q [\mathcal{G}; \mathcal{U}]^T \beta) \end{aligned} \quad (8)$$

$$\text{s. t. } \beta^T \beta = I$$

其中, $\text{tr}(\cdot)$ 表示迹运算, $\boldsymbol{\beta} = [\boldsymbol{\beta}_s, \boldsymbol{\beta}_t]$, $\boldsymbol{\beta} \in \mathbb{R}^{(n_1 \times n_2 \cdots \times n_k) \times (N_s + N_t)}$ 是正交系数矩阵, 是源域迁移至目标域的适配矩阵. $\mathcal{G} = [\mathcal{G}_s; \mathcal{G}_t]$, $\boldsymbol{Q} = (1 - \mu)\boldsymbol{Q}_0 + \mu \sum_{b=1}^B \boldsymbol{Q}_b$, $\boldsymbol{Q} \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ 是 MMD 矩阵, 其中 \boldsymbol{Q}_0 和 \boldsymbol{Q}_b 计算如式(9)和(10)所示.

$$(\boldsymbol{Q}_0)_{ij} = \begin{cases} 1/(N_s N_s), & x_i, x_j \in \mathcal{X}_s \\ 1/(N_t N_t), & x_i, x_j \in \mathcal{X}_t \\ -1/(N_s N_t), & \text{otherwise} \end{cases} \quad (9)$$

$$(\boldsymbol{Q}_b)_{ij} = \begin{cases} 1/(N_s^{(b)} N_s^{(b)}), & x_i, x_j \in \mathcal{X}_s^{(b)} \\ 1/(N_t^{(b)} N_t^{(b)}), & x_i, x_j \in \mathcal{X}_t^{(b)} \\ -1/(N_s^{(b)} N_t^{(b)}), & \begin{cases} x_i \in \mathcal{X}_s^{(b)}, & x_j \in \mathcal{X}_t^{(b)} \\ x_i \in \mathcal{X}_t^{(b)}, & x_j \in \mathcal{X}_s^{(b)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

其中, $N_s^{(b)}$ 是源域中 b 类别的样本个数, 同样, $N_t^{(b)}$ 是目标域中预测的伪标签为 b 类的样本个数.

3.4 张量子空间下的图适配

根据流形原理, 利用近邻点的几何属性, 使样本类内联系更加紧密, 本文增加了张量子空间下的图适配, 重构的源域和目标域 $\mathcal{Z} = \boldsymbol{\beta}\boldsymbol{X} = \boldsymbol{\beta}[\mathcal{G}; \boldsymbol{u}]$, 其中 $\mathcal{G} = [\mathcal{G}_s; \mathcal{G}_t]$, 将 \mathcal{Z} 带入 $R_{\mathcal{Z}}(\mathcal{X}_s, \mathcal{X}_t)$, 得到图适配如式(11)所示.

$$\begin{aligned} R_{\mathcal{Z}}(\mathcal{X}_s, \mathcal{X}_t) &= \sum_{i,j=1}^{N_s+N_t} (\boldsymbol{\beta}_i[\mathcal{G}; \boldsymbol{u}] - \boldsymbol{\beta}_j[\mathcal{G}; \boldsymbol{u}])^2 \boldsymbol{W}_{ij} \\ &= \sum_{i,j=1}^{N_s+N_t} (\boldsymbol{\beta}_i[\mathcal{G}; \boldsymbol{u}]) \boldsymbol{L}_{ij} (\boldsymbol{\beta}_j[\mathcal{G}; \boldsymbol{u}]) \\ &= \text{tr}(\boldsymbol{\beta}^T[\mathcal{G}; \boldsymbol{u}] \boldsymbol{L}[\mathcal{G}; \boldsymbol{u}] \boldsymbol{\beta}) \end{aligned} \quad (11)$$

$$\text{s. t. } \boldsymbol{\beta}^T \boldsymbol{\beta} = \boldsymbol{I}$$

其中, $\boldsymbol{W} \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ 是图关联矩阵, $\boldsymbol{L} \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ 是归一化的图拉普拉斯矩阵. \boldsymbol{W} 矩阵定义如下:

$$\boldsymbol{W}_{ij} = \begin{cases} \cos(x_i, x_j), & \text{if } x_i \in \mathcal{N}_p(x_j) \vee x_j \in \mathcal{N}_p(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

其中, $\mathcal{N}_p(x_i)$ 表示 x_i 的 p 近邻点的集合. $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2} \boldsymbol{W} \boldsymbol{D}^{-1/2}$ 是拉普拉斯矩阵, $\boldsymbol{D} \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ 是对角矩阵, 对角线上的元素表示为 $\boldsymbol{D}_{ii} = \sum_{j=1}^{N_s + N_t} \boldsymbol{W}_{ij}$.

3.5 目标函数优化过程

将式(7), (8), (11)带入公式(6), 得到本文的迁移学习特征表示模型最终的目标函数如式(13)所示.

$$\begin{aligned} \min_{\boldsymbol{u}, \mathcal{G}_s, \mathcal{G}_t, \boldsymbol{\beta}, \mathcal{M}_1, \mathcal{M}_2} & \|\boldsymbol{X}_s; \mathcal{M}_1\| - \|\mathcal{G}_s; \boldsymbol{u}\|_F^2 \\ & + \|\boldsymbol{X}_t; \mathcal{M}_2\| - \|\mathcal{G}_t; \boldsymbol{u}\|_F^2 \end{aligned} \quad (13)$$

$$\begin{aligned} & + \gamma \text{tr}(\boldsymbol{\beta}^T[\mathcal{G}; \boldsymbol{u}] \boldsymbol{Q}[\mathcal{G}; \boldsymbol{u}] \boldsymbol{\beta}) \\ & + \rho \text{tr}(\boldsymbol{\beta}^T[\mathcal{G}; \boldsymbol{u}] \boldsymbol{L}[\mathcal{G}; \boldsymbol{u}] \boldsymbol{\beta}) + \eta \|\boldsymbol{\beta}\|_F^2 \end{aligned}$$

$$\text{s. t. } \forall i, \boldsymbol{U}^{(i)T} \boldsymbol{U}^{(i)} = \boldsymbol{I}, \boldsymbol{M}_1^{(i)} \boldsymbol{M}_1^{(i)T} = \boldsymbol{I}, \boldsymbol{M}_2^{(i)} \boldsymbol{M}_2^{(i)T} = \boldsymbol{I}, \boldsymbol{\beta}^T \boldsymbol{\beta} = \boldsymbol{I}$$

其中, $\mathcal{G} = [\mathcal{G}_s; \mathcal{G}_t]$, γ, ρ, η 是权重参数. 目标函数的第一项和第二项是联合域对齐的张量子空间学习, 可以保存数据的结构信息, 得到优化的张量子空间. 第三项是张量子空间下的动态分布对齐, 动态适配源域和目标域的边缘分布和条件分布, 减小源域和目标域之间的差异. 第四项是张量子空间下的图适配, 利用近邻点的几何属性, 减小样本的类内距离, 使类内联系更加紧密. 第五项是 $\boldsymbol{\beta}$ 的 Frobenius 范数 $\|\boldsymbol{\beta}\|_F^2$, 获得问题更稳定的结果. 公式(13)是一个非凸函数, 本文采用变量交替迭代策略进行优化. 具体步骤如下.

(1) 固定 $\mathcal{M}_1, \mathcal{M}_2, \boldsymbol{\beta}$, 更新 $\boldsymbol{u}, \mathcal{G}_s$ 和 \mathcal{G}_t

通过式(1), $\boldsymbol{u}, \mathcal{G}_s$ 和 \mathcal{G}_t 是同时通过张量 Tucker 分解得到的, 所以将它们同时优化. 引入两个辅助变量 $\mathcal{A}_s = [\boldsymbol{X}_s; \mathcal{M}_1]$, $\mathcal{A}_t = [\boldsymbol{X}_t; \mathcal{M}_2]$, 优化问题(13)可以转换为式(14).

$$\begin{aligned} \min_{\boldsymbol{u}, \mathcal{G}_s, \mathcal{G}_t} & \|\mathcal{A}_s - [\mathcal{G}_s; \boldsymbol{u}]\|_F^2 + \|\mathcal{A}_t - [\mathcal{G}_t; \boldsymbol{u}]\|_F^2 \\ & + \gamma \text{tr}(\boldsymbol{\beta}^T[\mathcal{G}; \boldsymbol{u}] \boldsymbol{Q}[\mathcal{G}; \boldsymbol{u}] \boldsymbol{\beta}) + \rho \text{tr}(\boldsymbol{\beta}^T[\mathcal{G}; \boldsymbol{u}] \boldsymbol{L}[\mathcal{G}; \boldsymbol{u}] \boldsymbol{\beta}) \end{aligned} \quad (14)$$

$$\text{s. t. } \forall i, \boldsymbol{U}^{(i)T} \boldsymbol{U}^{(i)} = \boldsymbol{I}, \boldsymbol{\beta}^T \boldsymbol{\beta} = \boldsymbol{I}$$

(2) 固定 $\boldsymbol{u}, \mathcal{G}_s, \mathcal{G}_t, \boldsymbol{\beta}, \mathcal{M}_2$, 更新 \mathcal{M}_1

引入辅助变量 $\boldsymbol{V}_1 = [\mathcal{G}_s; \boldsymbol{u}]$, 得到 \mathcal{M}_1 的优化如式(15)所示.

$$\min_{\mathcal{M}_1} \|\boldsymbol{X}_s; \mathcal{M}_1\| - \boldsymbol{V}_1\|_F^2 \quad (15)$$

$$\text{s. t. } \forall i, \boldsymbol{M}_1^{(i)} \boldsymbol{M}_1^{(i)T} = \boldsymbol{I}$$

将式(15)按照模式- i 展开, 得到公式(16)

$$\|\{\boldsymbol{X}_s; \mathcal{M}_1\} - \boldsymbol{V}_1\}_{(i)}\|_F^2 = \|\boldsymbol{M}_1^{(i)} \boldsymbol{X}_s^{(i)} \boldsymbol{M}_1^{(i)T} - \boldsymbol{V}_1^{(i)}\|_F^2 \quad (16)$$

其中 $\boldsymbol{M}_1^{(i)T} = (\boldsymbol{I} \otimes \boldsymbol{M}_1^{(k)} \otimes \cdots \otimes \boldsymbol{M}_1^{(i+1)} \otimes \boldsymbol{M}_1^{(i-1)} \otimes \cdots \otimes \boldsymbol{M}_1^{(1)})^T$, \otimes 表示克罗克内积. 令 $\boldsymbol{J}_s^{(i)} = \boldsymbol{X}_s^{(i)} \boldsymbol{M}_1^{(i)T}$, 因为 \mathcal{M}_1 和 \mathcal{M}_2 是行正交, 所以为了让优化问题在标准的正交约束中, 即满足列正交, 令 $\boldsymbol{P}_1^{(i)} = \boldsymbol{M}_1^{(i)T}$, 带入式(16)运算如下:

$$\begin{aligned} \|\{\boldsymbol{X}_s; \mathcal{M}_1\} - \boldsymbol{V}_1\}_{(i)}\|_F^2 &= \|\boldsymbol{M}_1^{(i)} \boldsymbol{X}_s^{(i)} \boldsymbol{M}_1^{(i)T} - \boldsymbol{V}_1^{(i)}\|_F^2 \\ &= \|\boldsymbol{M}_1^{(i)} \boldsymbol{X}_s^{(i)} \boldsymbol{M}_1^{(i)T} - \boldsymbol{V}_1^{(i)T}\|_F^2 \\ &= \|\boldsymbol{J}_s^{(i)T} \boldsymbol{P}_1^{(i)} - \boldsymbol{V}_1^{(i)T}\|_F^2 \end{aligned} \quad (17)$$

最终 \mathcal{M}_1 的优化问题转换为 \boldsymbol{P}_1 的优化问题, 如式(18)所示.

$$\begin{aligned} \min_{\boldsymbol{P}_1} & \|\boldsymbol{J}_s^{(i)T} \boldsymbol{P}_1^{(i)} - \boldsymbol{V}_1^{(i)T}\|_F^2 \\ & = \text{tr}(\boldsymbol{P}_1^{(i)T} \boldsymbol{J}_s^{(i)} \boldsymbol{J}_s^{(i)T} \boldsymbol{P}_1^{(i)}) - 2\text{tr}(\boldsymbol{P}_1^{(i)T} \boldsymbol{J}_s^{(i)} \boldsymbol{V}_1^{(i)T}) + \text{tr}(\boldsymbol{V}_1^{(i)} \boldsymbol{V}_1^{(i)T}) \end{aligned} \quad (18)$$

$$\text{s. t. } \forall i, \mathbf{P}_1^{(i)\text{T}} \mathbf{P}_1^{(i)} = \mathbf{I}$$

(3) 固定 $\mathcal{U}, \mathcal{G}_s, \mathcal{G}_t, \beta, \mathcal{M}_1$, 更新 \mathcal{M}_2

引入辅助变量 $\mathcal{V}_2 = [\mathcal{G}_t; \mathcal{U}]$, \mathcal{M}_2 的更新和 \mathcal{M}_1 同理, 所以令 $\mathbf{J}_t^{(i)} = \mathbf{X}_t^{(i)} \mathbf{M}_2^{(i)\text{T}}, \mathbf{P}_2^{(i)} = \mathbf{M}_2^{(i)\text{T}}$, \mathcal{M}_2 的优化问题转换为 \mathbf{P}_2 的优化问题, 如式(19)所示.

$$\begin{aligned} \min_{\mathbf{P}_2} & \| \mathbf{J}_t^{(i)\text{T}} \mathbf{P}_2^{(i)} - \mathbf{V}_2^{(i)\text{T}} \|_F^2 \\ & = \text{tr}(\mathbf{P}_2^{(i)\text{T}} \mathbf{J}_t^{(i)} \mathbf{J}_t^{(i)\text{T}} \mathbf{P}_2^{(i)}) - 2\text{tr}(\mathbf{P}_2^{(i)\text{T}} \mathbf{J}_t^{(i)} \mathbf{V}_2^{(i)\text{T}}) + \text{tr}(\mathbf{V}_2^{(i)} \mathbf{V}_2^{(i)\text{T}}) \end{aligned} \quad (19)$$

$$\text{s. t. } \forall i, \mathbf{P}_2^{(i)\text{T}} \mathbf{P}_2^{(i)} = \mathbf{I}$$

(4) 固定 $\mathcal{U}, \mathcal{G}_s, \mathcal{G}_t, \mathcal{M}_1, \mathcal{M}_2$, 更新 β

引入辅助变量 $\mathcal{A} = [\mathcal{G}; \mathcal{U}]$, 将 \mathcal{A} 按照模式 $-(K+1)$ 水平展开为矩阵 $\mathbf{A}^{(K+1)} \in \mathbb{R}^{(n_1 \times n_2 \times \dots \times n_k) \times (N_s + N_t)}$, 然后采用特征分解的方法优化 β , 从而在目标函数中 β 的优化如式(20)所示.

$$\min_{\beta} \gamma \text{tr}(\beta^{\text{T}} \mathbf{A} \mathbf{Q} \mathbf{A}^{\text{T}} \beta) + \rho \text{tr}(\beta^{\text{T}} \mathbf{A} \mathbf{L} \mathbf{A}^{\text{T}} \beta) + \eta \|\beta\|_F^2 \quad (20)$$

$$\text{s. t. } \beta^{\text{T}} \beta = \mathbf{I}$$

式(20)可以用拉格朗日对偶法来求解, 其拉格朗日函数如式(21)所示.

$$l = \text{tr}(\beta^{\text{T}} (\mathbf{A} (\gamma \mathbf{Q} + \rho \mathbf{L}) \mathbf{A}^{\text{T}} + \eta \mathbf{I}) \beta) + \text{tr}((\mathbf{I} - \beta^{\text{T}} \beta) \xi) \quad (21)$$

其中, $\xi = \text{diag}(\xi_1, \dots, \xi_N) \in \mathbb{R}^{N \times N}$ ($N = N_s + N_t$) 为拉格朗日乘子, 其对角线上的元素为对应的特征值, 令 $\partial l / \partial \beta = 0$, 得到广义的特征分解如式(22)所示.

$$(\mathbf{A} (\gamma \mathbf{Q} + \rho \mathbf{L}) \mathbf{A}^{\text{T}} + \eta \mathbf{I}) \beta = \beta \xi \quad (22)$$

最终, 寻找最优的适配矩阵 β 的问题转化为求公式(22)的 N 个最小特征向量的问题.

通过张量子空间 \mathcal{U} , 核心张量 $\mathcal{G}_s, \mathcal{G}_t$, 对齐矩阵 $\mathcal{M}_1, \mathcal{M}_2$, 适配矩阵 $\beta = [\beta_s, \beta_t]$ 的优化更新, 得到张量子空间下源域和目标域的特征表示 χ_s^* 和 χ_t^* 如式(23)和式(24)所示.

$$\begin{aligned} \chi_s^* & = \beta_s [[\mathcal{G}_s; \mathcal{U}]; \mathcal{M}_1] = \mathcal{G}_s \times_1 (\mathbf{M}_1^{(1)} \mathbf{U}^{(1)}) \\ & \quad \times_2 \dots \times_k (\mathbf{M}_1^{(k)} \mathbf{U}^{(k)}) \times_{k+1} (\beta_s \mathbf{U}^{(k+1)}) \end{aligned} \quad (23)$$

$$\begin{aligned} \chi_t^* & = \beta_t [[\mathcal{G}_t; \mathcal{U}]; \mathcal{M}_2] = \mathcal{G}_t \times_1 (\mathbf{M}_2^{(1)} \mathbf{U}^{(1)}) \\ & \quad \times_2 \dots \times_k (\mathbf{M}_2^{(k)} \mathbf{U}^{(k)}) \times_{k+1} (\beta_t \mathbf{U}^{(k+1)}) \end{aligned} \quad (24)$$

3.6 算法流程

融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法流程如算法 1 所示

算法 1 融合联合域对齐和适配正则化的基于张量表示的迁移学习特征表示方法 (TTL-JDAAR)

输入 源域张量数据 χ_s , 源域标签 y_s , 目标域张量数据 χ_t , 张量子空间维度 $j_k, k=1, \dots, K+1$; 正则化参数 γ, ρ, η , 最大迭代次数 T

输出 张量子空间下源域和目标域的特征表示 χ_s^* 和 χ_t^*

1: 初始化 $\mathbf{M}_1^{(i)} = \mathbf{M}_2^{(i)} = \mathbf{I}, i=1, \dots, K$

2: FOR $t=1$ to T

3: 根据式(5)计算平衡参数 μ

4: 根据式(9)计算 \mathbf{Q}_0

5: 根据式(10)计算 \mathbf{Q}_i

6: 根据式(12)构建拉普拉斯矩阵 \mathbf{L}

7: 根据式(14)更新 $\{\mathcal{U}, \mathcal{G}_s, \mathcal{G}_t\}$

8: FOR $i=1$ to K

9: 根据式(18)和式(19)更新优化 $\mathbf{M}_1^{(i)}$ 和 $\mathbf{M}_2^{(i)}$

10: END FOR

11: 根据式(22)计算更新 β

12: 检查收敛性

13: END FOR

14: 通过式(23)和(24)计算张量子空间下源域和目标域的特征表示 χ_s^* 和 χ_t^*

15: RETURN 张量子空间下源域和目标域的特征表示 χ_s^* 和 χ_t^*

4 实验

4.1 数据集

本文选择几个通用数据集作为实验数据集, 具体包括 Office-Caltech10 数据集, Office-Home 数据集, 数字数据集 (MNIST, USPS, MNIST-M).

Office-Caltech10 数据集^[16] 包含四个子集: Amazon, DSLR, Web-cam 和 Caltech. 数据集由 10 个类别, 共计 2500 张图像组成. 实验分别任选其中两个数据子集作为源域和目标域, 排列组合产生 12 组分类迁移学习任务, 例如, A→D 表示 Amazon 作为源域, DSLR 作为目标域的迁移学习任务.

Office-Home 数据集^[17] 有 65 个类别, 共计 15500 张图像, 分别来自四个不同的数据子集 Artistic paintings, Clipart, Product, Real-World images. 实验分别任选其中两个数据子集作为源域和目标域, 排列组合产生 12 组分类迁移学习任务, 例如, Ar→Cl 表示 Artistic paintings 作为源域, Clipart 作为目标域的迁移学习任务.

数字数据集包括 MNIST, USPS 和 MNIST-M. MNIST 手写数字数据集有 10 个类别的样本, 分别为数字 0 至 9, 数据集包含 60000 张训练图像和 10000 张测试图像, 每张的图像的大小为 28×28 . USPS 是美国邮政手写数字数据集. 同样的 10 类数字样本, 包含 7291 张训练图像和 2007 张测试图像, 每张图像的大小为 16×16 . MNIST-M 数据集是 MNIST 数据集的一个变体, 将从 Berkeley Segmentation 数据集^[18] (BSDS500) 随机提取的图片替换掉 MNIST 的黑色背景. 因此训练和测试的样本大小和 MNIST 数据集相等. 实验分别任选其中两个数字数据集作为源域和目标域, 排列组合产生 6 组分类迁移学习任务, 例如, MNIST→MNIST-M 表示 MNIST 作为源域, MNIST-M 作为目标域的迁移学习任务.

4.2 实验设置

为了验证所提出的 TTL-JDAAR 方法的有效性, 本

文分别和一种非迁移学习方法 No Adaptation (NA) 和八种主流迁移学习特征表示方法进行了对比实验,其中八种主流迁移学习特征表示方法包括六种基于向量表示的迁移学习方法 (Transfer Component Analysis (TCA)^[2], Joint distribution alignment (JDA)^[3], Geodesic Flow Kernel (GFK)^[16], Correlation Alignment (CORAL)^[19], Joint Geometrical and Statistical Alignment (JGSA)^[20], Manifold Embedded Distribution Alignment (MEDA)^[15]) 和两种基于张量表示的迁移学习方法 (Naive tensor subspace learning (NTSL)^[14] 和 Tensor-aligned invariant subspace learning (TAISL)^[14]). 实验以迁移学习任务在目标域上的分类准确率 Accuracy 作为算法性能

的评价标准.

本文的实验参数统一设置为 $T = 10$, $\gamma = 10$, $\rho = 1$, $\eta = 1e^{-1}$. 对于基于张量表示的 NTSL, TAISL 和本文所提算法, 统一提取 vgg16 模型的 pool5 层特征作为数据的张量表示, 每一个样本图像都可以映射为 $7 \times 7 \times 512$ 的 3 阶张量, 第四阶对应样本数量. 设置张量子空间的维度 $j_1 = j_2 = 7$, $j_3 = 128$, $j_4 = 20$. 对比基于向量表示的其它方法采取同样的网络和输出层的特征.

4.3 实验结果和分析

本文在三组数据集上的实验结果分别如表 1, 表 2 和表 3 所示.

表 1 Office-Caltech10 数据集上的 12 个迁移任务的分类准确率 (%)

Task	NA	TCA	JDA	GFK	CORAL	JGSA	MEDA	NTSL	TAISL	TTL-JDAAR
A→C	77.3	84.2	82.3	84.7	88.7	86.7	87.8	78.5	80.3	89.5
C→A	88.8	91.0	87.9	91.4	93.1	93.1	94.0	89.5	89.8	94.9
A→D	82.6	84.1	80.7	84.7	81.5	86.6	89.6	83.7	85.9	94.9
D→A	80.9	83.3	89.2	83.0	83.5	92.3	93.6	87.9	87.9	94.5
A→W	73.6	83.1	90.7	84.1	81.7	94.2	97.2	76.5	77.6	88.1
W→A	73.8	88.5	88.8	86.3	83.8	94.9	95.1	85.7	86.5	94.8
C→D	86.4	79.6	81.6	85.4	87.3	90.5	91.3	86.8	90.3	98.7
D→C	70.4	76.1	80.7	78.3	79.9	85.9	86.1	79.8	83.8	90.3
C→W	79.4	80.3	86.1	82.4	88.8	90.2	96.0	79.7	85.7	96.5
W→C	63.1	82.2	81.7	82.5	79.5	86.5	88.5	79.6	82.7	90.4
D→W	91.1	96.6	97.1	98.0	96.3	97.6	97.3	95.9	95.9	97.6
W→D	94.8	96.8	97.5	99.4	98.7	98.1	98.7	97.7	97.7	99.4
MEAN	80.2	85.5	87.0	86.7	86.9	91.4	93.0	85.1	87.0	94.1

表 2 Office-Home 数据集上的 12 个迁移任务的分类准确率 (%)

Task	NA	TCA	JDA	GFK	CORAL	JGSA	MEDA	NTSL	TAISL	TTL-JDAAR
Cl→Ar	63.7	60.8	61.9	56.6	59.4	72.0	74.8	54.0	58.1	81.9
Ar→Cl	54.9	49.7	51.6	53.4	56.1	59.7	60.8	55.7	57.2	77.8
Pr→Re	91.2	89.5	92.0	88.8	88.7	94.5	96.6	85.9	88.3	97.5
Re→Pr	82.4	87.0	88.2	87.6	88.2	91.0	90.8	82.2	82.5	92.4
Ar→Pr	74.0	79.1	75.5	79.1	81.2	87.3	89.7	73.8	76.5	87.9
Pr→Ar	65.9	63.0	62.2	61.1	66.1	74.5	76.2	60.3	62.6	81.9
Ar→Re	91.6	85.6	82.7	87.7	89.6	90.7	93.5	91.0	88.5	96.0
Re→Ar	76.4	73.0	66.4	73.7	76.4	79.1	79.1	69.8	70.2	86.3
Re→Cl	61.5	56.2	51.7	58.2	60.1	63.6	74.1	57.7	61.5	74.0
Cl→Re	81.5	83.5	78.2	77.4	74.3	88.5	94.5	79.7	81.3	95.3
Cl→Pr	72.9	69.1	69.2	66.5	61.9	82.7	88.5	70.8	73.7	91.9
Pr→Cl	53.8	49.9	51.7	55.2	51.4	56.4	57.5	53.6	57.1	73.2
MEAN	72.5	70.5	69.3	70.4	71.1	78.3	81.3	69.5	71.5	86.3

表 3 数字数据集上的 6 个迁移任务的分类准确率 (%)

Task	NA	TCA	JDA	GFK	CORAL	JGSA	MEDA	NTSL	TAISL	TTL-JDAAR
MNIST→MNISTM	52.9	20.1	15.4	21.1	30.1	18.1	17.0	52.5	53.2	56.9
MNISTM→MNIST	89.3	28.1	20.7	32.3	60.9	45.8	56.9	89.2	91.6	92.3
MNIST→USPS	58.3	38.2	35.6	43.9	49.5	44.9	50.6	52.9	55.8	87.6
USPS→MNIST	63.9	31.8	34.5	38.4	45.4	46.1	53.3	59.9	65.9	89.2
MNISTM→USPS	40.9	29.3	17.2	27.0	27.2	32.5	41.8	35.5	36.8	82.5
USPS→MNISTM	22.0	13.4	13.1	14.0	17.6	10.9	10.7	19.2	24.5	32.8
MEAN	54.6	26.8	22.8	29.5	38.5	33.1	38.4	51.6	54.6	73.6

由表 1, 表 2, 表 3 可以看出, 本文提出的方法在 Office-Caltech10, OfficeHome, 数字数据集上的平均分类准确率有很大提高. 分析以上实验结果, 由于 NA 没有采用迁移学习方法, 直接使用将源域训练得到的分类器应用到目标域. TCA 通过最小化 MMD, 仅对齐边缘概率分布. JDA 对齐边缘概率分布和条件概率分布, 但没有考虑到边缘概率分布和条件概率分布在不同数据集上权重的差异. GFK 在 Grassmann 流形中使用测地线, 桥接两个域的子空间使源域更靠近目标域, 但是没有考虑样本的分布对齐. CORAL 使用协方差矩阵表示域的特征, 是一种二阶子空间对齐的方法, 但没有考虑样本的分布对齐和近邻点的几何分布. JGSA 在 JDA 的基础上增加了类内距离、类间距离和标签持久化, 减小几何和分布的漂移, 但是没有考虑图适配下的流形一致性. MEDA 是在 Grassmann 流形中通过结构风险最小化学习一个域不变分类器, 同时动态适配边缘概率分布和条件概率分布, 但是此方法是基于向量化的特征空间. 以上方法都是基于向量表示的方法, 存在无法自然表达数据重要结构信息的问题. NTSL 是通过张量分解进行朴素的张量子空间学习的方法, TAISL 是 NTSL 方法的扩展, 添加了对齐矩阵和正则项. NTSL 和 TAISL 是基于张量表示方法, 但是没有将目标域对齐到共享子空间, 也没有考虑样本的动态分布对齐和保持流行一致性. 实验结果显示, 本文提出的 TTL-JDAAR 基于张量表示, 融合联合域对齐, 动态分布对齐和图适配, 学习到的迁移学习特征表示取得了较高的分类准确率.

表 5 参数 ρ 在各数据集上的迁移学习任务的平均分类准确率 (%)

Dataset	0	0.1	1	10	20	30	40	1e+2	1e+3
Office-Caltech10	92.6	93.2	94.1	93.5	92.4	91.6	90.2	87.0	57.3
OfficeHome	84.3	85.2	86.3	84.0	79.9	77.2	75.6	64.4	32.5
数字数据集	67.6	73.1	73.6	69.5	55.2	44.7	41.5	25.4	13.9

对权值参数 η , 分别固定其它参数值 $T = 10, \gamma = 10, \rho = 1$, 调节 η 参数值, 各数据集上的平均分类准确率如表 6 所示. η 的大小决定了防止求解问题的过拟合

4.4 参数分析

为了检验本文提出的 TTL-JDAAR 的各参数对方法表示性能的影响, 分别对动态分布对齐参数 γ , 图适配正则参数 ρ , 适配矩阵 β 的 F 范数的权衡参数 η , 迭代次数 T , 在 3 组数据集, 30 个分类迁移学习任务进行对比实验分析.

首先对权值参数 γ , 分别固定其它参数值 $T = 10, \rho = 1, \eta = 1e^{-1}$, 调节 γ 参数值, 各数据集上的平均分类准确率如表 4 所示. γ 的大小决定了新特征空间下缩小源域和目标域分布差异在最小化目标函数中所占的权重. 从表 4 可以看出, 当 $\gamma \rightarrow 0$ 时, 即几乎忽略不同域的动态分布对齐, 和 $\gamma \rightarrow +\infty$ 时, 即几乎忽略不同域的联合域对齐和图适配, 都会导致分类准确率下降.

表 4 参数 γ 在各数据集上的迁移学习任务的平均分类准确率 (%)

Dataset	0	1	10	1e+2	1e+3	1e+4	1e+5
Office-Caltech10	91.8	93.7	94.1	93.9	93.9	92.0	92.0
OfficeHome	85.0	86.2	86.3	85.8	85.3	85	84.7
数字数据集	54.2	73.3	73.6	72.6	71.5	70.5	69.6

对权值参数 ρ , 分别固定其它参数值 $T = 10, \gamma = 10, \eta = 1e^{-1}$, 调节 ρ 参数值, 各数据集上的平均分类准确率如表 5 所示. ρ 的大小决定了新特征空间下数据集原几何特性的保留在最小化目标函数中所占的权重. 从表 5 可以看出, 当 $\rho \rightarrow 0$ 时, 即几乎忽略数据集的几何特性, 和 $\rho \rightarrow +\infty$ 时, 即几乎忽略不同域间联合域对齐和动态分布对齐, 都会导致分类准确率下降.

的能力, 获得求解问题稳定的解. 从表 6 可以看出, 当 $\eta \rightarrow 0$ 时, 即防止过拟合的能力越小, 和 $\eta \rightarrow +\infty$ 时, 即几乎忽略其它项的影响, 都会导致分类准确率下降.

表 6 参数 η 在各数据集上的迁移学习任务的平均分类准确率 (%)

Dataset	0	1e-10	1e-2	1e-1	1	1e+1	1e+2	1e+10
Office-Caltech10	87.4	93.7	93.8	94.1	93.9	91.5	89.6	89.0
OfficeHome	64.8	86.2	86.2	86.3	85.8	75.9	50.4	27.3
数字数据集	63.3	70.3	71.6	73.6	72.8	69.8	49.4	19.5

对权值参数 T , 分别固定其它参数值 $\gamma = 10, \rho = 1, \eta = 1e^{-1}$, 调节 T 参数值, 各数据集上的平均分类准确率如表 7 所示. 从表 7 可以看出, 本文算法在 Office-

Caltech10 数据集下 3 次可达到收敛, 在 OfficeHome 数据集下 6 次可达收敛, 在数字数据集下 8 次可达收敛.

表 7 不同迭代次数 T 下各数据集迁移学习任务的平均分类准确率 (%)

Dataset	1	2	3	4	5	6	7	8	9	10
Office-Caltech10	93.7	94.0	94.1	94.1	94.1	94.1	94.1	94.1	94.1	94.1
OfficeHome	84.9	85.8	86.2	86.2	86.2	86.3	86.3	86.3	86.3	86.3
数字数据集	60.2	62.6	63.0	63.2	63.4	63.4	63.6	73.6	73.6	73.6

4.5 张量子空间维度分析

为了检验张量子空间的维度对方法表示性能的影响, 本文在 OfficeHome 数据集上的 6 个分类迁移学习任务上, 分别对张量子空间的维度 j_1, j_2, j_3, j_4 进行实验分析, 且 $j_i \leq n_i$, 其中 n_i 是原始样本第 i 阶的大小.

首先检验张量子空间的维度 j_1, j_2 对性能的影响. 在实验中设置 $j_1 = j_2, j_3 = 128, j_4 = 20$, (j_1, j_2) 分别取值为 $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (7, 7)\}$, 实验结果如表 8 所示. 从表 8 可以看出, 当 $(j_1 = j_2) \rightarrow 7$ 时, 即张量子空间 $U^{(1)}$ 和 $U^{(2)}$ 第二维的维数越接近特征图的宽和高的值 7, 分类准确率会上升.

为了检验维度 j_3 对性能的影响, 设置 $j_1 = j_2 = 7, j_4 = 20, j_3$ 分别取值为 2, 4, 8, 16, 32, 64, 128, 256, 512, 实验结果如表 9 所示. 因为每个图像样本提取的特征是 $7 \times 7 \times 512$ 的大小, 512 对应的是提取特征的通道数, 即

表 9 j_3 不同取值下的分类准确率 (%)

j_3	2	4	8	16	32	64	128	256	512
Cl→Ar	37.9	61.0	69.3	78.1	81.1	81.5	81.9	81.5	80.6
Ar→Cl	33.0	44.8	56.0	67.2	69.6	75.8	77.8	77.0	75.4
Pr→Re	56.2	74.2	85.5	96.5	97.1	97.9	97.5	97.5	97.5
Re→Pr	62.7	74.9	83.2	89.3	90.2	90.8	92.4	91.3	90.8
Ar→Pr	29.4	49.4	73.1	84.1	87.8	87.8	87.9	87.2	87.4
Pr→Ar	32.6	57.3	69.1	75.0	79.4	82.3	81.9	81.9	79.4

为了检验维度 j_4 对性能的影响, 设置 $j_1 = j_2 = 7, j_3 = 128, j_4$ 分别取值为 10, 20, 30, 40, 50, 60, 70, 80, 90, 实验结果如表 10 所示. 因为源域和目标域第四维 n_4 和 m_4 表示的是图像的样本数量, 所以 $j_4 \leq \min(n_4, m_4)$. 从

每个样本提取的特征图数量, 所以张量子空间第三维 $j_3 < 512$. 从表 9 可以看出, 当 $j_3 \rightarrow 2$ 时, 即 $U^{(3)} \in \mathbb{R}^{n_3 \times j_3}$ 的第二维 j_3 越小, \mathcal{G}_s 和 \mathcal{G}_t 的第三维越小, 当 $j_3 \rightarrow 512$ 时, 即 $U^{(3)}$ 的第二维 j_3 越大, \mathcal{G}_s 和 \mathcal{G}_t 的第三维越大, 都会导致分类准确率的下降.

表 8 j_1, j_2 不同取值下的分类准确率 (%)

j_1, j_2	1,1	2,2	3,3	4,4	5,5	6,6	7,7
Cl→Ar	81.0	81.5	81.5	81.6	81.8	81.8	81.9
Ar→Cl	68.5	72.9	72	75.6	76.9	77.8	77.8
Pr→Re	97.1	97.1	97.3	97.5	97.5	97.5	97.5
Re→Pr	91.0	91.5	91.8	92.0	92.0	92.2	92.4
Ar→Pr	85.0	86.1	87.4	87.8	87.8	87.9	87.9
Pr→Ar	80.7	81.1	81.4	81.7	81.8	81.7	81.9

表 10 可以看出, 当 $j_4 \rightarrow 10$ 时, 即 $U^{(4)} \in \mathbb{R}^{n_4 \times j_4}$ 的第二维 j_4 越小, \mathcal{G}_s 和 \mathcal{G}_t 的第四维越小, 当 $j_4 \rightarrow 90$ 时, 即 $U^{(4)}$ 的第二维 j_4 越大, \mathcal{G}_s 和 \mathcal{G}_t 的第四维越大, 都会导致分类准确率的下降.

表 10 j_4 不同取值下的分类准确率 (%)

j_4	10	20	30	40	50	60	70	80	90
Cl→Ar	77.7	81.9	81.3	80.9	80.0	80.0	79.9	78.9	78.3
Ar→Cl	71.6	77.8	74.1	71.6	71.4	73.8	74.6	73.6	73.4
Pr→Re	96.6	97.5	97.2	97.2	97.1	96.9	97.5	97.2	97.2
Re→Pr	89.8	92.4	92.0	91.1	91.0	91.4	91.0	91.4	91.2
Ar→Pr	76.8	87.9	87.1	86.5	81.0	80.5	80.4	81.6	81.5
Pr→Ar	74.3	81.5	81.0	80.0	81.3	80.1	80.1	79.9	79.5

5 总结

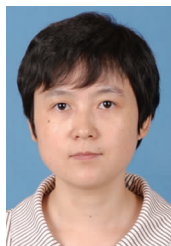
张量可以自然地表示向量表示不易表达的信息. 本文提出了一种融合联合域对齐张量表示和域适配的迁移学习特征表示方法. 该方法利用联合域对齐张量子空间学习, 将源域和目标域同时对齐潜在共享子空间, 缓解单域对齐易造成数据扭曲过大的问题. 在张量子空间学习中, 利用动态分布对齐, 缩小源域和目标域边缘概率分布和条件概率分布. 并且在张量子空间学习中引入图适配, 保存数据的流行一致性. 几个公共数据集 (Office-Caltech10、Office-Home 数据集和数字数据集) 上的大量实验验证了本文方法的有效性.

参考文献

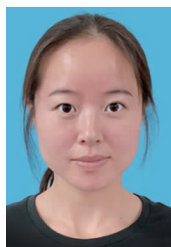
- [1] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis [J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199 – 210.
- [2] Pan S J, Kwok J T, Yang Q. Transfer learning via dimensionality reduction [A]. Association for the Advance of Artificial Intelligence [C]. Menlo Park, CA: AAAI, 2008. 677 – 682.
- [3] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation [A]. IEEE International Conference on Computer Vision [C]. Washington: IEEE, 2013. 2200 – 2207.
- [4] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning [A]. IEEE International Conference on Data Mining [C]. Washington: IEEE, 2017. 1129 – 1134.
- [5] Zhang L, Liu Y, He Z, et al. Anti-drift in E-nose: A subspace projection approach with drift reduction [J]. Sensors and Actuators B: Chemical, 2017, 253: 407 – 417.
- [6] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning [A]. 2006 Conference on Empirical Methods in Natural Language Processing [C]. Sydney: Association for Computational Linguistics, 2006. 120 – 128.
- [7] Satpal S, Sarawagi S. Domain adaptation of conditional probability models via feature subsetting [A]. European Conference on Principles of Data Mining and Knowledge Discovery [C]. Berlin Heidelberg: Springer-Verlag, 2007. 224 – 235.
- [8] Fernando B, Habrard A, Sebban M, et al. Unsupervised visual domain adaptation using subspace alignment [A]. IEEE Conference on Computer Vision [C]. Washington: IEEE, 2013. 2960 – 2967.
- [9] Ghifary M, Balduzzi D, Kleijn W B, et al. Scatter component analysis: A unified framework for domain adaptation and domain generalization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1414 – 1430.
- [10] Ben X, Gong C, Zhang P, et al. Coupled bilinear discriminant projection for cross-view gait recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, (99): 1 – 1.
- [11] Shashua A, Levin A. Linear image coding for regression and classification using the tensor-Rank principle [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2001. 42 – 49.
- [12] Ben X, Zhang P, Lai Z, et al. A general tensor representation framework for cross-view gait recognition [J]. Pattern Recognition, 2019, 90: 87 – 98.
- [13] Kolda T, Bader B. The TOPHITS model for higher-order web link analysis [A]. Workshop on Link Analysis, Counterterrorism and Security [C]. Bethesda: Link Analysis, Counterterrorism and Security, 2006. 7: 26 – 29.
- [14] Lu H, Zhang L, Cao Z, et al. When unsupervised domain adaptation meets tensor representations [A]. IEEE International Conference on Computer Vision [C]. Washington: IEEE, 2017. 599 – 608.
- [15] Wang J, Feng W, Chen Y, et al. Visual domain adaptation with manifold embedded distribution alignment [A]. ACM Multimedia Conference on Multimedia Conference [C]. New York: ACM, 2018. 402 – 410.
- [16] Gong B, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Washing-

- ton: IEEE, 2012. 2066 – 2073.
- [17] Venkateswara H, Eusebio J, Chakraborty S, et al. Deep hashing network for unsupervised domain adaptation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2017. 5018 – 5027.
- [18] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011. 33(5): 898 – 916.
- [19] Sun B, Feng J, Saenko K. Return of frustratingly easy domain adaptation [A]. Thirtieth AAAI Conference on Artificial Intelligence [C]. Phoenix, Arizona: AAAI, 2016. 2058 – 2065.
- [20] Zhang J, Li W, Ogunbona P. Joint geometrical and statistical alignment for visual domain adaptation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2017. 1859 – 1867.

作者简介



赵 鹏 女, 1976 年生于安徽合肥. 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为机器学习、图像理解.
E-mail: zhaopeng_ad@163.com



王美玉 女, 1994 年生于安徽黄山. 硕士研究生, 主要研究方向为迁移学习、图像标注.



纪 霞 女, 1983 年生于安徽宣城. 博士, 讲师, 硕士生导师, CCF 会员, 主要研究方向为智能信息处理.



刘慧婷 女, 1978 年生于安徽阜阳. 博士, 副教授, 硕士生导师, 主要研究方向为机器学习、智能推荐.