

基于迁移鲁棒稀疏编码的图像表示方法

赵 鹏^{1),2)} 王 维²⁾ 刘慧婷^{1),2)} 纪 霞^{1),2)}

¹⁾(安徽大学计算智能与信号处理教育部重点实验室 合肥 230039)

²⁾(安徽大学计算机科学与技术学院 合肥 230601)

摘 要 图像表示是图像处理和图像理解研究中的关键问题之一. 在图像的低层表示上有很多重要的研究工作, 例如 HOG, SIFT 等. 然而在图像的低层表示和高层语义间仍然存在着巨大的鸿沟. 因而, 很多机器学习的方法被用来学习图像的高层表示, 例如主成分分析, 稀疏编码, 非负矩阵分解以及低秩表示等. 传统机器学习假设标记图像和未标记图像服从同一分布, 图像表示的误差服从高斯分布. 然而现实中图像数据更新速度快, 而且图像生成环境存在差异性, 导致未标记图像与已标记图像不服从同一分布, 因而需要重新标记数据和训练模型. 并且图像数据容易出现异常, 例如遮挡、腐蚀等等, 从而不能再用高斯分布来估计误差. 迁移学习允许标记图像(训练数据)和未标记图像(测试数据)服从不同的分布. 基于迁移学习的图像表示方法学习一个新的好的特征空间. 在这个新的特征空间下, 可以较好地描述标记图像和未标记图像的语义信息. 并且在这个新的特征空间下, 从训练集中标记图像上学习到的统计模型(例如分类模型), 可以较好地迁移到测试集中未标记图像上, 从而充分利用已标记图像, 将学习到的知识迁移到未标记的图像集上. 该文提出了一种基于迁移鲁棒稀疏编码的图像表示方法, 引入权值矩阵削弱异常点对分类的干扰, 使用稀疏编码获得数据的高级语义, 利用最小化最大均值差异缩小源域和目标域图像集之间的分布差异以及图拉普拉斯项保留图像集的几何特性. 该文的主要贡献在于: 一是通过权值矩阵泛化残差分布, 使得所提出的基于迁移鲁棒稀疏编码的图像表示方法能大大减少异常点对编码和字典学习的影响; 二是在鲁棒字典学习过程中, 采用正则化参数代替迁移稀疏编码中的字典约束, 从而将其转化为无约束优化问题, 避免了拉格朗日求解法的复杂性. 在几个通用迁移学习数据集上的对比实验结果表明, 该文所提出的图像表示方法在分类上的平均准确率比其它 6 种相关主流方法均有不同程度的提高, 证明了其有效性和鲁棒性.

关键词 迁移学习; 鲁棒稀疏编码; 图像表示; 最大均值差异; 异常点

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2017.02421

An Image Representation Method Based on Transfer Robust Sparse Coding

ZHAO Peng^{1),2)} WANG Wei²⁾ LIU Hui-Ting^{1),2)} JI Xia^{1),2)}

¹⁾(Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230039)

²⁾(School of Computer Science and Technology, Anhui University, Hefei 230601)

Abstract Image feature representation method is one of the crucial problems in image process and image understanding. There have many important researches on low-level feature representation of images, such as HOG, SIFT, and so on. But there is still a big gap between the low-level feature representation and high-level semantic feature representation. So many machine learning methods are employed to learn the high-level feature representation of images for capturing the semantic information of images, such as Principle Component Analysis (PCA), Sparse Coding, non-negative matrix factorization, low-rank representation and so on. Traditional machine learning assumes that the labeled images and the unlabeled images are drawn from the same distribution and the residuals of image representation follow the Gaussian distribution. However, images update

收稿日期:2016-09-21;在线出版日期:2017-05-16. 本课题得到国家自然科学基金(61602004,61472001)、安徽省高校自然科学研究重点项目(KJ2016A041)、安徽省自然科学基金(1408085MF122,1508085MF127)、安徽大学信息保障技术协同创新中心公开招标课题(ADXXBZ2014-5,ADXXBZ2014-6)资助. 赵 鹏,女,1976年生,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为机器学习、智能信息处理. E-mail: zhaopeng_ad@163.com. 王 维,男,1992年生,硕士研究生,主要研究方向为迁移学习、稀疏表示. 刘慧婷,女,1978年生,博士,副教授,主要研究方向为数据挖掘、机器学习. 纪 霞,女,1983年生,博士,讲师,主要研究方向为数据挖掘、智能信息处理.

very quickly in the real world. The image generation environments are so diverse that the unlabeled images and the labeled images don't follow the same distribution. Moreover, there are often some outliers appearing in images, such as occlusion, corruption and so on. The feature representation residuals of the images can not be estimated with Gaussian distribution. So the data should be relabeled and the model should be retrained. Transfer learning methods allow the distribution of the labeled images (also named the training data) and the unlabeled images (also named the test data) to be different. The feature representation methods based on transfer learning aim to design a new "good" feature representation space, which can describe the semantic information of the labeled images and the unlabeled images better. In the new feature representation space, the statistic models (e. g. classification) learned from the training data (the labeled images) can be transferred to the test data (the unlabeled images). So the feature representation methods based on the transfer learning take full advantage of the labeled images and realize the knowledge transfer across image sets with different distributions. This paper proposes an image representation method based on transfer robust sparse coding (TRSC). In this method, the weight matrix is used to weaken the interference of outliers to classification. The high level semantic content of the image is obtained with sparse coding. The differences between the source images and the target images are reduced by minimizing the maximum mean discrepancy. The geometrical properties of the image database are preserved by graph Laplacian term. There are two main contributions in this paper. One is that the weight matrix is employed to generalize the residual distribution, which weakens the inference of outliers to the code and dictionary learning in the proposed image representation method based on Transfer Robust Sparse Coding (TRSC). The other is that the regularization parameter takes the place of the dictionary constraint of Transfer Sparse Coding (TSC) in the robust dictionary learning, which transfers the robust dictionary learning to the optimization problem and avoids the complexity of Lagrange solver. Experimental results on the common transfer learning databases show the proposed method significantly improves the average accuracies in classification compared to other 6 related representation methods, which proves the effectiveness and robustness of the proposed method.

Keywords transfer learning; robust sparse coding; image representation; maximum mean discrepancy; outliers

1 引 言

传统的机器学习方法建立在训练样本和测试样本服从同一分布的假设上,当测试样本分布改变时,就需要重新标记训练样本,重新学习、建模.在信息量迅速增长并不断更新的大数据时代,迁移学习引起了越来越多学者的关注^[1].迁移学习与传统机器学习最重要的区别是迁移学习放宽了源域和目标域样本同分布的假设,目标域样本特征分布发生变化,迁移学习可以获得良好的学习效果.

特征表示方法是迁移学习中的关键研究领域之一.现有的图像特征表示方法主要包括基于稀疏编码的特征表示,基于非负矩阵分解的特征表示和基

于低秩重构的特征表示. Olshausen 等人^[2]模仿生物视觉神经提出稀疏编码,由于其简洁性和可解释性,使其在图像特征提取领域被广泛应用^[3-5]. Long 等人^[6]首次把稀疏编码引入迁移学习,提出迁移稀疏编码(Transfer Sparse Coding, TSC). Lee 等人^[7]提出非负矩阵分解. 由于非负矩阵分解能够挖掘图像中的隐藏因子^[8-9],在同构迁移学习中, Long 等人^[8]提出联合图正则化的迁移学习,通过矩阵非负二分解和三分解,提取图像隐藏因子,联合图正则化减小分布差距. 在异构迁移学习中, Yang 等人^[9]针对文本到图像的迁移学习,提出鲁棒非负联合矩阵分解. 低秩图像重构近来广受关注^[10-12], Jhuo 等人^[13]提出通过低秩重构来实现鲁棒视觉域适应.

本文在迁移学习中融合了稀疏编码. 稀疏编码

的一个关键问题是如何在提高编码精度的同时保持信号不失真. 对此研究者在稀疏编码模型中引入了不同的约束. Gao 等人^[14]引入了拉普拉斯项. Zheng 等人^[15]将其进一步拓展为图正则化稀疏编码算法 (Graph Sparse Coding, GSC). Wright 等人^[16]首次把稀疏编码引入人脸识别领域, 提出基于稀疏编码的人脸识别 (Sparse Representation-based Classification, SRC). Yang 等人^[17]针对人脸图像中存在的遮挡和腐蚀问题, 在 SRC 的基础上进一步提出了鲁棒稀疏编码 (Robust Sparse Coding, RSC). Mukherjee 等人^[18]针对图像分类提出一个带有字典原子关联度的非负稀疏编码方法.

现实图像数据往往容易存在异常 (例如遮挡、腐蚀等), 为了解决迁移学习中的这一类问题, 本文提出了一个新的基于迁移鲁棒稀疏编码的图像表示方法 (Transfer Robust Sparse Coding, TRSC). TRSC 模型分别使用权值矩阵削弱异常点对迁移学习的干扰, 稀疏编码获取图像的高级语义, 最小化最大均值差异 (Maximum Mean Discrepancy, MMD) 缩小图像集之间的分布差异, 利用图拉普拉斯项保留图像集的几何特性, 把迁移学习中的图像表示问题转化为对模型的优化求解. 对于本文目标函数的优化, 难点是 l_1 正则的最小二乘问题不可导. Lee 等人^[19]提出特征符号搜索法, 将这个不可导问题转化为无约束二次规划问题, TSC 模型^[6]、GSC 模型^[15]以及针对稀疏编码的在线字典学习模型^[20]等均采用了该方法求解目标函数优化问题.

不同于之前的相关工作, 本文提出一个一般化模型 TRSC. 与 RSC 不同, 本文提出鲁棒字典学习, 且不要求源域数据无异常, 直接在编码过程中自动获取异常点的位置. 并且 TRSC 将 RSC 中单个测试数据的编码学习推广到了整个目标域数据集的编码学习, 从而大大提高算法的效率; 与 TSC 不同, 本文主要从两方面做了进一步改进, 首先通过权值矩阵泛化残差分布, 使得 TRSC 能大大减少异常点对编码和字典学习的影响, 其次, 在鲁棒字典学习过程中, 采用正则化参数代替 TSC 中的字典约束, 从而将其转化为无约束优化问题, 避免了拉格朗日求解法的复杂性.

本文第 2 节介绍 TSC 模型; 第 3 节提出 TRSC 模型和它的优化方法及算法流程; 第 4 节在几个通用迁移学习数据集上, 通过实验对所提出的算法进行性能测试, 并对实验结果和参数进行分析; 最后对本文的工作进行总结.

2 相关工作

迁移稀疏编码模型在 GSC 的基础上进一步最小化最大均值差异, 使得该模型能够解决两个数据集分布差异大的问题, 它的数学模型如式 (1) 所示

$$\min_{\mathbf{B}, \mathbf{S}} \|\mathbf{X} - \mathbf{BS}\|_F^2 + \text{tr}(\mathbf{S}(\mu\mathbf{M} + \gamma\mathbf{L})\mathbf{S}^T) + \lambda \|\mathbf{S}\|_1, \quad \text{s. t. } \|\mathbf{B}\|_F^2 \leq \nu \quad (1)$$

其中 $\mathbf{X} \in R^{m \times n}$ 是待编码的图像矩阵, 包含源域 $\mathbf{X}_s \in R^{m \times n_s}$ 和目标域 $\mathbf{X}_t \in R^{m \times n_t}$ ($n = n_s + n_t$), m 为每幅图像的维度, n 为图像的个数, $\mathbf{B} \in R^{m \times k}$ 是字典矩阵, $\mathbf{S} \in R^{k \times n}$ 是编码矩阵.

式 (1) 共包括 3 项. 第 1 项是稀疏编码, 获得图像数据的高级语义; 第 2 项分别利用 $\mathbf{M}$ 矩阵最小化源域和目标域的差异和 $\mathbf{L}$ 矩阵保留图像数据集的几何特征. 其中无参数距离矩阵 $\mathbf{M} \in R^{n \times n}$ 度量不同数据集分布的差异, 图拉普拉斯矩阵 $\mathbf{L} \in R^{n \times n}$ 保留数据集的几何特性, 即如果两个数据在几何结构上的距离越近, 那么它们的编码应该越接近. 第 3 项利用正则化参数 λ , 调节编码的稀疏度和对矩阵 $\mathbf{X}$ 的近似程度. 最后对字典基向量的约束是为了控制模型的复杂度.

式 (1) 的优化, 可以分为以下两个交替迭代步骤:

① 固定字典 $\mathbf{B}$, 最优化编码 $\mathbf{S}$, 问题转化为带有 l_1 正则的最小二乘优化问题, 优化目标函数转换为式 (2)

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{BS}\|_F^2 + \text{tr}(\mathbf{S}(\mu\mathbf{M} + \gamma\mathbf{L})\mathbf{S}^T) + \lambda \|\mathbf{S}\|_1 \quad (2)$$

② 固定编码 $\mathbf{S}$, 最优化字典 $\mathbf{B}$, 问题转化为带有 l_2 约束的最小二乘优化问题, 优化目标函数转换为式 (3)

$$\min_{\mathbf{B}} \|\mathbf{X} - \mathbf{BS}\|_F^2, \quad \text{s. t. } \|\mathbf{B}\|_F^2 \leq \nu \quad (3)$$

下面简单介绍式 (1) 中第二部分的推导过程. TSC 采用不同域样本的一阶矩的差值来度量源域和目标域间概率分布差异^[6]. 最小化源域和目标域差异即转化为最小化公式 (4). s_i 为第 i 幅图像的编码

$$\min \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} s_i - \frac{1}{n_t} \sum_{j=1}^{n_t} s_j \right\|_2^2 = \sum_{i,j=1}^n s_i^T s_j \mathbf{M}_{ij} = \text{tr}(\mathbf{SMS}^T) \quad (4)$$

其中:

$$\mathbf{M}_{ij} = \begin{cases} 1/n_s^2, & \mathbf{x}_i, \mathbf{x}_j \in D_s \\ 1/n_t^2, & \mathbf{x}_i, \mathbf{x}_j \in D_t \\ -\frac{1}{n_s n_t}, & \text{其他} \end{cases} \quad (5)$$

TSC 利用图正则化保留数据集的几何特征. 令 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 代表 n 个图像数据点, 构建一个具有 n 个顶点的 p 近邻图 G , 每个顶点代表一个图像数据点. $\mathbf{H}$ 是 G 的权值矩阵, 如果数据点 $\mathbf{x}_i$ 属于 $\mathbf{x}_j$ 的 p 近邻或者 $\mathbf{x}_j$ 属于 $\mathbf{x}_i$ 的 p 近邻, 那么 $\mathbf{H}_{ij} = 1$, 否则 $\mathbf{H}_{ij} = 0$. 令 $d_i = \sum_{j=1}^n \mathbf{H}_{ij}$, $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$, 则图拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{H}$. 编码后保持图 G 中数据点的几何特性转为最小化公式(6)

$$\min \frac{1}{2} \sum_{i,j=1}^n \|\mathbf{s}_i - \mathbf{s}_j\|_2^2 \mathbf{H}_{ij} = \text{tr}(\mathbf{S} \mathbf{L} \mathbf{S}^T) \quad (6)$$

根据式(6), 当 $\mathbf{H}_{ij} \neq 0$ 时, 即两点之间几何上接近, 此时为了目标函数的最小化, 编码 $\mathbf{s}_i, \mathbf{s}_j$ 应该接近; 当 $\mathbf{H}_{ij} = 0$ 时, 即两点无近邻关系, 此时对编码距离无约束.

3 基于迁移鲁棒稀疏编码的图像表示方法

3.1 迁移鲁棒稀疏编码模型

TSC 假设误差服从高斯分布, 然而当图像存在遮挡, 腐蚀或者其它异常的情况时, 误差将偏离此分布, 因而学习获得的编码系数与字典基向量不具有鲁棒性.

为了提高编码学习和字典学习的鲁棒性, 本文在 TSC 基础上提出一个迁移鲁棒稀疏编码模型 (TRSC), 其目标函数如式(7)所示

$$\min_{\mathbf{B}, \mathbf{S}} \|\mathbf{W}^{1/2} \text{vec}[(\mathbf{X} - \mathbf{B}\mathbf{S})]\|_2^2 + \text{tr}(\mathbf{S}(\mu\mathbf{M} + \gamma\mathbf{L})\mathbf{S}^T) + \lambda \|\mathbf{S}\|_1 + \xi \|\mathbf{B}\|_F^2 \quad (7)$$

下面给出式(7)的推导过程. 为了讨论问题方便, 首先只考虑稀疏编码部分, 即

$$\min_{\mathbf{B}, \mathbf{S}} \|\mathbf{X} - \mathbf{B}\mathbf{S}\|_F^2 \quad (8)$$

令 $\mathbf{E} = \mathbf{X} - \mathbf{B}\mathbf{S}$, $\mathbf{E} \in R^{m \times n}$ 为残差矩阵, 假设每个残差矩阵中每个元素 e_{ij} 独立同分布, 且密度函数为 $h_\sigma(e_{ij})$, 其中 σ 为概率分布的参数.

残差矩阵的极大似然估计为

$$L_\sigma(\mathbf{e}_{11}, \mathbf{e}_{21}, \dots, \mathbf{e}_{nm}) = \prod_{i=1}^m \prod_{j=1}^n h_\sigma(e_{ij}) \quad (9)$$

考虑对数似然, 则最大化公式(9)等价于最小化公式(10)

$$-\ln L_\sigma = \sum_{i=1}^m \sum_{j=1}^n \rho_\sigma(e_{ij}) \quad (10)$$

其中 $\rho_\sigma(e_{ij}) = -\ln h_\sigma(e_{ij})$, 考虑异常点对稀疏编码的影响, 将式(8)转化为求解式(11)的最小化问题:

$$\min_{\mathbf{B}, \mathbf{S}} \sum_{i=1}^m \sum_{j=1}^n (\rho_\sigma(\mathbf{X}_{ij} - (\mathbf{B}\mathbf{S})_{ij}))^2 \quad (11)$$

一般地, 假设密度函数 $h_\sigma(e_{ij})$ 对称, 且

$$h_\sigma(e_{ij}) < h_\sigma(e_{kl}), \text{ if } |e_{ij}| > |e_{kl}| \quad (12)$$

因此 $\rho_\sigma(e_{ij})$ 有如下性质: $\rho_\sigma(0)$ 是全局最小值, 不失一般性, 假设该点的值为 0.

且

$$\rho_\sigma(e_{ij}) = \rho_\sigma(-e_{ij}) \quad (13)$$

$$\rho_\sigma(e_{ij}) < \rho_\sigma(e_{kl}), \text{ if } |e_{ij}| > |e_{kl}| \quad (14)$$

令

$$F_\sigma(\mathbf{E}) = \sum_{i=1}^m \sum_{j=1}^n \rho_\sigma(e_{ij})^2 \quad (15)$$

用它在 $\hat{\mathbf{E}}$ 点的一阶泰勒展开式来近似它, 得

$$\tilde{F}_\sigma(\mathbf{E}) = F_\sigma(\hat{\mathbf{E}}) + (\text{vec}(\mathbf{E} - \hat{\mathbf{E}}))^T \left(\frac{\partial F_\sigma(\hat{\mathbf{E}})}{\partial \text{vec}(\hat{\mathbf{E}})} \right) + R_1(\mathbf{E}) \quad (16)$$

$$F'_\sigma(\hat{\mathbf{E}}) = \left(\frac{\partial F_\sigma(\hat{\mathbf{E}})}{\partial \text{vec}(\hat{\mathbf{E}})} \right) = [2\rho'_\sigma(\hat{e}_{11}); 2\rho'_\sigma(\hat{e}_{21}); \dots; 2\rho'_\sigma(\hat{e}_{nm})] \quad (17)$$

公式中的 vec 表示矩阵列向量化. 在稀疏编码中, 一般期望保真项严格凸^[17], 因而令一阶残差项为式(18)所示

$$R_1(\mathbf{E}) = 0.5 (\text{vec}(\mathbf{E} - \hat{\mathbf{E}}))^T \mathbf{W} \cdot (\text{vec}(\mathbf{E} - \hat{\mathbf{E}})) \quad (18)$$

其中 $\mathbf{W} \in R^{mn \times mn}$ 为对角矩阵. $F_\sigma(\mathbf{E})$ 在 $\mathbf{0}$ 点处取得最小值, 因此 $\tilde{F}_\sigma(\mathbf{E})$ 在该点也取得最小值, 即 $\tilde{F}'_\sigma(\mathbf{0}) = \mathbf{0}$, 而

$$\tilde{F}'_\sigma(\mathbf{0}) = \left(\frac{\partial F_\sigma(\hat{\mathbf{E}})}{\partial \text{vec}(\hat{\mathbf{E}})} \right) - \mathbf{W} \text{vec}(\hat{\mathbf{E}}) \quad (19)$$

从而

$$\mathbf{w}_{ii} = (F'_\sigma(\hat{\mathbf{E}}))_i / (\text{vec}(\hat{\mathbf{E}}))_i \quad (20)$$

根据 $\rho_\sigma(e_{ij})$ 的性质可得 $\mathbf{W}$ 矩阵的元素都为非负值, 且

$$\begin{aligned} \tilde{F}_\sigma(\mathbf{E}) &= F_\sigma(\hat{\mathbf{E}}) + (\text{vec}(\mathbf{E} - \hat{\mathbf{E}}))^T \left(\frac{\partial F_\sigma(\hat{\mathbf{E}})}{\partial \text{vec}(\hat{\mathbf{E}})} \right) + R_1(\mathbf{E}) \\ &= F_\sigma(\hat{\mathbf{E}}) + 0.5 \|\mathbf{W}^{1/2} \text{vec}(\mathbf{E})\|_2^2 + \\ &\quad 0.5 (\text{vec}(\hat{\mathbf{E}}))^T (\text{vec}(\hat{\mathbf{E}})) \\ &= 0.5 \|\mathbf{W}^{1/2} \text{vec}(\mathbf{E})\|_2^2 + b \end{aligned} \quad (21)$$

其中 $b = F_\sigma(\hat{\mathbf{E}}) + 0.5 (\text{vec}(\hat{\mathbf{E}}))^T (\text{vec}(\hat{\mathbf{E}}))$, b 只与 $\hat{\mathbf{E}}$ 有关, 即为常数. 由于常数项不影响最小化问题的求解, 从而得到本文所提出的迁移鲁棒稀疏编码模型的鲁棒稀疏编码部分:

$$\min_{\mathbf{B}, \mathbf{S}} \|\mathbf{W}^{1/2} \text{vec}[(\mathbf{X} - \mathbf{B}\mathbf{S})]\|_2^2 \quad (22)$$

$\mathbf{W}$ 为一个对角权值矩阵, 它的每个对角元素分别对应某幅图像的某个像素点, 误差大, 则权值小,

误差小,则权值大. 由于逻辑函数与此性质吻合,因此本文将权值函数定义为

$$w_{ii} = \frac{\exp(\beta\delta - \beta e_{kl}^2)}{1 + \exp(\beta\delta - \beta e_{kl}^2)}, m(l-1) + k = i \quad (23)$$

式(23)进一步化简为

$$w_{ii} = \frac{1}{1 + \frac{1}{\exp(\beta\delta - \beta e_{kl}^2)}}, m(l-1) + k = i \quad (24)$$

把 $\mathbf{W}$ 矩阵表示成一个分块矩阵的形式,如式(25)所示

$$\begin{bmatrix} \mathbf{W}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{W}^n \end{bmatrix} \in R^{mn \times mn} \quad (25)$$

其中, $\mathbf{W}^i \in R^{m \times m}$ 即第 i 幅图像对应的权值对角矩阵, 对角线上每个元素对应第 i 幅图像上每个像素的权值.

计算每幅图像 i 的残差平方向量 $\Psi_i = [(e_{1i})^2, (e_{2i})^2, \dots, (e_{mi})^2]$, $i=1, \dots, n$. 然后将该向量的各分量按升序排列获得一个新的有序序列 $(\Psi_i)_a$, 令 $q = \lfloor \tau n \rfloor$, 其中 $\tau \in (0, 1]$ 为可调参数. 令 δ 为序列 $(\Psi_i)_a$ 的第 q 个残差分量的平方, 由式(24)可得此时第 q 个残差的像素对应的权值为 0.5, 即为该图像的权值中心点. 因而当 q 取值越大, 则该幅图像就有越多的像素点对应的权值大于 0.5. 令 $\beta = \frac{c}{\delta}$, 它控制权值从 1 到 0 的下降速度.

进一步将式(22)与 TSC 模型中的最小化最大均值差异与最小化图正则化项合并, 同时将字典约束转化为目标函数中的正则项, 从而得到本文所提出的模型 TRSC 的目标函数(7).

由式(7)不难得出, 当 $\mathbf{W}$ 矩阵的对角元素都为 2 时, 模型转化为 TSC, 即二范式, 误差服从高斯分布. 当 $\mathbf{W}$ 矩阵的对角元素为 $\frac{1}{|e_{ij}|}$ 时, 模型转化为一范式, 误差服从拉普拉斯分布^[17]. TRSC 模型对图像数据进行了归一化, 为了避免出现权值无穷大, 故将权值范围归一化为 $[0, 1]$. TSC 相当于 TRSC 中给每个像素点赋予同样的权值的特殊情况, 显然不能处理异常点. 一范式模型虽然同样给残差大的像素点赋予小的权值, 但是无界, 当残差无限小时, 权值无限大, 模型稳定性差. 因此 TRSC 一般模型泛化了残差分布的各种情况, 而且边界的约束使得模型稳定性较高.

本文将目标函数式(7)转化为如下两个子问题的求解.

① 编码优化问题, 即固定字典 $\mathbf{B}$, 最优化编码 $\mathbf{S}$, 问题转化为带有 l_1 正则的最小二乘优化问题, 优化目标函数转换为式(26)

$$\min_{\mathbf{S}} \|\mathbf{W}^{1/2} \text{vec}[(\mathbf{X} - \mathbf{B}\mathbf{S})]\|_2^2 + \text{tr}(\mathbf{S}(\mu\mathbf{M} + \gamma\mathbf{L})\mathbf{S}^T) + \lambda \|\mathbf{S}\|_1 \quad (26)$$

② 鲁棒字典学习问题, 即固定编码 $\mathbf{S}$, 最优化字典 $\mathbf{B}$, ξ 是字典正则化参数, 取代了 TSC 子问题②中的字典约束, 将其转化为一个无约束的最小二乘优化问题, 优化目标函数转换为式(27)

$$\min_{\mathbf{B}} \|\mathbf{W}^{1/2} \text{vec}[(\mathbf{X} - \mathbf{B}\mathbf{S})]\|_2^2 + \xi \|\mathbf{B}\|_F^2 \quad (27)$$

3.2 编码优化

编码优化子问题的目标函数式(26)等价式为(28)

$$\min_{\mathbf{s}_i} \sum_{i=1}^n \|(\mathbf{W}^i)^{1/2} (\mathbf{x}_i - \mathbf{B}\mathbf{s}_i)\|_2^2 + \sum_{i,j=1}^n (\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij}) \mathbf{s}_i^T \mathbf{s}_j + \lambda \sum_{i=1}^n \|\mathbf{s}_i\|_1 \quad (28)$$

进一步转化为每一幅图像编码对应的目标函数的最小化

$$\min_{\mathbf{s}_i} f(\mathbf{s}_i) = \|(\mathbf{W}^i)^{1/2} (\mathbf{x}_i - \mathbf{B}\mathbf{s}_i)\|_2^2 + \lambda \|\mathbf{s}_i\|_1 + (\mu\mathbf{M}_{ii} + \gamma\mathbf{L}_{ii}) \mathbf{s}_i^T \mathbf{s}_i + \mathbf{s}_i^T \boldsymbol{\ell}_i \quad (29)$$

其中

$$\boldsymbol{\ell}_i = 2 \sum_{j \neq i} (\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij}) \mathbf{s}_j.$$

式(29)在 $\mathbf{s}_i$ 的分量为 0 时不可导, 从而使得标准的无约束优化方法不可行. 本文采用特征符号搜索法来求解式(29). 不可导问题的非光滑优化方法求解中, 有一个必要条件: 若某个向量是局部极小点, 那么零向量是它次微分集合中的一个元素^[21].

令

$$g(\mathbf{s}_i) = \|(\mathbf{W}^i)^{1/2} (\mathbf{x}_i - \mathbf{B}\mathbf{s}_i)\|_2^2 + (\mu\mathbf{M}_{ii} + \gamma\mathbf{L}_{ii}) \mathbf{s}_i^T \mathbf{s}_i + \mathbf{s}_i^T \boldsymbol{\ell}_i \quad (30)$$

则

$$f(\mathbf{s}_i) = g(\mathbf{s}_i) + \lambda \sum_{j=1}^k |\mathbf{s}_i^{(j)}| \quad (31)$$

$\mathbf{s}_i^{(j)}$ 为 $\mathbf{s}_i$ 的第 j 个分量, 令 $\nabla_i^{(j)} |\mathbf{s}_i|$ 为 $\mathbf{s}_i$ 的第 j 个分量的次微分值, 如果 $|\mathbf{s}_i^{(j)}| > 0$, 那么 $\nabla_i^{(j)} |\mathbf{s}_i| = \text{sign}(\mathbf{s}_i^{(j)})$, 如果 $|\mathbf{s}_i^{(j)}| = 0$, 那么不可导, 并且 $\nabla_i^{(j)} |\mathbf{s}_i|$ 可以在 $[-1, 1]$ 之间取值.

$f(\mathbf{s}_i)$ 取得最小值的优化条件是 $\nabla_i f(\mathbf{s}_i) = 0$, 即

$$\nabla_i^{(j)} f(\mathbf{s}_i) = \begin{cases} \nabla_i^{(j)} g(\mathbf{s}_i) + \lambda \text{sign}(\mathbf{s}_i^{(j)}) = 0, & \text{若 } |\mathbf{s}_i^{(j)}| \neq 0 \\ |\nabla_i^{(j)} g(\mathbf{s}_i)| \leq \lambda, & \text{其他} \end{cases} \quad (32)$$

当式(32)的第 2 个条件无法满足时: ① 当 $\nabla_i^{(j)} g(\mathbf{s}_i) > \lambda$, 此时无论 $\text{sign}(\mathbf{s}_i^{(j)})$ 取什么值, $\nabla_i^{(j)} f(\mathbf{s}_i) > 0$, 为了最小化目标函数, 则应该降低 $\mathbf{s}_i$

的值,则 $\text{sign}(s_i^{(j)}) = -1$; ② 当 $\nabla_i^{(j)} g(s_i) < -\lambda$, 此时无论 $\text{sign}(s_i^{(j)})$ 取什么值, $\nabla_i^{(j)} f(s_i) < 0$, 则应该增加 s_i 的值, 则 $\text{sign}(s_i^{(j)}) = 1$. $\nabla_i g(s_i)$ 的计算如式(33)所示

$$\nabla_i g(s_i) = -2\mathbf{B}^T((\mathbf{W}^i)^{1/2})^T((\mathbf{W}^i)^{1/2})^T(x_i - \mathbf{B}s_i) + 2(\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij})s_i + \ell_i \quad (33)$$

下面给出特征符号搜索法过程中使用到的激活集和符号集的定义.

定义 1. 每个 s_i 对应一个激活集 $\mathbf{A}_i: \mathbf{A}_i = \{j | s_i^{(j)} = 0 \wedge |\nabla_i^{(j)} g(s_i)| > \lambda\}$.

定义 2. s_i 每个分量元素的符号构成一个符号集 $\boldsymbol{\theta}_i: \boldsymbol{\theta}_i = [\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(k)}]$, $\theta_i^{(j)} \in \{-1, 0, 1\}$ 代表 $s_i^{(j)}$ 的符号.

同时在特征符号搜索法优化过程中引入 $\hat{\mathbf{B}}, \hat{s}_i, \hat{\ell}_i$ 和 $\hat{\boldsymbol{\theta}}_i$. 例如当 $\mathbf{A}_i = \{1, 3, 5\}$ 时, $\hat{\mathbf{B}}$ 为 $\mathbf{B}$ 的第 1, 3, 5 列形成的子矩阵, $\hat{s}_i$ 为 s_i 的第 1, 3, 5 个分量形成的列向量, 同理 $\hat{\ell}_i, \hat{\boldsymbol{\theta}}_i$ 分别为 $\ell_i, \boldsymbol{\theta}_i$ 对应分量形成的列向量. 从而编码优化子问题进一步转化为无约束二次优化问题.

$$\min_{\hat{s}_i} \hat{f}(\hat{s}_i) = \|(\mathbf{W}^i)^{1/2}(x_i - \hat{\mathbf{B}}\hat{s}_i)\|_2^2 + (\mu\mathbf{M}_{ii} + \gamma\mathbf{L}_{ii})\hat{s}_i^T\hat{s}_i + \hat{s}_i^T\hat{\ell}_i + \lambda\hat{\boldsymbol{\theta}}_i^T\hat{s}_i \quad (34)$$

$$\frac{\partial \hat{f}(\hat{s}_i)}{\partial \hat{s}_i} = -2\hat{\mathbf{B}}^T((\mathbf{W}^i)^{1/2})^T((\mathbf{W}^i)^{1/2})(x_i - \hat{\mathbf{B}}\hat{s}_i) + 2(\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij})\hat{s}_i + \hat{\ell}_i + \lambda\hat{\boldsymbol{\theta}}_i \quad (35)$$

令

$$\frac{\partial \hat{f}(\hat{s}_i)}{\partial \hat{s}_i} = 0$$

$$-2\hat{\mathbf{B}}^T((\mathbf{W}^i)^{1/2})^T((\mathbf{W}^i)^{1/2})(x_i - \hat{\mathbf{B}}\hat{s}_i) + 2(\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij})\hat{s}_i + \hat{\ell}_i + \lambda\hat{\boldsymbol{\theta}}_i = 0 \quad (36)$$

$$\hat{s}_i^{\text{new}} = (\hat{\mathbf{B}}^T((\mathbf{W}^i)^{1/2})^T((\mathbf{W}^i)^{1/2})\hat{\mathbf{B}} + (\mu\mathbf{M}_{ij} + \gamma\mathbf{L}_{ij})\mathbf{I})^{-1} (\hat{\mathbf{B}}^T((\mathbf{W}^i)^{1/2})^T((\mathbf{W}^i)^{1/2})x_i - (\hat{\ell}_i - \lambda\hat{\boldsymbol{\theta}}_i)/2) \quad (37)$$

$\hat{s}_i^{\text{new}} (i=1, 2, \dots, n)$ 即为优化后的编码.

3.3 鲁棒字典学习

鲁棒字典学习子问题目标函数式(27)等价于式(38)

$$\min_{\mathbf{B}} \|\mathbf{Q} \otimes (\mathbf{X} - \mathbf{B}\mathbf{S})\|_{\text{F}}^2 + \xi \|\mathbf{B}\|_{\text{F}}^2 \quad (38)$$

其中 $\otimes$ 代表矩阵点乘操作, 即对应元素相乘. 定义符号 diag^{-1} 为逆对角化, 即将一个对角矩阵的对角元素按序排成一个列向量, 上式中 $\mathbf{Q} \in R^{m \times n}$, 且 $\mathbf{Q} = \text{unvec}_{m,n}(\text{diag}^{-1}(\mathbf{W}^{1/2}))$. 由于 $\otimes$ 的存在, 不同于 TSC 中字典学习, 鲁棒字典学习过程如下.

式(38)等价于式(39)

$$\min_{\mathbf{B}} \|\mathbf{Q}^T \otimes (\mathbf{X}^T - \mathbf{S}^T \mathbf{B}^T)\|_{\text{F}}^2 + \xi \|\mathbf{B}^T\|_{\text{F}}^2 \quad (39)$$

为了表达简洁, 不妨令 $\mathbf{J} = \mathbf{B}^T, \mathbf{K} = \mathbf{Q}^T, \mathbf{O} = \mathbf{X}^T$,

$\mathbf{U} = \mathbf{S}^T$, 则式(39)等价于式(40)

$$\min_{\mathbf{J}} \|\mathbf{K} \otimes (\mathbf{O} - \mathbf{U}\mathbf{J})\|_{\text{F}}^2 + \xi \|\mathbf{J}\|_{\text{F}}^2 \quad (40)$$

进一步转化为式(41)

$$\min_{\mathbf{J}_i} \sum_{i=1}^m \|\mathbf{K}_i \otimes (\mathbf{O}_i - \mathbf{U}\mathbf{J}_i)\|_{2,2}^2 + \xi \sum_{i=1}^m \|\mathbf{J}_i\|_2^2 \quad (41)$$

其中 $\mathbf{K}_i, \mathbf{O}_i, \mathbf{J}_i$ 分别对应该矩阵的第 i 列. 式(41)归为对每个子问题的求解, 即

$$\min_{\mathbf{J}_i} \|\text{diag}(\mathbf{K}_i)(\mathbf{O}_i - \mathbf{U}\mathbf{J}_i)\|_2^2 + \xi \|\mathbf{J}_i\|_2^2 \quad (42)$$

由于式(42)有解析解, 从而学习到一个鲁棒字典, 具有泛化残差分布的能力.

式(42)的解析解如下

$$\mathbf{J}_i = (\mathbf{U}^T \text{diag}(\mathbf{K}_i) \text{diag}(\mathbf{K}_i) \mathbf{U} + \xi \mathbf{I})^{-1} \mathbf{U}^T \text{diag}(\mathbf{K}_i) \text{diag}(\mathbf{K}_i) \mathbf{O}_i \quad (43)$$

因而所求字典矩阵 $\mathbf{B}$ 为

$$\mathbf{B} = \mathbf{J}^T \quad (44)$$

3.4 TRSC 算法流程

输入: 数据矩阵 $\mathbf{X}$, 最大均值差异矩阵 $\mathbf{M}$, 图拉普拉斯矩阵 $\mathbf{L}$, 正则化参数 $\mu, \gamma, \lambda, \xi$, 权值参数 τ, c , 迭代次数 NT .

输出: 编码矩阵 $\mathbf{S}^* = [s_1^*, \dots, s_n^*]$

1. 字典随机初始化
2. 权值矩阵初始化
3. 编码学习

FOR $k=1$ to NT

(1) FOR $i=1$ to n

特征符号搜索法进行编码学习获得 s_i^* .

ENDFOR

(2) 根据式(44)获得字典 $\mathbf{B}$

(3) 根据式(23)进行权值更新

ENDFOR

4 实验结果与分析

4.1 实验设置

为了验证所提出的 TRSC 方法的有效性, 本文分别和相关的 6 种主流图像特征表示方法进行了对比实验: 主成分分析(PCA)、稀疏编码(SC)^[19]、鲁棒稀疏编码(RSC)^[17]、图正则化稀疏编码(GSC)^[15]、不含图正则化的迁移稀疏编码(TSC_{MMD})^[6]以及迁移稀疏编码(TSC)^[6]. 实验分别用它们提取源域和目标域的特征, 然后用逻辑回归方法(LR)对源域进行训练获得分类器, 最后对目标域样本进行分类.

参考文献[6], 本文将逻辑回归分类方法中的权衡参数 C 设置为 100, 为了获得过完备字典, 字典基原子的个数 $k=200$. 对于字典约束正则化参数 ξ , 由于它限制字典基的取值范围, 实验发现当 $\xi=0.001$ 时, 字典基元素与文献[6]中拉格朗日法获得的字典

的取值范围基本一致. TRSC 在与其它算法的对比实验中,参数统一设置为 $\gamma=1, \mu=1e5, \lambda=0.1, NT=10$.

本文选择几个通用的迁移学习数据集作为实验数据集,包括字符数据集 USPS 和 MNIST^[6]、手绘草图 Sketch^[22] 和自然图像数据集 MSRC, VOC2007^[6], Caltech-256, Office^[23-25]. 所有算法随机抽取样本 10 次,取 10 次运行结果的平均值.

实验采用目标域上的分类准确率 *Accuracy* 作为算法性能的评价标准^[15, 26],其计算如式(45)所示

$$Accuracy = \frac{|\{x: x \in D_t \wedge \hat{y}(x) = y(x)\}|}{|\{x: x \in D_t\}|} \quad (45)$$

D_t 表示目标域, $y(x)$ 表示真实标签, $\hat{y}(x)$ 表示预测标签.

4.2 USPS 和 MNIST 数据集

USPS 和 MNIST 数据集的数字书写风格和明暗度不一样,分布不同. 实验从 USPS 中随机取 1800 幅图像,从 MNIST 中随机取 2000 幅图像,分别作为源域和目标域,将所有图像统一缩放成 10×10 大小,每幅图像表示成 100 维的向量. 为了验证算法鲁棒性,实验分别对两个数据集进行了不同程度的随机遮挡,从而既加大了分布的差别,又产生了异常图像. 图 1 分别给出了 MNIST 和 USPS 数据集,随机 5% 遮挡后的 MNIST 和 USPS 数据集以及随机 10% 遮挡后的 MNIST 和 USPS 数据集的示

例. 定义 6 个分类迁移学习任务 U_M, M_U, 5U_5M, 5M_5U, 10U_10M, 10M_10U 分别表示取不同程度遮挡的 MNIST 数据集或者 USPS 数据集作为源域或者目标域. 例如 5U_5M 表示取随机 5% 遮挡的 USPS 数据集作为源域,随机 5% 遮挡的 MNIST 数据集作为目标域. 实验结果如表 1 所示.

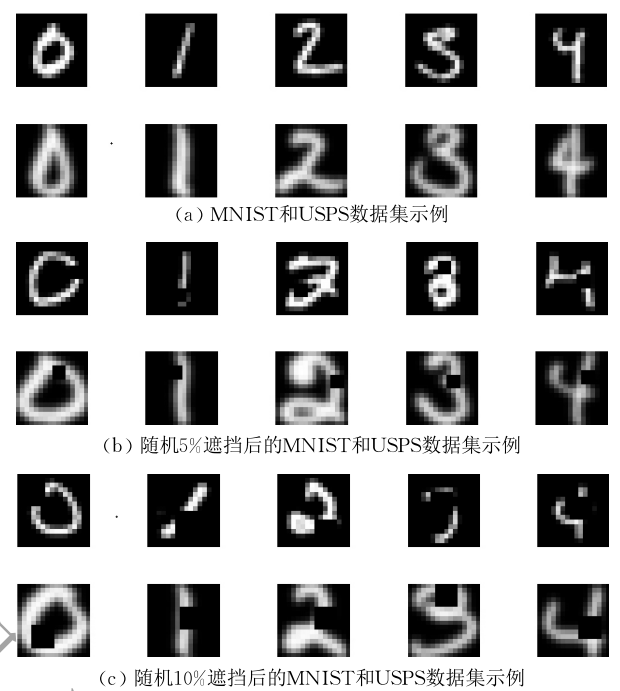


图 1 USPS 数据集和 MNIST 数据集示例

表 1 字符数据集上的平均分类准确率

方法	数据集/%						平均值
	M_U	5M_5U	10M_10U	U_M	5U_5M	10U_10M	
PCA	37.78	17.89	10.44	31.75	32.25	28.45	26.43
SC	43.66	37.15	31.91	38.96	37.23	29.95	36.48
GSC	59.29	56.15	49.08	44.26	41.83	35.34	47.66
RSC	31.95	29.90	28.95	43.33	39.22	33.11	34.41
TSC _{MMD}	49.52	45.29	41.63	44.47	41.72	35.28	42.99
TSC	54.14	53.19	48.74	52.32	50.77	41.73	50.15
TRSC	66.17	65.72	63.44	60.95	59.80	47.30	60.56

字符数据集的特点是每幅图像比较简单规则,从表 1 中可以看出无论是传统方法(PCA, SC, GSC, RSC)还是迁移学习方法(TSC_{MMD}, TSC, TRSC),分类准确率随着遮挡程度的加大而下降. 由于 RSC 是把源域作为字典,只进行了编码学习,所以当源域存在遮挡,且源域与目标域存在分布差异时,将不再能体现其鲁棒性. TSC_{MMD} 仅考虑目标域和源域的分布差异, GSC 仅考虑数据集的局部几何特性,实验结果显示, GSC 的平均准确率高出 TSC_{MMD}, 表明在字符数据集中,几何特性的保留对于分类准确率起着更为重要的作用. 而本文所提的算法 TRSC 模型不但考虑了源域和目标域的分布差异和数据集局部几何特性,且考虑了图像存在的异常情况,在此 6 组

实验结果中, SC 提取了图像数据的高级语义,平均准确率相比 PCA 提高了 10.05%. 把 SC 作为基准方法,因为后面的 5 种方法都是建立在 SC 的基础之上,与其平均准确率相比, GSC, TSC_{MMD}, TSC, TRSC 分别提高了 11.18%, 6.51%, 13.67%, 24.08%, 而 RSC 下降了 2.07%. TRSC 相比 PCA, GSC, RSC, TSC_{MMD}, TSC 分别提高了 34.13%, 12.9%, 26.15%, 17.57%, 10.41%, 体现了 TRSC 的鲁棒性和有效性.

4.3 Sketch, MSRC 和 VOC2007 数据集

Sketch 数据集是手绘草图^[22], 而 MSRC 和 VOC2007 数据集是自然拍摄的照片,且拍摄环境不一样,特征分布各不相同,数据示例如图 2 所示.

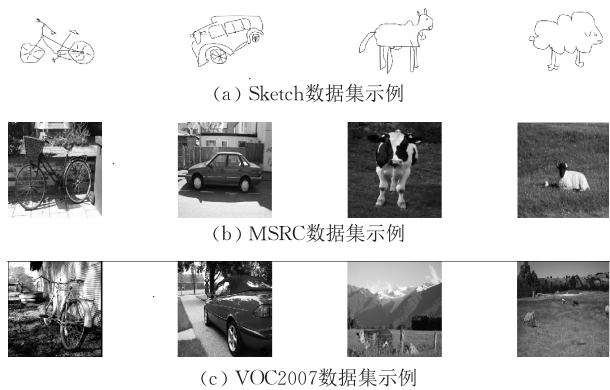


图 2 Sketch, MSRC, VOC2007 数据集示例

实验在 Sketch 和 MSRC 中随机取了 7 类,其中 Sketch 数据集有 560 幅图像,MSRC 数据集有 1365 幅图像;在 VOC2007 和 MSRC 中随机取了 6 类,其中 VOC2007 数据集有 1530 幅图像,MSRC 数据集有 1269 幅图像.把所有图像统一成 10×10 大小,每幅图像表示成 100 维的向量.定义 4 个分类迁移学习任务: S_M (Sketch 作为源域, MSRC 作为目标域), M_S (MSRC 作为源域, Sketch 作为目标域), V_M (VOC2007 作为源域, MSRC 作为目标域)和 M_V (MSRC 作为源域, VOC2007 作为目标域).实验结果如表 2 所示.

表 2 Sketch, MSRC 和 VOC2007 数据集上的平均分类准确率

方法	数据集 / %				平均值
	S_M	M_S	V_M	M_V	
PCA	27.61	14.29	37.19	20.98	25.02
SC	19.46	14.29	42.94	26.87	25.89
GSC	26.13	14.29	44.15	22.59	26.79
RSC	11.75	14.82	28.53	18.76	18.47
TSC _{MMD}	16.00	12.64	41.47	31.39	25.38
TSC	34.57	21.75	48.56	34.46	34.84
TRSC	54.15	30.95	55.16	46.58	46.71

与字符数据集不同,在自然图像数据集中,图像异常点比较多,对于分类准确率的影响比较大.从表 2 中可以看出,SC 比 PCA 平均准确率提高了 0.87%.相比 SC,在只考虑源域和目标域分布差异时,TSC_{MMD} 的平均准确率下降了 0.51%,表明只考虑分布差异的迁移学习方法不能在异常点较多的数据集获得好的迁移学习效果.只考虑数据集的

几何局部特性的 GSC 的平均准确率比 SC 提高了 0.90%.而 RSC 相比于 SC,平均准确率下降了 7.42%,主要是因为 RSC 对于字典的依赖性太强,不能自学习获得字典,因而不能较好处理含异常数据较多的字典的问题. TSC 平均准确率比 SC 提高了 8.95%,而 TRSC 平均准确率比 PCA, SC, GSC, TSC_{MMD}, RSC, TSC 分别提高了 21.69%, 20.82%, 19.92%, 21.33%, 28.24%, 11.87%.分析原因,主要由于异常点带来了以下两点影响:首先在 TSC 迭代过程中,大部分残差由图像中的异常点产生,学习过程中为了目标函数的下降,而导致模型强行表示异常点,降低了学习到的编码的分类效果;其次异常点干扰了局部特性和分布差异的衡量标准,因而干扰了模型学习保留图像正常点的局部特性和降低正常点的分布差异. TRSC 模型引入了能够动态调整的权值矩阵,对异常点赋予低权值,从而实现了鲁棒性.

4.4 Caltech-256 和 Office 数据集

Office 数据集集中的 3 个子数据集 Amazon, Dslr 和 Webcam,共计包括 31 类 4652 幅图像. Caltech-256 数据集包括 256 类 30607 幅图像.数据示例如图 3 所示.本文分别在这 4 个数据集中随机取 10 类,其中 Amazon 数据集包括 958 幅图像, Caltech-256 数据集包含 1123 幅图像, Dslr 数据集包括 157 幅图像, Webcam 数据集包括 295 幅图像.把所有图像统一成 10×10 大小,每幅图像表示成 100 维的向量.实验分别任选其中两个数据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.分别用 A, C, D, W 表示 Amazon, Caltech-256, Dslr, Webcam 数据集,例如 A_C 表示 A 作为源域, C 作为目标域的迁移学习任务.实验结果如表 3 所示.从表 3 中可以看出 Caltech-256 和 Office 同样作为自然图像集,在 12 组实验结果中,SC 比 PCA 平均准确率提高了 0.86%.相比 SC, TSC_{MMD} 平均准确率下降了 2.27%, GSC 平均准确率提高了 0.13%, RSC 平均准确率下降了 1.09%, TSC 平均准确率下降了 0.06%,而 TRSC 平均准确率相比 PCA, SC, GSC, TSC_{MMD}, RSC, TSC 分别提高了 10.35%, 9.49%, 9.35%, 11.76%, 10.57%, 9.55%.在分类迁移学习任务 D_W 和 W_D 中,迁移学习方法平均准确率有所下



图 3 Office, Caltech-256 数据集示例

表 3 Amazon, Caltech-256, Dslr 和 Webcam 数据集上的平均分类准确率

(单位: %)

方法	数据集												平均值
	A_C	A_D	A_W	C_A	C_D	C_W	D_A	D_C	D_W	W_A	W_C	W_D	
PCA	17.28	33.12	26.78	15.87	13.38	7.80	15.45	12.11	52.20	15.76	9.53	60.51	23.32
SC	15.91	24.71	22.92	16.98	13.89	8.07	18.42	12.00	62.27	14.44	10.40	70.19	24.18
GSC	14.79	28.22	24.71	18.50	17.52	13.56	14.24	15.33	56.20	14.71	11.72	62.29	24.32
RSC	15.94	22.93	15.25	18.79	14.01	11.86	19.62	11.67	64.07	11.90	9.97	61.15	23.10
TSC _{MMD}	16.99	17.77	15.83	18.68	23.06	18.98	16.45	12.12	41.12	16.33	11.62	54.01	21.91
TSC	18.07	27.64	22.14	19.34	23.31	13.86	14.39	12.00	49.93	18.58	13.07	57.13	24.12
TRSC	33.87	32.10	30.68	27.11	38.34	37.80	21.46	17.93	54.58	27.34	24.87	57.96	33.67

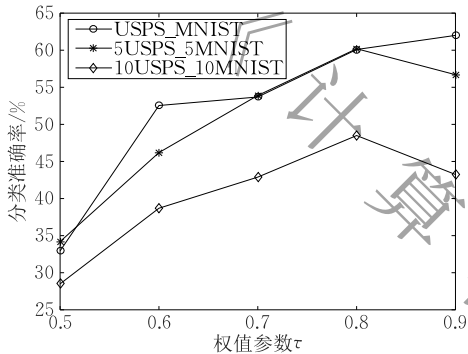
降,一方面从图 3(c)和(d)看出两个数据集分布差异非常小,不能体现迁移学习算法的有效性;另一方面源域和目标域数据集较小,分别为 157,295,不利于迁移学习训练获得一个有效的模型,降低了迁移效率。

4.5 参数分析

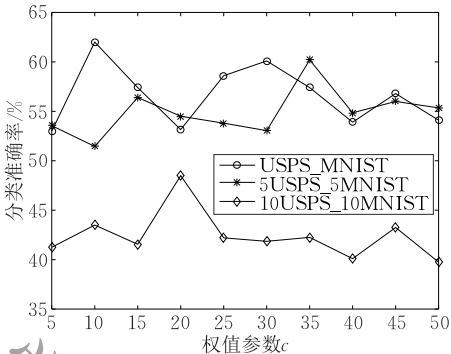
为了检验 TRSC 中各参数对方法表示性能的

影响,本文分别对权值矩阵参数 τ, c , 迭代次数 NT, 图正则化参数 γ , 最大均值差异正则化参数 μ , 稀疏正则化参数 λ 进行实验分析。

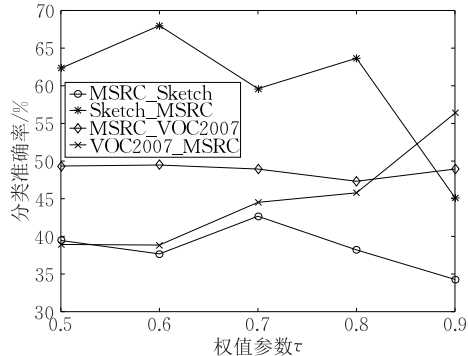
首先对权值参数 τ, c , 在 9 个数据集, 13 个分类迁移学习任务进行对比实验, 实验中分别固定一个参数值, 调节另一个参数值, 实验结果如图 4 所示。



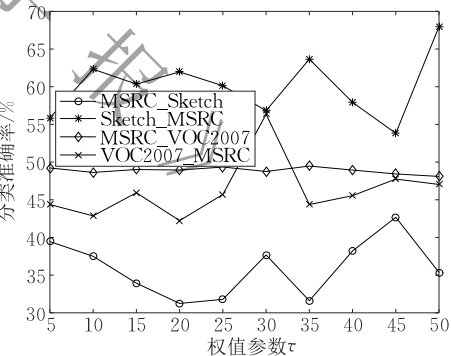
(a) τ 对 USPS 和 MNIST 上的分类准确率的影响



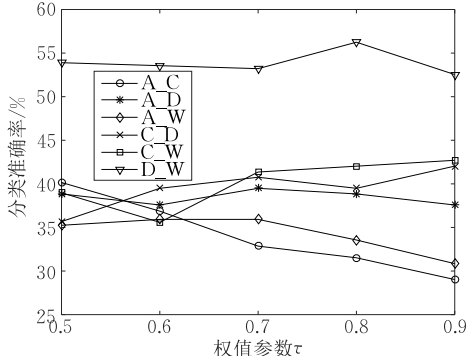
(b) c 对 USPS 和 MNIST 上的分类准确率的影响



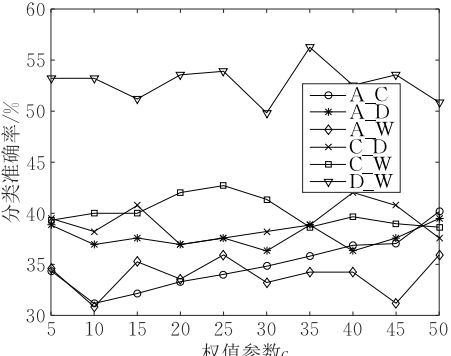
(c) τ 对 MSRC, VOC2007, Sketch 上的分类准确率的影响



(d) c 对 MSRC, VOC2007, Sketch 上的分类准确率的影响



(e) τ 对 Caltech-256 和 Office 上的分类准确率的影响



(f) c 对 Caltech-256 和 Office 上的分类准确率的影响

图 4 权值参数 τ 和 c 对各数据集分类准确率的影响

从图 4 中可以看出, τ 对分类准确率影响相对比较大, 它决定权值的中心, 当 τ 越大, 如 0.9, 则有 90% 的像素点的权值在 0.5 以上, 说明异常点相对比较少, 从图 4(a), (c), (e) 可以看出字符数据集偏向于 τ 越大, 效果越好, 这是由于字符数据集相对于自然图像集异常点少. 参数 c 决定权值在 0~1 之间的波动幅度, 从图 4(b), (d), (f) 可以看出, c 取值适中效果相对比较好, 稳定性更高.

为了检验迭代次数 NT 对 TRSC 表示方法性能的影响, 在 6 个数据集上 3 个分类迁移学习任务上进行对比实验, 实验结果如图 5 所示. 从图 5 可以看出, TRSC 在迭代次数 NT 为 10 时开始收敛, 表明 TSRC 算法的收敛速度快, 同时也验证了 TRSC 中用正则化约束代替字典不等式约束是有效的.

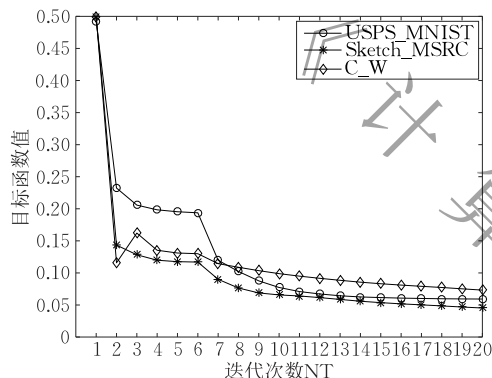
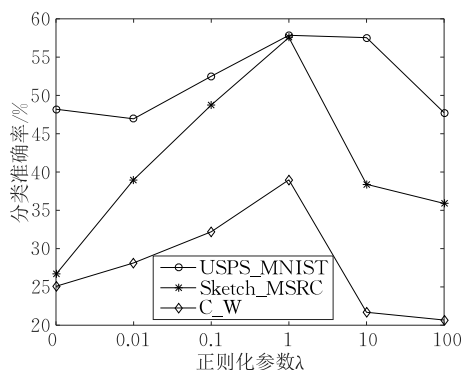


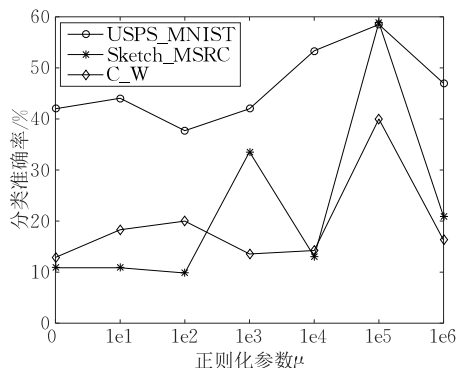
图 5 迭代次数 NT 对目标函数收敛的影响

最后为了分析图正则化参数 γ , 最大均值差异参数 μ 和稀疏正则化参数 λ 对表示方法性能的影响, 分别在 6 个数据集上 3 个分类迁移学习任务上进行对比实验, 实验结果如图 6 所示.

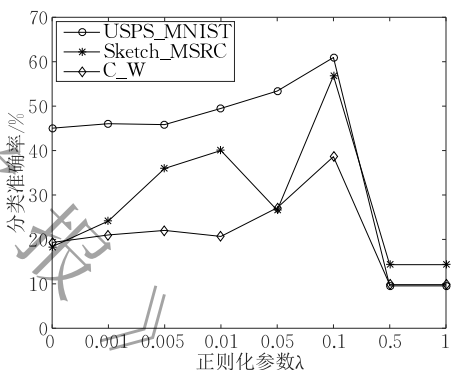
对于图正则化参数 γ , 它的大小决定了新特征空间中数据集原几何特性的保留在最小化目标函数中所占的权重. 从图 6(a) 可以看出, 当 $\gamma \rightarrow 0$ 时, 即几乎忽略数据集的几何特性, 和 $\gamma \rightarrow +\infty$ 时, 即几乎忽略不同域间数据的分布差异和稀疏结构, 都会导致分类准确率下降. 从图 6(b) 可以看出, 当 $\mu \rightarrow 0$ 时, 即几乎忽略不同域间数据差异, 和 $\mu \rightarrow +\infty$ 时, 即几乎忽略数据集的几何特性和稀疏结构, 都会导致分类准确率下降. 从图 6(c) 可以看出, 稀疏正则化参数越小, 对编码稀疏度惩罚越小, 导致其中的非零分量增加, 容易使得模型以更小的残差学习获得图像的近似, 从而异常点也被很好地表示, 丢失了异常点的信息; 当稀疏正则化参数超过一定数值后, 随着稀疏正则化参数的增大, 对编码稀疏度惩罚越大, 导致非零分量减少, 极端的情况就是趋向于全部为



(a) 图正则化参数 γ 对各数据集上的分类准确率的影响



(b) 最大均值差异正则化参数 μ 对各数据集上的分类准确率的影响



(c) 稀疏正则化参数 λ 对各数据集上的分类准确率的影响

图 6 γ, μ, λ 对分类准确率的影响

零, 此时正常点和异常点都不能被表示. 这两种情况都使得模型对异常点信息不敏感, 即正常点和异常点混淆不清, 从而分类准确率大幅度下降. TRSC 是在综合考虑了异常点信息, 不同域间差异, 数据集的几何特性, 在保留数据几何特性和缩小不同数据集的分布差异的同时, 削弱异常点对分类的干扰, 实现了迁移鲁棒稀疏编码和字典的学习.

5 总 结

本文针对迁移学习领域存在的图像异常, 提出了迁移鲁棒稀疏编码模型, 与鲁棒稀疏编码不同的是, 将单个测试样本的模型推广到样本矩阵, 提高了

算法效率,更符合大数据的需求,且提出了鲁棒字典学习,使得模型更加一般化。在问题的求解上,编码的学习仍然采用特征符号搜索法,而字典的学习,通过把不等式约束转化为正则化项,使其成为一个无约束最优化问题,避免了用拉格朗日法求解的复杂性。TRSC 模型在实现领域适配的同时,通过图正则化保持了数据几何特性,稀疏编码提取图像高级语义,权值矩阵削弱异常点对分类产生的影响。从实验中的数字和自然图像两类数据集可以看出,TRSC 算法在异常点较多的数据集上效果更明显,验证了 TRSC 算法在处理异常点时的有效性。

参 考 文 献

- [1] Pan S J, Yang Qiang. A survey on transfer learning. *IEEE Transaction on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359
- [2] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for nature images. *Nature*, 1996, 381(6583): 607-609
- [3] Zhao Zhong-Qiu, Ji Hai-Feng, Gao Jun. Sparse coding based multi-scale spatial latent semantic analysis for image classification. *Chinese Journal of Computers*, 2014, 37(6): 1251-1260(in Chinese)
(赵仲秋, 季海峰, 高隽等. 基于稀疏编码多尺度空间潜在语义分析的图像分类. *计算机学报*, 2014, 37(6): 1251-1260)
- [4] Dash K S, Puan N B, Panda G. A sparse concept code spatio-spectral feature representation for handwritten character recognition//*Proceedings of the International Conference on Signal Processing and Communications*. Bengaluru, India, 2016: 1-5
- [5] Xu Bing-Xin, Yin Qian, Guo Ping, et al. Image representation via sub-dictionary based sparse coding//*Proceedings of the International Joint Conference on Neural Networks*. Vancouver, Canada, 2016: 5113-5118
- [6] Long Ming-Sheng, Ding Gui-Guang, Wang Jian-Min, et al. Transfer sparse coding for robust image representation//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, Oregon, USA, 2013: 407-414
- [7] Lee D D, Seung H S. Algorithms for non-negative matrix factorization//*Proceedings of the Neural Information Processing Systems*. Vancouver, Canada, 2001: 556-562
- [8] Long Ming-Sheng, Wang Jian-Min, Ding Gui-Guang, et al. Transfer learning with graph co-regularization. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(7): 1805-1818
- [9] Yang Liu, Jing Li-Ping, Michael K N. Robust and non-negative collective matrix factorization for text-to-image transfer learning. *IEEE Transactions on Image Processing*, 2015, 24(12): 4701-4714
- [10] You Cong-Zhe, Wu Xiao-Jun, Palade V, et al. Manifold locality constrained low-rank representation and its applications//*Proceedings of the International Joint Conference on Neural Networks*. Vancouver, Canada, 2016: 3264-3217
- [11] Wang Rong, Chen Cai-Kou, Li Jing-Shan, et al. Face recognition via non-negative sparse low-rank representation classification//*Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*. Changsha, China, 2016: 1609-1614
- [12] Lee J, Choe Y. Low rank matrix recovery via augmented lagrange multiplier with non-convex minimization//*Proceedings of the IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop*. Bordeaux, France, 2016: 1-5
- [13] Jhuo I H, Liu Dong, Lee D T, et al. Robust visual domain adaption with low-rank reconstruction//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA, 2012: 2168-2175
- [14] Gao Sheng-Hua, Tsang I W H, Chia L T, et al. Local features are not lonely laplacian sparse coding for image classification//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San-Francisco, USA, 2010: 3555-3561
- [15] Zheng Miao, Bu Jia-Jun, Chen Chun, et al. Graph regularized sparse coding for image representation. *IEEE Transactions on Image Processing*, 2011, 20(5): 1327-1336
- [16] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation. *Journal of Inner Mongolia Agricultural University*, 2010, 31(2): 210-227
- [17] Yang Meng, Zhang Lei, Yang Jian, et al. Robust sparse coding for face recognition//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs, USA, 2011: 625-632
- [18] Mukherjee L, Hall A, et al. Non-Negative sparse coding with regularizer for image classification//*Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. Hawaii, USA, 2015: 852-859
- [19] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms//*Proceedings of the NIPS*. Vancouver, Canada, 2006: 801-808
- [20] Mairal J, Bach F, Ponce J, et al. Online dictionary learning for sparse coding//*Proceedings of the ICML*. Montreal, Canada, 2009: 689-696
- [21] Fletcher R. Practical methods of optimization. RI, USA: A Wiley-Interscience Publication, 1987
- [22] Eitz M, Hays J, Alexa M. How do humans sketch objects? *Acm Transactions on Graphics*, 2012, 31(4): 44-54
- [23] Long Ming-Sheng, Wang Jian-Min, Ding Gui-Guang, et al. Transfer joint matching for unsupervised domain adaption//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Ohio, USA, 2014: 1410-1417
- [24] Long Ming-Sheng, Wang Jian-Min, Sun Jia-Guang, et al. Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(6): 1-1

- [25] Long Ming-Sheng, Wang Jian-Min, Ding Gui-Guang, et al. Transfer feature learning with joint distribution adaptation// Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 2200-2207

- [26] Quanz B, Huan J, Mishra M. Knowledge transfer with low-quality data: A feature extraction issue//Proceedings of the ICDE. Hannover, Germany, 2011, 24(10): 769-779



ZHAO Peng, born in 1976, Ph. D., associate professor. Her research interests include machine learning and intelligent information processing.

WANG Wei, born in 1992, M. S. candidate. His main research interests include transfer learning and sparse coding.

LIU Hui-Ting, born in 1978, Ph. D., associate professor. Her research interests include data mining and machine learning.

Ji Xia, born in 1983, Ph. D., lecturer. Her research interests include data mining and intelligent information processing.

Background

Machine learning has already achieved great success in many research areas including predication, annotation and classification. Traditional machine learning assumes that the training data and the future data follow the same distribution. However, in real world, this assumption often does not hold. The environments and conditions of the data generation are so diverse that the training data and the test data are not drawn from the same distribution. So many traditional machine learning methods can not work well in real-world application.

In contrast, transfer learning allows the distributions in the source domain and the target domain to be different. Transfer learning focuses on how to transfer the knowledge obtained from the source domain to the target domain. The transfer learning based on the feature space expects to find a “good” feature representation space, in which the difference between the source domain and the target domain is reduced. One of the major challenges in real-world transfer learning is how to deal with the outliers. This paper proposes an image

representation method based on transfer robust sparse coding (TRSC). In this method, the weight matrix is used to weaken the interference of outliers to classification. The high level semantic content of the image is obtained with sparse coding. The differences between the source images and the target images are reduced by minimizing the maximum mean discrepancy. The geometrical properties of the image database are preserved by graph Laplacian term. Experimental results on the common transfer learning databases show the effectiveness and robustness of the proposed method.

This paper is supported by the National Natural Science Foundation of China (Grant Nos. 61602004, 61472001), the Natural Science Foundation of the Education Department of Anhui Province (Grant No. KJ2016A041), the Natural Science Foundation of Anhui Province (Grant Nos. 1408085MF122, 1508085MF127) and the Foundation of Center of Information Support & Assurance Technology (Grant Nos. ADXXBZ2014-5, ADXXBZ2014-6).