



智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于迁移学习的无监督跨域人脸表情识别

莫宏伟, 傅智杰

引用本文:

莫宏伟, 傅智杰. 基于迁移学习的无监督跨域人脸表情识别[J]. 智能系统学报, 2021, 16(3): 397–406.

MO Hongwei, FU Zhijie. Unsupervised cross-domain expression recognition based on transfer learning[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(3): 397–406.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202008034>

您可能感兴趣的其他文章

基于深度学习的空间非合作目标特征检测与识别

Feature detection and recognition of spatial noncooperative objects based on deep learning

智能系统学报. 2020, 15(6): 1154–1162 <https://dx.doi.org/10.11992/tis.202006011>

可能性匹配知识迁移原型聚类算法

Possibility-matching based knowledge transfer prototype clustering algorithm

智能系统学报. 2020, 15(5): 978–989 <https://dx.doi.org/10.11992/tis.201810028>

鲁棒的正则化编码随机遮挡表情识别

Recognition of facial expression in case of random shielding based on ro-bust regularized coding

智能系统学报. 2018, 13(2): 261–268 <https://dx.doi.org/10.11992/tis.201609002>

行人重识别研究综述

Survey on pedestrian re-identification research

智能系统学报. 2017, 12(6): 770–780 <https://dx.doi.org/10.11992/tis.201706084>

知识迁移的极大熵聚类算法及其在纹理图像分割中的应用

A maximum entropy clustering algorithm based on knowledge transfer and its application to texture image segmentation

智能系统学报. 2017, 12(2): 179–187 <https://dx.doi.org/10.11992/tis.201603005>

基于最小最大概率机的迁移学习分类算法

Transfer learning classification algorithms based on minimax probability machine

智能系统学报. 2016, 11(1): 84–92 <https://dx.doi.org/10.11992/tis.201505024>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202008034

基于迁移学习的无监督跨域人脸表情识别

莫宏伟, 傅智杰

(哈尔滨工程大学 自动化学院, 黑龙江 哈尔滨 150001)

摘 要: 本文主要研究了基于迁移学习的无监督跨域人脸表情识别。在过去的几年里, 提出的许多方法在人脸表情识别方面取得了令人满意的识别效果。但这些方法通常认为训练和测试数据来自同一个数据集, 因此其具有相同的分布。而在实际应用中, 这一假设通常并不成立, 特别当训练集和测试集来自不同的数据集时, 即跨域人脸表情识别问题。为了解决这一问题, 本文提出将一种基于联合分布对齐的迁移学习方法 (domain align learning) 应用于跨域人脸表情识别, 该方法通过找到一个特征变换, 将源域和目标域数据映射到一个公共子空间中, 在该子空间中联合对齐边缘分布和条件分布来减小域之间的分布差异, 然后对变换后的特征进行训练得到一个域适应分类器来预测目标域样本标签。为了验证提出算法的有效性, 在 CK+、Oulu-CASIA NIR 和 Oulu-CASIA VIS 这 3 个不同的数据库上做了大量实验, 实验结果证明所提算法在跨域表情识别上是有效性的。

关键词: 表情识别; 无监督; 跨域; 迁移学习; 特征变换; 联合对齐; 公共子空间; 域适应

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2021)03-0397-10

中文引用格式: 莫宏伟, 傅智杰. 基于迁移学习的无监督跨域人脸表情识别 [J]. 智能系统学报, 2021, 16(3): 397-406.

英文引用格式: MO Hongwei, FU Zhijie. Unsupervised cross-domain expression recognition based on transfer learning[J]. CAAI transactions on intelligent systems, 2021, 16(3): 397-406.

Unsupervised cross-domain expression recognition based on transfer learning

MO Hongwei, FU Zhijie

(Automation College, Harbin Engineering University, Harbin 150001, China)

Abstract: This paper primarily studies unsupervised cross-domain facial expression recognition based on transfer learning. In recent years, many proposed methods have achieved satisfactory results in facial expression recognition. However, these methods usually assume that the training and test data come from the same data set and therefore have the same distribution. In practical applications, this assumption is usually untrue, especially when the training and test sets come from different data sets, also known as the cross-domain facial expression recognition problem. To solve this problem, we propose a migration learning method (domain align learning) based on joint distributed alignment for cross-domain facial expression recognition. By determining a feature transform, the source and target domain data are mapped onto a common subspace, wherein edge distribution and conditional distribution are aligned jointly to reduce the distribution difference between domains, and then a domain adaptive classifier is obtained by training the transformed features to predict the target domain sample label. To verify the effectiveness of the proposed algorithm, many experiments are performed on three databases, CK+, Oulu-CASIA NIR, and Oulu-CASIA VIS. The experimental results show the effectiveness of the proposed algorithm in cross-domain facial expression recognition.

Keywords: expression recognition; unsupervised; cross-domain; transfer learning; feature transformation; joint alignment; public subspace; domain adaptive

收稿日期: 2020-08-28.

基金项目: 国家自然科学基金项目 (60035117).

通信作者: 莫宏伟. E-mail: honwei2004@126.com.

人脸表情自动识别是情感计算和计算机视觉领域的研究热点之一^[1-5]。近年来, 随着人工智能的迅猛发展, 人脸表情识别在多媒体娱乐、人机

交互、机器智能等领域有着广泛的应用前景。近年来,许多研究者致力于人脸表情识别研究,并提出了许多有效的方法。

Zheng 等^[6]提出了核典型关联分析(kernel canonical correlation analysis, KCCA)的人脸表情识别方法,通过对输入图像中人面部34个关键点进行定位,然后将这些关键点通过Gabor小波变换转换成带有标注信息的图向量(labeled graph, LG)来表示人脸特征。另外,针对每个训练人脸表情图像,将6种表情类型的标签转化成一个六维语义表达向量,进而学习LG向量与语义表达向量之间的相关关系,利用这种关系推断出测试图像的语义表达。在不同人脸表情数据库上的实验结果表明了该方法的有效性。与目标检测等计算机视觉任务相比,图像描述不仅需要检测图像中的内容,还需要理解图像中目标物体之间的具体关系,并使用自然语言正确地表达出来。Uddin 等^[7]提出了一种新的方法来对基于时间序列的视频中的人脸表情进行识别,首先从时间序列人脸中提取局部方向模式特征,然后进行主成分分析和线性判别分析,使特征具有较强的鲁棒性。最后,利用隐马尔可夫模型对不同的面部表情进行局部特征建模和识别,取得了较好的识别效果。Yu^[8]提出了一种基于图像的静态面部表情识别方法,该方法包含一个基于3个最先进的人脸检测器集成的人脸检测模块,以及一个基于多层深度卷积神经网络(convolutional neural network, CNN)级联的分类模块。为了自动决定级联CNN的权重分配问题,提出了通过最小化对数似然损失和最小化铰链损失2种自适应训练卷积神经网络权值的优化方案。Baccouche 等^[9]设计了一种时空卷积稀疏自编码器,在不需要任何监督信息的情况下学习稀疏移位不变表示。

虽然上述人脸表情识别方法取得了较好的识别效果,但需要注意的是,这些方法通常认为训练数据样本和测试数据样本的数据分布相同。而在实际应用中,可能无法满足相同的分布假设。特别是当训练数据和测试数据来自2个不同的数据库,且这2个数据库是在不同的数据收集环境下收集的。对于这样的跨域人脸表情识别问题,传统的人脸表情识别方法可能无法达到很好的识别效果。近年来,迁移学习在图像分类^[10-11]和标注^[12-13]、目标识别^[14-17]和特征学习^[18-20]等方面都有很好的应用前景。且在跨域识别问题上展现出较大的优势,越来越受到研究者的关注。

在跨域人脸表情识别问题中,源域和目标域来自不同的数据库,服从不同的数据分布,因此

需要解决的主要问题就是如何减小不同域之间的分布差异。近几年,一些研究者通过对源域和目标域数据进行权重的重用来减小域之间的分布差异^[21],或者通过找到一个共享的特征表示在保留原始数据属性的同时来减小域之间的分布差异^[22-23]。但是大多数存在的方法仅对齐了边缘分布而忽略了条件分布的重要性,且往往需要目标域中有少量标签样本。

因此,为了解决源域和目标域数据来自2个不同的数据集,服从不同的边缘分布和条件分布,且目标域没有标记数据的无监督跨域表情识别问题,本文提出将联合分布对齐的迁移学习方法应用于跨域表情识别。该方法通过找到一个特征变换,将源域和目标域数据映射到一个公共子空间中,在该子空间中联合对齐边缘分布和条件分布来减小域之间的分布差异,然后对变换后的特征进行训练得到一个域适应分类器来预测目标域样本标签,提高跨域人脸表情识别的准确率。

1 提出方法

1.1 问题描述

令 $X_s \in \mathbf{R}^{m \times n_s}$ 表示源域样本集合, $X_t \in \mathbf{R}^{m \times n_t}$ 表示目标域样本集合, $P_s(X_s)$ 和 $P_t(X_t)$ 分别表示源域数据和目标域数据的边缘分布, $P_s(Y_s|X_s)$ 和 $P_t(Y_t|X_t)$ 分别表示源域和目标域的条件分布。 m 表示源域和目标域中样本的维度, n_s 和 n_t 分别表示源域和目标域样本数据的个数。因此,给定一个带有标签的源域数据 $D_s = \{(x_i, y_i)\}_{i=1}^{n_s}, x_i \in \mathbf{R}^m$, 一个无标签的目标域数据 $D_t = \{(x_j)\}_{j=1}^{n_t}, x_j \in \mathbf{R}^m$, 源域和目标域样本具有相同的特征空间和标签空间,即 $X_s = X_t, Y_s = Y_t$ 。但源域和目标域样本服从不同的数据分布 $P_s(X_s) \neq P_t(X_t), P_s(Y_s|X_s) \neq P_t(Y_t|X_t)$ 。

1.2 核心思想

分布对齐的核心思想是通过找到一个特征变换矩阵,将源域和目标域样本映射到一个公共子空间,即可再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS),通过引入无参数的最大均值差异MMD来度量源域和目标域数据之间边缘分布和条件分布的距离,在该子空间中在最大化投影后源域和目标域数据方差的同时,联合对齐边缘分布和条件分布,最小化域之间的分布距离,即

$$\text{Dist}(D_s, D_t) \approx \text{Dist}(P_s(X_s), P_t(X_t)) + \text{Dist}(P_s(Y_s|X_s), P_t(Y_t|X_t)) \quad (1)$$

然后对变换后的特征进行训练得到一个域适应分类器,来对目标域中的数据标签进行预测。分布对齐原理示意图如图1所示。

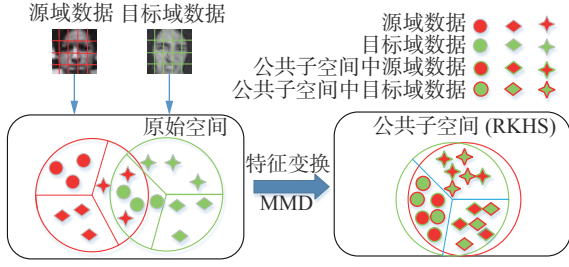


图1 分布对齐原理示意

Fig. 1 Schematic diagram of the distribution alignment

1.3 特征变换

令 $X = [X_s, X_t] \in \mathbb{R}^{m \times (n_s + n_t)}$ 为输入矩阵, n_s 和 n_t 分别为源域和目标域中的样本个数, $H = I - \frac{1}{n_s + n_t} \mathbf{1}\mathbf{1}^T$ 为中心矩阵, 其中 $I \in \mathbb{R}^{(n_s + n_t) \times (n_s + n_t)}$ 为单位矩阵, $\mathbf{1}$ 是一个 $(n_s + n_t) \times (n_s + n_t)$ 的全 1 矩阵, 则输入数据的协方差矩阵可以表示为 XHX^T 。本文的主要目标就是找到一个正交变换矩阵 $W \in \mathbb{R}^{m \times k}$ 使得源域和目标域数据映射到公共子空间中的数据方差最大, 即

$$\max_{W^T W = I} \text{tr}(W^T X H X^T W) \quad (2)$$

1.4 边缘分布对齐

在进行分布对齐之前, 需要解决的主要问题就是找到一个合适的度量准则来计算 2 个域之间的分布差异。而在机器学习中存在很多度量准则可以度量不同分布之间的差异, 例如欧氏距离、余弦相似度、KL 散度等。但这些方法通常计算起来比较复杂。因此为了对齐边缘分布, 减小边缘分布 $P(X_s)$ 和 $P(X_t)$ 之间的差异, 采用无参数的最大均值差异 MMD 来度量不同分布之间的距离。

MMD 的主要目标就是在 RKHS 中计算经过变换后的源域样本和目标域样本均值之间的距离。因此, 边缘分布对齐的目标函数可以写成:

$$\min_{W^T W = I} \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} W^T x_i - \frac{1}{n_t} \sum_{j=1}^{n_t} W^T x_j \right\|_H^2 \quad (3)$$

式中: x_i, x_j 分别为源域和目标域中的第 i 个样本和第 j 个样本; H 表示可再生核希尔伯特空间。

通过矩阵技巧和迹优化, 式 (3) 可以被改写为

$$\min_{W^T W = I} \text{tr}(W^T X L X^T W) \quad (4)$$

其中 L 是 MMD 矩阵, 计算如下:

$$L_{ij} = \begin{cases} \frac{1}{n_s n_s}, & x_i, x_j \in D_s \\ \frac{1}{n_t n_t}, & x_i, x_j \in D_t \\ -\frac{1}{n_s n_t}, & \text{其他} \end{cases} \quad (5)$$

因此, 通过式 (4) 使得域之间的边缘分布 $P(X_s)$ 和 $P(X_t)$ 尽可能地接近, 减小了域之间的边

缘分布差异。

1.5 条件分布对齐

减小边缘分布的差异通常并不能保证域之间的条件分布差异也随之减小。实际上最小化条件分布 $P_s(Y_s|X_s)$ 和 $P_t(Y_t|X_t)$ 之间的差异对于跨域识别问题来说也是至关重要的。然而, 对齐条件分布并不是很容易, 在目标域没有标签数据, 对条件分布 $P_t(Y_t|X_t)$ 直接进行求解是行不通的。因此可以使用类条件分布的充分统计量 $P_t(X_t|Y_t)$ 来近似 $P_t(Y_t|X_t)$ 。近来的一些工作开始通过内核映射空间中的样本选择、联合训练^[24]、循环验证^[25]、核密度估计^[26]等进行条件分布对齐。但是这些方法往往都需要在目标域中有一些标签数据, 所以这些方法并不能解决我们的问题。

为了解决这一问题, 提出利用目标域数据的伪标签, 即通过将在有标签的源域数据上训练的基分类器应用于无标签的目标域数据, 可以很容易地预测目标域数据的伪标签。由于不同域之间的分布差异, 这些伪标签可能不是太准确, 需要在实验中进行迭代优化。在源域带标签数据上训练的基分类器可以是标准的学习器, 例如支持向量机 (SVM)、NN 等。

有了真实标签的源域数据和带有伪标签的目标域数据之后, 就可以在标签空间 Y 的每个类中匹配类条件分布 $P_s(x_s|y_s = c)$ 和 $P_t(x_t|y_t = c)$, $c = \{1, 2, \dots, C\}$ 。为了测量类条件分布 $P_s(x_s|y_s = c)$ 和 $P_t(x_t|y_t = c)$ 之间的距离, 本文对 MMD 进行了修改。因此, 条件分布对齐的目标函数可以写成:

$$\min_{W^T W = I} \left\| \frac{1}{n_s^{(c)}} \sum_{x_i \in D_s^{(c)}} W^T x_i - \frac{1}{n_t^{(c)}} \sum_{x_j \in D_t^{(c)}} W^T x_j \right\|_H^2 \quad (6)$$

式中: $D_s^{(c)} = \{x_i : x_i \in D_s \wedge y(x_i) = c\}$ 是源域数据中属于第 c 类的样本集合; $y(x_i)$ 是数据 x_i 的真实标签; $n_s^{(c)} = |D_s^{(c)}|$ 是源域数据中属于第 c 类的样本个数。对应的, $D_t^{(c)} = \{x_j : x_j \in D_t \wedge \hat{y}(x_j) = c\}$ 是目标域数据中属于第 c 类样本集合, $\hat{y}(x_j)$ 是样本 x_j 的伪标签, $n_t^{(c)} = |D_t^{(c)}|$ 是目标域数据中属于第 c 类样本的个数。

通过进一步优化, 式 (6) 可以简化为

$$\min_{W^T W = I} \text{tr}(W^T X L_c X^T W) \quad (7)$$

式中: L_c 是类标签的 MMD 矩阵, 计算如下:

$$(L_c)_{ij} = \begin{cases} \frac{1}{n_s^{(c)} n_s^{(c)}}, & x_i, x_j \in D_s^{(c)} \\ \frac{1}{n_t^{(c)} n_t^{(c)}}, & x_i, x_j \in D_t^{(c)} \\ -\frac{1}{n_s^{(c)} n_t^{(c)}}, & \begin{cases} x_i \in D_s^{(c)}, x_j \in D_t^{(c)} \\ x_j \in D_s^{(c)}, x_i \in D_t^{(c)} \end{cases} \\ 0, & \text{其他} \end{cases} \quad (8)$$

因此,通过式(7)减小了域之间的条件分布差异。但是在本文要解决的问题中,目标同时是最小化域间的边缘分布和条件分布的差异。因此将式(4)、(7)与式(2)合并到一起得到 DaL 的优化问题,即

$$\min_{W^T X H X^T W = I} \text{tr}(W^T X L X^T W) + \sum_{c=1}^C \text{tr}(W^T X L_c X^T W) + \lambda \|W\|_F^2 \quad (9)$$

式中: λ 是正则化参数; $\|W\|_F^2$ 表示转换矩阵 W 的 Frobenius 范数。

1.6 模型优化

令 $\varphi = \text{diag}(\varphi_1, \varphi_2, \dots, \varphi_k) \in \mathbf{R}^{k \times k}$ 为拉格朗日乘子,根据约束优化理论,可以推导出式(9)的拉格朗日函数,即

$$\text{tr}(W^T X L X^T W) + \sum_{c=1}^C \text{tr}(W^T X L_c X^T W) + \text{tr}((I - W^T X H X^T W)\varphi) \quad (10)$$

令式(10)对变换矩阵 W 求偏导数,并令其等于零可得到其广义特征分解:

$$(X L X^T + X \sum_{c=1}^C L_c X^T + \lambda I) W = X H X^T W \varphi \quad (11)$$

最后,选择式(11)的前 k 个最小特征值对应的特征向量作为变换矩阵 W 的解。算法的伪代码如算法1所示。

算法1 分布对齐算法 (DaL)

输入 数据矩阵 X , 标签矩阵 y_s , 子空间维度 k , 以及正则化参数 λ ;

输出 变换矩阵 W , 域适应分类器 f 。

1) 通过式(5)构建 MMD 矩阵 L , 并令 $\{L_c = 0\}_{c=1}^C$ 。

2) 通过对式(11)进行特征分解选择最小的 k 个特征向量构建变换矩阵 W ;

3) 在 $\{(W^T x_i, y_i)\}_{i=1}^{n_s}$ 上训练一个标准分类器,并用其去更新目标域伪标签 $\{\hat{y}_j = f(W^T x_j)\}_{j=1}^{n_t}$, 通过式(8)构造 MMD 矩阵 L_c 。

4) 重复步骤2)、3)直至收敛;

5) 在标注数据 $\{(W^T x_i, y_i)\}_{i=1}^{n_s}$ 上训练的域适应分类器 f 。

2 实验与分析

2.1 数据库介绍

1) CK+数据库

CK+(the extended cohn-kanda dataset) 数据库^[27] 是美国的匹兹堡大学团队和卡内基梅隆大学团队合作在 CK(cohn-kanade) 数据库^[28] 上进行扩充

而来的。CK+数据库包含 123 名年龄在 18~50 岁来自不同人种的成年人,包括非裔美国人、亚洲以及拉丁美洲人的 593 个图像序列,如图2所示。在 CK 数据库的基础上对视频序列和采集对象进行扩充,分别增加了 22% 和 27%,且在该数据库中的视频序列中包括生气、厌恶、恐惧、高兴、难过、吃惊和中性等 7 种表情,每种表情均有完整的面部编码。该数据库中表情图像的采集对象中有 65% 是女性, 35% 是男性。



图2 CK+数据库部分表情样本

Fig. 2 Partial expression samples of CK+ Database

2) Oulu-CASIA NIR & VIS Database 数据库^[29]

该数据库主要由 Oulu-CASIA NIR(近红外相机)和 Oulu-CASIA VIS(可见光相机)两部分组成,如图3、4所示,均通过对年龄范围为 25~58 岁的 80 位测试者进行表情采集,其中男性和女性的比例分别为 73.8% 和 26.2%。在 80 位测试者中,50 位芬兰测试者由奥卢大学进行采集,剩余的测试者由中国科学院模式识别实验室完成采集整理。整个数据库包含的面部表情为高兴、生气、厌恶、惊讶、恐惧和悲伤,所有这些表情都是在 3 种不同光源条件(正常、强光、黑暗)下,分别通过一个可见光摄像机和一个近红外摄像机获得的。

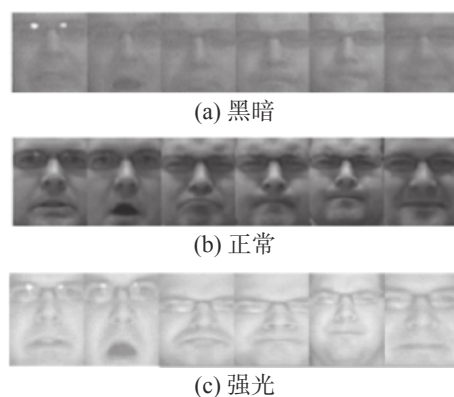


图3 Oulu-CASIA VIS 分别在黑暗、正常、强光条件下数据库中表情样本

Fig. 3 Expression samples of Oulu-CASIA VIS under dark, normal and strong light conditions respectively

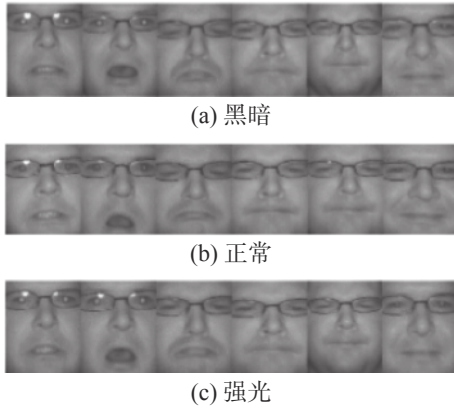


图4 Oulu-CASIA NIR 分别在黑暗、正常、强光条件下数据库中表情样本

Fig. 4 Expression samples of Oulu CASIA NIR under dark, normal and strong light conditions respectively

2.2 实验设置

为了比较,本文选择了近年来在跨域人脸表情识别中常使用的迁移学习算法,包括核均值匹配(kernel mean matching, KMM)^[30]、Kullback-Leibler 重要性估计过程(kullback-leibler importance estimation process, KLIEP)^[31]、选择性迁移机(selective transfer machine, STM)^[5]等与本文提出的分布对齐(DaL)进行对比,这些方法的参数设置将在后面进行介绍。此外,使用线性支持向量机(SVM)作为这4种基准对比方法的分类器,即KLIEP+SVM、KMM+SVM、STM+SVM和DaL+SVM,同时将无迁移的SVM作为基准方法与这4种算法进行对比。

在本文中,采用圆形LBP^[32]特征提取算子来提取表情特征,在实验中选择2种不同的半径和周围像素点的圆形LBP算子,即 $R=1, P=8$ 和 $R=2, P=16$,分别用 LBP_1^8 和 LBP_2^{16} 表示。选择CK+, Oulu-CASIA NIR和Oulu-CASIA VIS这3个数据库进行验证本文提出算法的有效性,在这3个数据库中采集的表情样本来自不同的人种,且是在不同的实验环境下使用不同的设备进行采集的,因此这3个数据库中表情样本具有不同的特征分布。在进行实验之前,首先对这3个数据库中的表情图像进行预处理,将这3个数据库中的表情图像大小进行归一化,统一调整为 64×64 ,并选用这3个数据库中共有的惊讶(surprise)、高兴(happy)、悲伤(sad)、愤怒(angry)、恐惧(fear)和厌恶(disgust)等6种表情类型的样本进行实验,因此在这3个数据库中的表情图像具有相同的特征空间和标签空间。

在实验过程中,每次选取这3个数据库中的任意2个分别作为源域(训练集)和目标域(测试

集),因此可以得到6组对比实验,即

- 1) CK+作为源域, Oulu-CASIA NIR 作为目标域, 用 C&N 表示;
- 2) Oulu-CASIA NIR 作为源域, CK+作为目标域, 用 N&C 表示;
- 3) CK+作为源域, Oulu-CASIA VIS 作为目标域, 用 C&V 表示;
- 4) Oulu-CASIA VIS 作为源域, CK+作为目标域, 用 V&C 表示;
- 5) Oulu-CASIA NIR 作为源域, Oulu-CASIA VIS 作为目标域, 用 N&V 表示;
- 6) Oulu-CASIA VIS 作为源域, Oulu-CASIA NIR 作为目标域, 用 V&N 表示。

在本文实验中所用到的3个数据库中的表情样本构成如表1所示。实验中所有方法采用的参数设置及评价指标如下:在实验中,线性支持向量机(SVM)以固定的 $C=0.2$,且在实验过程中,为了公平对比,所有方法均采用线性核函数。对于KMM来说主要有2个参数 B 和 ε ,分别设置为 $B=1\,000$, $\varepsilon = \sqrt{n} - \frac{1}{\sqrt{n}}$,其中 n 为训练样本的个数。STM中的参数 B 和 ε 设置和KMM一样。在DaL中主要包括2个参数子空间维度 k 和正则化参数 λ ,在实验中设置为 $k=100$, $\lambda=0.1$,迭代次数 $N=20$ 。在本文中,使用测试数据的分类精度作为评价指标,即

$$A = \frac{|x : x \in D_t \wedge \hat{y}(x) = y(x)|}{|x : x \in D_t|} \quad (12)$$

式中: A 表示准确率; D_t 表示目标域样本; $\hat{y}(x)$ 表示各对比算法预测的目标域样本标签; $y(x)$ 表示目标域样本的真实标签。

表1 实验所用数据库样本构成

Table 1 Composition of database samples used in the experiment

实验数据库	愤怒	厌恶	恐惧	高兴	悲伤	吃惊
CK+	45	60	25	70	28	85
Oulu-CASIA NIR	240	240	240	240	240	240
Oulu-CASIA VIS	240	240	240	240	240	240

2.3 实验结果与分析

不同对比方法在本文设置的6种实验场景下的平均准确率分别如表2、3所示。从实验结果可以看出:

- 1) 本文提出的DaL在不同场景下的识别效果相对于无迁移学习的传统机器学习算法SVM有大幅提升,且均高于KMM、KLIEP和STM,表明DaL在跨域人脸表情识别的有效性。

2) 从表2和表3的1、2和3、4四组实验中可以看出,在源域为CK+而目标域分别为Oulu-CASIA VIS和Oulu-CASIA NIR时的识别准确率均低于在源域分别为Oulu-CASIA VIS和Oulu-CASIA NIR而目标域为CK+时的识别准确率,并结合表1可以看出,出现这种情况的原因:1)可能是因为在CK+数据库中的表情样本数量较少;2)可能是因为在CK+数据库中样本数量不均衡造成的。

3) 通过表2和3中5、6两组实验可以看出,在源域为Oulu-CASIA VIS,目标域为Oulu-CASIA NIR上的识别效果要高于在源域为Oulu-CASIA NIR,目标域为Oulu-CASIA VIS上的识别效果,出现这种现象的原因可能是由于Oulu-CASIA VIS和Oulu-CASIA NIR数据库中表情图像使用的采集设备不同,在Oulu-CASIA VIS和Oulu-CASIA NIR数据库中的表情图像分别是通过可见光相机和近红外相机拍摄的,表明模型从由可见光相机拍摄的表情图像上提取的特征要比近红外图像上提取的质量更高。

4) 通过对比表2和表3中6组实验可以看出,通过圆形LBP特征提取方法在 $R=2, P=16$ 时提取特征的识别效果均高于 $R=1, P=8$ 时的识别效果,这一现象表明,在扩大特征提取半径和增加临近像素点时,LBP提取的特征相对较好。

5) 通过对比表2和3中的6组实验可以看出,SVM在5、6两组实验上的识别效果相对于前4组实验来说有了较大提升。这是因为,Oulu-CASIA VIS和Oulu-CASIA NIR数据库中的样本之间的差异,要小于CK+中样本与Oulu-CASIA VIS和Oulu-CASIA NIR之间的差异。在Oulu-CASIA VIS和Oulu-CASIA NIR中样本的采集对象相同,因此样本间的特征差异相对较小。所以,对于无迁移学习的传统机器学习算法SVM来说,在5、6两组上的识别率相对较好。

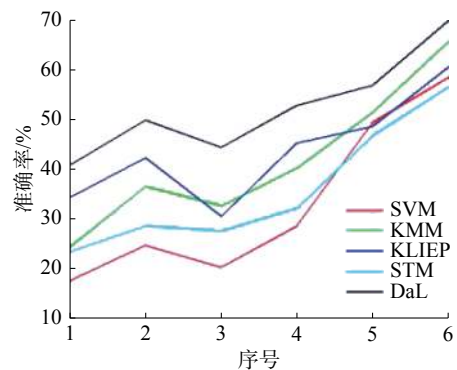
表2 不同方法在6种不同实验场景下的平均准确率(LBP_1^8)

序号	实验场景	SVM	KMM	KLIEP	STM	DaL
1	C&N	17.68	24.35	34.26	23.5	40.76
2	N&C	24.65	36.52	42.3	28.64	49.48
3	C&V	20.32	32.68	30.49	27.56	44.35
4	V&C	28.6	40.18	45.24	32.15	52.74
5	N&V	49.35	51.4	48.54	46.7	56.78
6	V&N	58.4	65.64	60.48	56.5	69.74

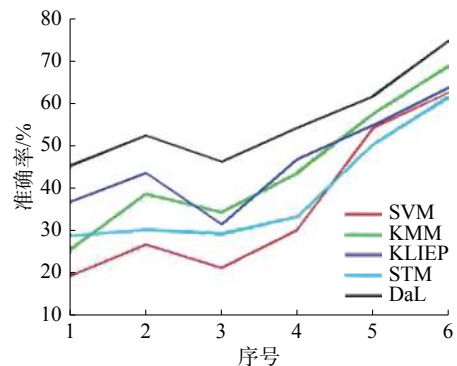
表3 不同方法在6种不同实验场景下的平均准确率(LBP_2^{16})

序号	实验场景	SVM	KMM	KLIEP	STM	DaL
1	C&N	19.32	25.46	36.84	28.76	45.24
2	N&C	26.79	38.64	43.5	30.26	52.38
3	C&V	21.25	34.3	31.5	29.32	46.18
4	V&C	30.28	43.64	46.78	33.26	54.13
5	N&V	54.2	57.5	54.78	50.24	61.64
6	V&N	62.5	68.76	63.54	61.45	74.58

为了更加直观地表示各对比算法的识别效果,给出了这5种算法在不同场景下的平均准确率曲线,如图5所示。从图5中可以清晰看出,本文提出的DaL方法在跨域人脸表情识别上的有效性。



(a) $R=1, P=8$ 时的识别准确率



(b) $R=2, P=16$ 时的识别准确率

图5 各对比算法在不同实验场景下的平均准确率曲线
Fig. 5 Average accuracy curve of each comparison algorithm in different experimental scenarios

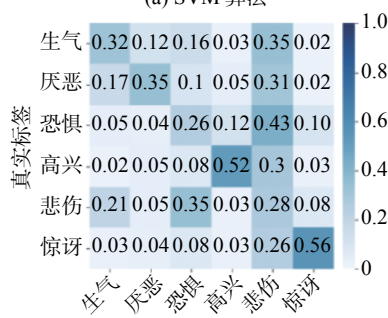
为了进一步验证提出方法的有效性,本文选择表3中序号为3和5的这一组对比实验进行分

析,并给出了在 C&V 和 N&V 两种实验场景下各对比算法在 6 种表情上的识别率混淆矩阵,分别如图 6 和图 7 所示。

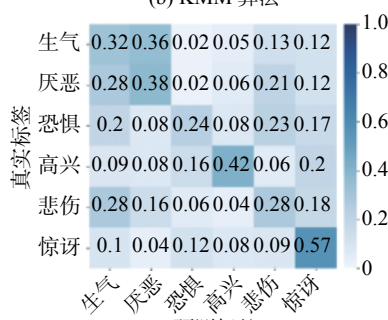
通过图 6 和图 7 可以看出,在 CK+作为源域, Oulu-CASIA VIS 作为目标域时,在 DaL 中各种表情的识别率均高于各基准对比算法,并且不同算法在 6 种表情上的识别率差异相对较大。例如,在 DaL 中“惊讶”、“高兴”这 2 种表情的识别率相对较高,分别为 72%、60%,而“悲伤”和“恐惧”这 2 种表情的识别率较低,分别为 36%、32%。



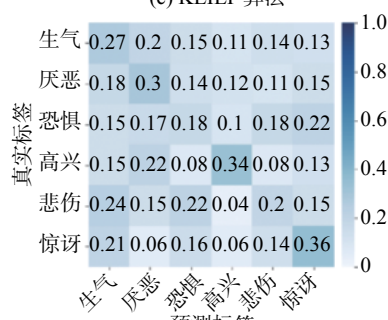
(a) SVM 算法



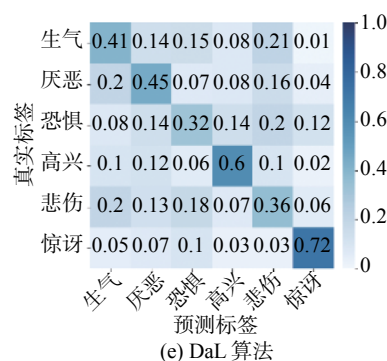
(b) KMM 算法



(c) KLIEP 算法



(d) STM 算法

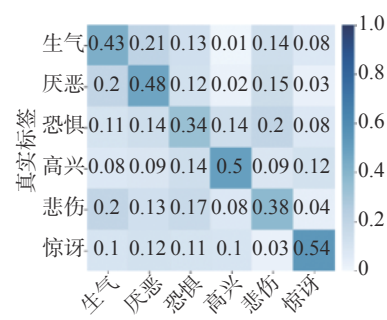


(e) DaL 算法

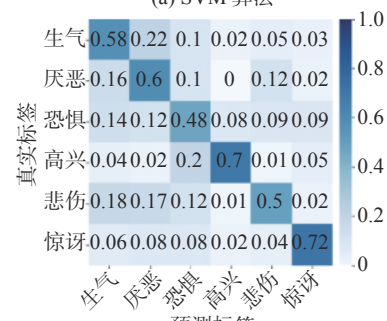
图 6 C&V 实验场景下各对比方法在 6 种表情上的识别率混淆矩阵

Fig. 6 Confusion matrix of the recognition rate of each comparison method on six expressions in the C&V experimental scene

结合表 2 可知,出现这种现象主要是因为 CK+数据库中“吃惊”表情的样本数量最多,有 85 张表情图像,而“恐惧”仅有 25 张表情图像。另外,还可以看到,“愤怒”、“厌恶”和“悲伤”这 3 种表情比较容易误判,结合图 4 中各表情样本,出现这种情况原因可能是因为受试者在表达这 3 种表情时的面部运动变化不大,因此提取的特征较为相似,不易于区分。此外,当源域为 Oulu-CASIA NIR,目标域为 Oulu-CASIA VIS 时,“悲伤”和“恐惧”这 2 种表情的识别准确率有了大幅提升,平均提高了 22% 左右,且其他 4 种表情的识别率也有较大提升。这进一步说明了,数据库中样本数量的多少对跨域人脸表情识别的效果具有一定的影响。但无论 CK+和 Oulu-CASIA NIR 哪个数据库作为源域, DaL 的识别准确率均高于各基准对比算法,验证了该算法在跨域人脸表情识别上的有效性。



(a) SVM 算法



(b) KMM 算法

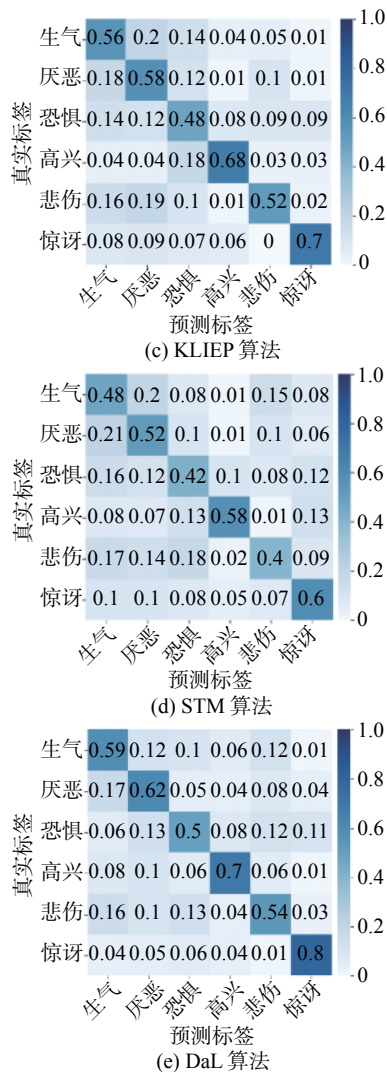


图7 N&V 实验场景下各对比方法在6种表情上的识别率混淆矩阵

Fig. 7 Confusion matrix of the recognition rate of each comparison method on six expressions in the N&V experimental scene

3 结束语

本文提出将一种基于分布对齐的迁移学习方法应用到跨域人脸表情识别中,通过找到一个特征变换矩阵,将源域和目标域样本映射到一个公共子空间,引入无参数的最大均值差异 MMD 来度量源域和目标域数据之间边缘分布和条件分布的距离,在该子空间中最大化投影后源域和目标域数据方差的同时,联合对齐边缘分布和条件分布,最小化域之间的分布距离,然后对迁移后的特征进行训练得到一个域适应分类器,来对目标域中的数据标签进行预测,与4种基准方法在不同实验场景下的实验结果表明,本文提出的算法在跨域人脸表情识别上具有优势。但是还有一些不足之处,需要进一步的研究:1)对于跨域人脸表情识别来说,实验中的数据样本数量的多少

对于实验效果会产生很大的影响。因此,对于如何建立大样本人脸表情数据库将会是下一步需要进行的工作。2)在本文中对条件分布进行对齐时,由于目标域数据中没有标签数据,因此在实验中通过目标域数据的伪标签来进行对齐。但由于域之间的分布差异问题,这种方法预测的伪标签可能不是很准确。因此,对于在条件分布对齐时目标域中无标签数据的问题,需要进一步探索新方法来解决。

参考文献:

- [1] ZHENG Wenming, TANG Hao, HUANG T S. Emotion recognition from non-frontal facial images[M]//KONAR A, CHAKRABORTY A. Emotion Recognition: A Pattern Analysis Approach. Hoboken, USA: John Wiley & Sons, 2015: 183–213.
- [2] WANG Yongjin, GUAN Ling, VENETSANOPOULOS A N. Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition[J]. IEEE transactions on multimedia, 2012, 14(3): 597–607.
- [3] YAN Jingjie, ZHENG Wenming, XU Guanming, et al. Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech[J]. IEEE transactions on multimedia, 2016, 18(7): 1319–1329.
- [4] ZENG Zhihong, PANTIC M, ROISMAN G I, et al. A survey of affect recognition methods: audio, visual, and spontaneous expressions[J]. IEEE transactions on pattern analysis and machine intelligence, 2009, 31(1): 39–58.
- [5] CHU W S, DE LA TORRE F, COHN J F. Selective transfer machine for personalized facial expression analysis[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(3): 529–545.
- [6] ZHENG Wenming, ZHOU Xiaoyan, ZOU Cairong, et al. Facial expression recognition using kernel canonical correlation analysis (KCCA)[J]. IEEE transactions on neural networks, 2006, 17(1): 233–238.
- [7] UDDIN M Z. A local feature-based facial expression recognition system from depth video[M]//DELIGIANNIDIS L, ARABNIA H R. Emerging Trends in Image Processing, Computer Vision and Pattern Recognition. Amsterdam: Elsevier, 2015: 407–419.
- [8] YU Zhiding, ZHANG Cha. Image based static facial expression recognition with multiple deep network learning[C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. New York, USA: ACM, 2015:433–442.

- [9] BACCOUCHE M, MAMALET F, WOLF C, et al. Spatio-temporal convolutional sparse auto-encoder for sequence classification[C]//Proceedings of the of British Machine Vision Conference. Guildford, UK: BMVA Press, 2012:12.
- [10] WANG Hua, NIE Feiping, HUANG Heng, et al. Dyadic transfer learning for cross-domain image classification[C]//IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011:551–556.
- [11] LUO Jie, TOMMASI T, CAPUTO B. Multiclass transfer learning from unconstrained priors[C]//IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011:1863–1870.
- [12] ROY S D, MEI Tao, ZENG Wenjun, et al. SocialTransfer: cross-domain transfer learning from social streams for media applications[C]//Proceedings of the 20th ACM international conference on Multimedia. Nara, Japan: ACM, 2012:649–658.
- [13] WANG Shuhui, JIANG Shuqiang, HUANG Qingming, et al. Multi-feature metric learning with knowledge transfer among semantics and social tagging[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012:2240–2247.
- [14] AYTAR Y, ZISSERMAN A. Tabula rasa: model transfer for object category detection[C]//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011:2252–2259.
- [15] GOPALAN R, LI Ruonan, CHELLAPPA R. Domain adaptation for object recognition: an unsupervised approach[C]//2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011:999–1006.
- [16] GUILLAUMIN M, FERRARI V. Large-scale knowledge transfer for object localization in ImageNet[C]//Proceedings of 2012 IEEE Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012:3202–3209.
- [17] LAMPERT C H, NICKISCH H, HARMELING S. Learning to detect unseen object classes by between-class attribute transfer[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009:951–958.
- [18] JHUO I H, LIU D, LEE D T, et al. Robust visual domain adaptation with low-rank reconstruction[C]//2012 IEEE Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012:2168–2175.
- [19] LAMPERT C H, KRÖMER O. Weakly-paired maximum covariance analysis for multimodal dimensionality reduction and transfer learning[C]//Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece: Springer-Verlag, 2010:566–579.
- [20] QIU Qiang, PATEL V M, TURAGA P, et al. Domain adaptive dictionary learning[C]//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer-Verlag, 2012:631–645.
- [21] DAI Wenyuan, YANG Qiang, XUE Guirong, et al. Boosting for transfer learning[C]//Proceedings of the 24th International Conference on Machine Learning. Corvallis, USA: ACM, 2007:193–200.
- [22] 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.
- [23] GONG Boqing, SHI Yuan, SHA Fei, et al. Geodesic flow kernel for unsupervised domain adaptation[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012:2066–2073.
- [24] ZHONG Erheng, FAN Wei, PENG Jing, et al. Cross domain distribution adaptation via kernel mapping[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France: ACM, 2009.
- [25] BRUZZONE L, MARCONCINI M. Domain adaptation problems: A DASVM classification technique and a circular validation strategy[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 32(5): 770–787.
- [26] QUANZ B, HUAN J, MISHRA M. Knowledge transfer with low-quality data: A feature extraction issue[J]. IEEE transactions on knowledge and data engineering, 2012, 24(10): 1789–1802.
- [27] LUCEY P, COHN J F, KANADE T, et al. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, USA: IEEE, 2010: 94–101.
- [28] KANADE T, COHN J F, TIAN Yingli. Comprehensive database for facial expression analysis[C]//Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble, France: IEEE, 2020: 46–53.
- [29] ZHAO Guoying, HUANG Xiaohua, TAINI M, et al. Facial expression recognition from near-infrared videos[J]. Image and vision computing, 2011, 29(9): 607–619.
- [30] HUANG J, GRETTON A, BORGBARDT K, et al. Cor-

recting sample selection bias by unlabeled data[J]. Advances in neural information processing systems, 2006, 19: 601–608.

- [31] SUGIYAMA M, NAKAJIMA S, KASHIMA H, et al. Direct importance estimation with model selection and its application to covariate shift adaptation[C]//Proceedings of 21st Annual Conference on Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2008: 1433–1440.

- [32] 刘斌, 米强, 徐岩. LBP 和 MB-LBP 加权融合的人脸识别[J]. 计算机工程与设计, 2018, 39(2): 551–556.
LIU Bin, MI Qiang, XU Yan. Face recognition based on weighted fusion of LBP and MB-LBP features[J]. Computer engineering and design, 2018, 39(2): 551–556.

作者简介:



莫宏伟, 教授, 博士生导师, 博士, 主要研究方向为类脑计算与人工智能、机器视觉与机器认知、人机混合智能。主持国家自然科学基金等项目 20 余项。出版专著 6 部, 发表学术论文 80 余篇。



傅智杰, 硕士研究生, 主要研究方向为深度学习、计算机视觉、医学影像。

CAAI 第六届全国大数据与社会计算学术会议

China National Conference on Dig Data and Social Computing

2021 年 8 月 21–22 日, 由中国人工智能学会主办, CAAI 社会计算与社会智能专委会、重庆工商大学承办的“CAAI 第六届全国大数据与社会计算学术会议”(以下简称 BDSC 2021)将在重庆召开。

BDSC 创建于 2012 年, 旨在搭建全国大数据与社会计算学术交流平台, 培育社会计算与社会智能学科发展, 助力社会计算与社会智能领域人才成长, 建设“有仁、有信、有情”的学术共同体, 已成为全国大数据与社会计算领域的知名学术交流品牌。

本次会议的主题为“数字社会的重构与转型”, 面向国家经济社会发展战略, 立足数字社会构建, 通过跨学科交叉视野剖析数字社会的机遇与挑战。会议面向全国开展大数据与社会计算领域的学术征文, 重点探讨数字社会的基础性、前瞻性和战略性理论及其应用, 讨论数字社会领域前沿进展, 交流新的学术思想和新方法, 探索数字社会对人类发展的意义, 展望数字社会未来的发展趋势。

本次大会将组织专家对所有投稿论文进行双盲评审, 优秀论文将有机会被推荐到 11 本高水平期刊进入期刊快速评审通道, 所有录用论文在大会上进行墙报 (poster) 交流。

大会将邀请人工智能、IoT+5G 网络、地理信息以及社会科学、系统科学等交叉领域的重量级学者做大会报告, 举办系统科学的前沿讲习班。大会致力于跨越传统学科分界, 呈现大数据智能时代社会研究新境界, 将重磅发布极具创新性的大规模社会计算试验场和社会计算开放数据集, 并举办全国社会计算大赛。

征文主题:

1) 社会系统建模与仿真; 2) 人工智能与认知科学; 3) 社会网络与群体行为; 4) 社会地理与城市计算; 5) 计算人口新范式、新方法; 6) 数字基础建设与智能社会; 7) 数据价值评估与流通服务; 8) 数字社会与公共安全; 9) 数字政府与公共大数据; 10) 数字技术与社会韧性; 11) 数据与社会治理; 12) 数字平台建设与管理; 13) 计算社会与系统工程; 14) 数据伦理与隐私保护。

重要时间:

投稿截止日期: 2021 年 6 月 15 日

审稿通知日期: 2021 年 7 月 15 日

审稿修改日期: 2021 年 8 月 1 日

录用通知日期: 2021 年 8 月 7 日

会议召开日期: 2021 年 8 月 21–22 日

投稿链接: <https://easychair.org/conferences/?conf=bdsc2021>

大会网站: <http://idke.ruc.edu.cn/BDSC2021/>