

基于路径签名的时间序列领域自适应方法*

蔡瑞初¹, 颜嘉文¹, 陈道鑫¹, 李梓健¹, 郝志峰²

¹(广东工业大学 计算机学院, 广东 广州 510006)

²(汕头大学 工学院, 广东 汕头 515063)

通信作者: 蔡瑞初, E-mail: cairuichu@gmail.com; 郝志峰, E-mail: haozhifeng@stu.edu.cn



摘要: 近年来深度学习因其在各个场景下的优异性能而受到越来越多研究者的重视, 但是这些方法通常依赖独立同分布假设. 领域自适应则是为了缓解分布偏移带来的影响而提出的问题, 它利用带标签的源域数据和不带标签的目标域数据能够训练得到在目标域数据上性能较好的模型. 现有的领域自适应方法大多针对静态数据, 而时间序列数据的方法需要捕捉变量之间的依赖关系. 现有的方法虽然采用针对时间序列数据的特征提取器, 例如递归神经网络, 以学习变量间的依赖关系, 但是往往将冗余的信息也一同提取. 这些冗余信息容易和语义信息耦合, 进而影响模型的预测性能. 基于上述问题, 提出一种基于路径签名的时间序列领域自适应方法 (path-signature-based time-series domain adaptation, PSDA). 该方法一方面利用路径签名变换来捕捉变量间的稀疏依赖关系, 排除冗余相关关系的同时保留语义相关关系, 从而有利于提取时序数据中具有判别力的特征; 另一方面通过约束源域和目标域之间的依赖关系一致性来保留领域之间不变的依赖关系, 排除领域变化的依赖关系, 从而有利于提取时序数据中具有泛化性的特征. 基于以上方法, 进一步提出一个距离度量标准和泛化性边界理论, 并且在多个时间序列领域自适应标准数据集上获得了最好的实验效果.

关键词: 迁移学习; 时间序列领域自适应; 路径签名

中图法分类号: TP18

中文引用格式: 蔡瑞初, 颜嘉文, 陈道鑫, 李梓健, 郝志峰. 基于路径签名的时间序列领域自适应方法. 软件学报, 2025, 36(2): 570-589. <http://www.jos.org.cn/1000-9825/7149.htm>

英文引用格式: Cai RC, Yan JW, Chen DX, Li ZJ, Hao ZF. Time-series Domain Adaptation Based on Path Signature. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 570-589 (in Chinese). <http://www.jos.org.cn/1000-9825/7149.htm>

Time-series Domain Adaptation Based on Path Signature

CAI Rui-Chu¹, YAN Jia-Wen¹, CHEN Dao-Xin¹, LI Zi-Jian¹, HAO Zhi-Feng²

¹(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)

²(College of Engineering, Shantou University, Shantou 515063, China)

Abstract: Recently, deep learning has received increasing attention from researchers due to its excellent performance in various scenarios, but these methods often rely on the independent and identically distribution assumption. Domain adaptation is a problem proposed to mitigate the impact of distribution offset, which uses labeled source domain data and unlabeled target domain data to achieve better performance on target data. Existing methods are devised for static data, while the methods for time series data need to capture the dependencies between variables. Although these methods use feature extractors for time series data, such as recurrent neural networks, to learn the dependencies between variables, they often extract redundant information. This information is easily entangled with semantic information, affecting the model performance. To solve these problems, this study proposes a path-signature-based time-series domain adaptation (PSDA). On the one hand, this method uses path signature transformation to capture sparse dependencies between variables and

* 基金项目: 科技创新 2030—“新一代人工智能”重大项目 (2021ZD0111501); 国家优秀青年科学基金 (62122022); 国家自然科学基金 (61876043, 61976052, 62206064)

收稿时间: 2023-06-28; 修改时间: 2023-08-15, 2023-11-22; 采用时间: 2024-01-08; jos 在线出版时间: 2024-07-03

CNKI 网络首发时间: 2024-07-04

eliminate redundant correlations while preserving semantic dependencies, thereby facilitating the extraction of discriminative features from temporal data. On the other hand, the invariant dependency relationships are preserved by constraining the consistency of dependency relationships among different domains, and the changing dependency relationships between domains are excluded, which is conducive to extracting generalized features from temporal data. Based on the above methods, the study further proposes a distance metric and generalized boundary theory and obtains the best experimental results on multiple time series domain adaptation standard datasets.

Key words: transfer learning; time-series domain adaptation; path signature

由于训练集和测试集数据分布的不同, 研究者提出了无监督领域自适应^[1-8]以缓解分布偏移. 无监督领域自适应在无目标域标签的情况下, 通过寻找源域与目标域的域不变信息, 建立从源域到目标域的知识转移, 以达到高泛化性的目的. 目前已经有许多工作通过最小化源域与目标域之间的距离以达到领域自适应的目的. 主要有两类方法: 基于最大平均差异 (maximum mean discrepancy, MMD)^[9]的无监督领域自适应方法^[10,11]和基于对抗学习 (adversarial learning) 的无监督领域自适应方法^[12,13]. 基于最大平均差异是利用源域和目标域之间的均值差异, 通过最小化源域和目标域特征分布的最大距离, 提取领域不变信息. 基于对抗学习的方法则是借鉴了生成对抗网络的思想, 设计了梯度反转层, 从而提取领域不变表示. 实际应用场景中, 通常有不同类型的数据, 如图像数据、网络数据和时间序列数据等. 其中时间序列数据的领域自适应还利用了底层动态信息, 对数据进行分析和研究, 预测序列的未来趋势. 而目前对于时间序列数据的领域自适应研究远少于非时间序列数据.

近年来, 针对时间序列数据的领域自适应方法备受学者关注. 一种最直接的方法是将非时间序列领域自适应方法扩展到时间序列场景, 例如将时间序列特征提取网络 (如长短时记忆网络^[14]、循环神经网络^[15]和傅里叶变换等) 直接应用于经典的非时间序列领域自适应方法中. 其中 Da Costa 等人^[16]使用递归神经网络 RNN 作为特征提取器与经典基于对抗学习的领域自适应方法 DANN 结合寻找领域不变表示, Liu 等人^[17]利用对抗性光谱核匹配与经典 MMD 度量来解决时间序列数据中的非平稳性和非单调性问题. 随后, 学者们发现时间序列数据包含着潜在的时间动态信息, 可以利用这些动态信息在源域和目标域之间进行对齐^[18-20]. Wu 等人^[19]将时间序列信息解耦为周期项和趋势项, 认为数据内部蕴含的长期趋势信息更有利于预测工作, 而周期信息更有利于发现序列的自相关, He 等人^[18]使用傅里叶变换并借助编码器与解码器的结构范式寻找时域和频域上的领域不变表示. 这些方法的目标都是学习变量之间不变的依赖.

虽然以上方法利用了现有的时序建模方法来提取变量之间的依赖并且取得了良好的性能, 但是他们提取出来的依赖是冗余的, 而冗余依赖容易导致模型提取的表征丧失判别力. 图 1(a) 为针对静态数据领域自适应方法 DANN 提取的特征, 数据特征并未按标签聚集成簇 (不同类别的特征容易重叠), 这是因为记忆网络 RNN 提取的特征包含冗余的依赖信息, 导致特征没有分辨力. 传统针对静态数据的领域自适应方法之所以在时序数据上性能下降是因为他们通常采用协变量偏移假设^[21] (即假设不同领域的边缘分布不同 $P_S(x) \neq P_T(x)$, 但条件分布在领域间保持稳定 $P_S(y|x) = P_T(y|x)$), 但由于时间序列数据变量间存在复杂的依赖关系, 导致协变量偏移的假设难以满足. 例如对于多变量时间序列的生成过程 $x_t^1 = (x_{t-1}^1)^2$, $x_t^2 = 2(x_{t-1}^1)^3 + 3x_{t-1}^2$. 给定 $x_0^1 = 2, x_0^2 = 1$, 则有 $x_3^2 = 8747$; 如果 $x_0^1 = 2.1, x_0^2 = 1$, 则有 $x_3^2 \approx 15419.95$. 可以看出, x_0^1 的微小变化会导致未来时间步变量 x_3^2 的巨大变化. 在这个场景下, 协变量偏移假设是不成立的, 因为 $P(x_3^2 | x_0^1 = 2, x_0^2 = 1) \neq P(x_3^2 | x_0^1 = 2.1, x_0^2 = 1)$. 为了解决以上问题, 本文将变量之间的依赖考虑进模型中. 如图 1(b) 所示, 本文发现通过在模型中引入领域不变的稀疏依赖关系, 可以提取出具有分辨力的特征 (如图 1(c) 所示).

为了提取到时间序列数据中的不变依赖关系并将其应用于时间序列领域自适应问题中, 本文提出了一种基于路径签名的时间序列领域自适应方法 (path-signature-based time-series domain adaptation, PSDA). 首先, 本文发现利用路径签名变换的平移不变性和重参数化不变性可以有效地排除不同领域时序数据中由时延和偏置等因素引起的影响, 从而可以更好地提取到时序特征. 基于这些特征, 本文利用自注意力机制学习变量之间的依赖关系. 其次, 本方法通过约束源域和目标域之间的依赖关系一致性来保留领域之间不变的依赖关系, 排除领域变化的依赖关系, 从而为分类器提供领域不变的特征. 基于以上方法, 本文进一步提出了一个距离度量标准和泛化性边界理论, 该理论发现可以用时间序列数据的相关结构相似性来分析算法泛化误差. 本方法在多个真实数据集上验证 (UCI HAR, WISDM 和 PPG), 相比现有针对静态数据和时序数据的对比方法, 本文提出的方法体现出较大优势.

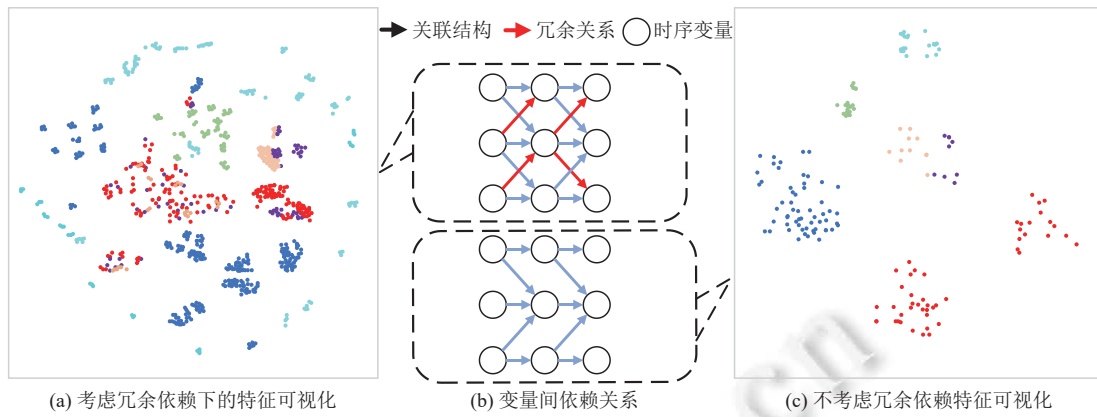


图1 考虑变量依赖和没有考虑变量依赖的特征可视化示意图

1 相关工作

1.1 非时间序列领域自适应

无监督领域自适应 (unsupervised domain adaptation, UDA) 的目的是将所学习的知识从有标记的源域转移到未标记的目标域。除了对领域进行预训练和微调, 现有的大部分工作都是基于在诸如最大平均差异 (MMD)^[9]等限制条件的帮助下从原始数据中学习数据的域不变表示, Tzeng 等人^[22]提出深度领域混淆方法最小化源域和目标域之间的 MMD 距离. Sun 等人^[23]与 Chen 等人^[24]在此基础上分别对齐源域和目标域的二阶及高阶统计量. 另一些无监督领域自适应方法是通过借用生成对抗网络的思想来提取领域不变表示, Ganin 等人^[25]提出利用梯度反转层对特征提取器网络进行反向训练以欺骗域分类器, 通过对抗训练提取域不变表示, 随后也产生了一系列的工作遵循这种对抗性的训练范式^[26-29]. Xie 等人^[30]提出保留语义信息在领域自适应工作中的重要性, 借助了伪标签对齐源域和目标域中各类的质心, 解决了无监督域自适应错对齐的问题. Chen 等人^[31]发现分类对特征尺度有鲁棒性, 而回归没有. 而调整深度表示的分布将改变特征尺度, 阻碍域自适应回归. 遂提出通过不存在特征缩放的表示空间的正交基来缩小领域距离. Zhu 等人^[32]通过对齐相关子域分布的局部最大平均差异 (LMMD) 来最小化源域和目标域之间的差异.

如果考虑变量间的因果关系, 则领域自适应可以由因果机制来进行指导. Zhang 等人^[21]讨论了领域自适应中目标位移、条件位移和广义目标位移这 3 种情况. Cai 等人^[33]应用遵循数据生成过程的因果模型, 通过在恢复的隐空间上提取解耦的语义表示来解决领域自适应问题. 最近, Cai 等人^[34]认为变量间的稀疏关联结构在不同域间是稳定的, 并提出了用于自适应时间序列分类和回归任务的稀疏关联结构对齐方法. Li 等人^[35]提出了因果机制转移网络 (CMTN), 捕获和转移动态和时间因果机制, 以缓解不同领域之间的时间滞后和不同值范围的影响.

1.2 时间序列领域自适应

大部分现有的领域自适应方法都应用于图像领域^[36,37], 然而现实生活中有许多领域自适应问题需要考虑时间信息, 因而出现了时间序列领域自适应. 由于很难在所有时间步中对齐观察到的数据的条件分布, 在图像领域中表现很好的领域自适应方法并不能很好的直接应用到时间序列数据集中, 因此时间序列领域自适应 (time-series unsupervised domain adaptation) 获得越来越多研究者的关注. Da Costa 等人^[16]采用了最直接的方法, 简单地将特征提取器替换为基于 RNN 的特征提取器来提取时间序列数据的表示. Liu 等人^[17]提出利用基于傅里叶变换的对抗性光谱核匹配来解决时间序列数据中的非平稳性和非单调性问题. He 等人^[18]利用傅里叶变换, 通过考虑时间和频率特征, 将其跨域对齐. Cai 等人^[34]提出利用因果关系的稳定性探索数据的领域不变结构, 并在领域间进行对齐. Wu 等人^[19]将时间序列信息区分为周期项和趋势项, 交替进行预测结果优化和序列分解, 逐步将两种信息分离. Jin 等人^[38]引入领域不变的潜在特征 (查询和键) 和重新训练领域特定的特征 (值), 以实现对源域和目标域的预测器的联合训练. Ragab 等人^[39]采用自监督预训练方法和伪标签考虑了特征学习和域对齐过程中数据的时间动态.

Li 等人^[40]利用因果结构的稳定性提出因果条件偏移假设, 重构数据间的格兰杰因果关系, 并基于格兰杰因果结构进行跨领域的时间序列数据的预测工作.

1.3 路径签名方法

近年来, 路径签名作为一个新兴的数据变换方法^[41]受到机器学习的广泛应用, Yang 等人^[42]将路径签名应用于视觉领域, 从姿态空间和时间空间同时进行应用, 提取视频特征. Glad 等人^[43]利用路径签名方法及其相关性质进行因果联合动力系统的分析, 提出了一种新颖的因果发现方法. Morrill 等人^[44]提出了一种通用的广义签名方法, Bonnier 等人^[45]提出了一种结合了签名变换与现代深度学习框架的优势的新方法. 通过在签名变换之前学习流的增强, 可以以依赖于数据的方式选择签名的项, 使路径签名操作本质上作为神经网络中的一个层. De Curtò 等人^[46]使用路径签名和对数签名利用谐波分析工具, 并将其应用到评分函数中指导模型的学习. 赵艺等人^[47]引入路径签名预处理来增强数据, 解决时空图卷积网络中不易学习语义信息和描述时间信息方面存在不足的问题. 目前仍没有工作将路径签名方法应用于时间序列的领域自适应工作中.

2 路径签名变换的背景知识

本文采用路径签名变换方法来提取变量的时序特征, 为了更好地让读者理解本方法, 本文先简单介绍路径签名变换的基础知识. 按顺序排序的数据可以被看作一个自然的路径结构, 例如特定位置的气压变化可以被认为是 \mathbb{R} 的路径; 纸上的笔的运动可以被认为是 \mathbb{R}^2 的路径, 路径签名便是一种描述这种顺序结构的数学工具. 路径签名方法的产生来自粗糙路径理论^[48,49], 是将多元时间序列解释为潜在连续路径的离散化, 可以应用它产生一个已知的实值特征向量. 具体来说, 路径签名是无限长的沿维度组合进行积分的集合, 每一个积分项都是按相应维度组合的顺序进行积分的.

令 $x = \{x_{t-N+1}, \dots, x_{t-1}, x_t\} \in \mathbb{R}^{N \times M}$ 表示一个时间步长为 N 的, 时间跨度为 $[t-N+1, t]$ 的, M 维多元时间序列样本数据, 也称路径, 其中 $x_t \in \mathbb{R}^M$, 相应的标签 $y \in \mathbb{R}$. 本文进一步令多维时间序列 x 的路径签名为各阶迭代积分的集合, 如公式 (1) 所示. 序列 x 的路径签名由 1 和各个阶次的签名项 $S^k(x)$ 组成, 其中 1 代表零阶签名项, 各阶次的签名项由路径按指定维度顺序进行迭代积分组成.

$$\text{Sig}(x) = \{1, S^1(x), S^2(x), \dots, S^k(x), \dots\} \quad (1)$$

其中, $S^k(x)$ 表示序列 x 的第 k 阶路径签名, 1 表示第零阶路径签名. 由于路径签名是一个无限长的序列, 一般地, 学者们在实践中定义一个截断数 μ , 然后在 μ 阶范围内应用路径签名方法. 第 k 阶的路径签名序列 $S^k(x)$ 通常有 M^k 项, 由路径 x 按指定维度顺序进行迭代积分组成, 形式化如公式 (2) 所示:

$$S^k(x) = \left(\left\{ S'(x)^{(i_1, \dots, i_k)} \right\}, (i_1, \dots, i_k) \in [0, M-1] \right) \quad (2)$$

其中, $S'(x)^{(i_1, \dots, i_k)}$ 表示按 i_1, \dots, i_k 的维度顺序进行积分的单个签名项, 形式化如公式 (3) 所示:

$$S'(x)^{(i_1, \dots, i_k)} = \int \dots \int_{t-N+1 \leq u_1 < \dots < u_k \leq t} dx_{u_1}^{i_1} \dots dx_{u_k}^{i_k} \in \mathbb{R} \quad (3)$$

本文通过如图 2 的例子说明路径签名可以更有效地提取时序信息. 相比现有的基于神经网络的方法, 路径签名变换方法具有以下几个优点.

1) 首先, 路径签名变换具有唯一性, 即:

$$S^\mu(x_1) = \text{Sig}^\mu(x_2), \mu \rightarrow \infty, \text{当且仅当 } x_1 = x_2 \quad (4)$$

相比其他基于神经网络的特征提取, 不同的时间序列数据可能提取到相似的特征而丧失判别力, 基于路径签名的方法可以充分保留时序信息.

2) 其次, 路径签名变换具有平移不变性, 即:

$$\text{Sig}(g(t) + c_1) = \text{Sig}(g(t) + c_2), c_1, c_2 \in \mathbb{R} \quad (5)$$

其中, $g(t)$ 是时间序列数据生成函数, c_1, c_2 是振幅常数. 公式 (5) 表示, 同一个生成函数, 由领域引起的不同振幅变

化信息可以由路径签名变换所排除. 如图 2 所示, 虽然不同人群心率变化范围 (c_1, c_2) 不一样, 但是路径签名变换可以排除这种变化的干扰, 仅提取出和心率机制相关的信息.

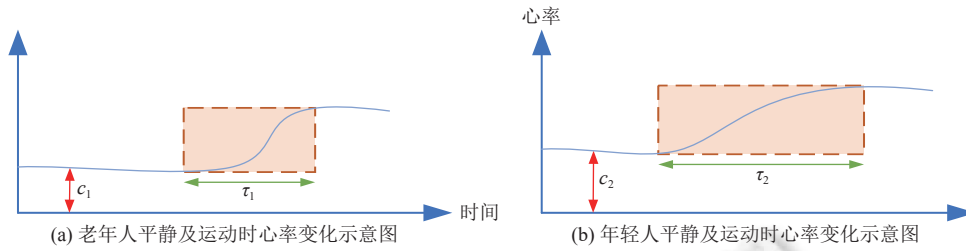


图 2 不同人群的心率变化示意图

3) 第三, 路径签名具有重参数化不变性:

$$Sig(g(t)) = Sig(g(\phi(t))) = Sig(x_u), u = \varphi(t) \tag{6}$$

其中, $\varphi(\cdot)$ 为对 t 变换的重参数化函数, 例如对时间序列的缩放. 公式 (6) 表示, 同一个生成函数, 即使受到由领域引起的不同程度的缩放, 路径签名变换依然可以有效提取领域不变信息. 如图 2 所示, 不同人群心率变化速度不同 (τ_1, τ_2), 但是路径签名变换可以排除这种变化的干扰, 仅提取出和心率机制相关的信息.

路径签名^[45]是时间序列建模方法的推广, 路径签名方法的产生来自粗糙路径理论^[48,49], 是将多元时间序列解释为潜在连续路径的离散化, 从而获得产生一个已知的实值特征向量, 并且用于时间序列数据的下游任务^[42-44]. 本文提出的方法将路径签名用于时间序列数据特征学习和相关结构提取, 并解决了时间序列数据的领域自适应问题.

3 基于路径签名的时间序列领域自适应方法

3.1 问题定义

给定源域分布 $P_S(x, y)$ 与目标域分布 $P_T(x, y)$, 且 $P_S(x, y) \neq P_T(x, y)$, 但是不同领域变量之间的依赖是相同的. 在每一个领域中, 有 $x = \{x_{t-N+1}, \dots, x_{t-1}, x_t\} \in \mathbb{R}^{N \times M}$ 表示一个具有 N 个时间步长的 M 维多元时间序列样本数据, 其中 $x_t \in \mathbb{R}^M$, 相应的标签 $y \in \mathbb{R}$. 在无监督领域自适应的限制下, 假设每一个源域样本 x_S 都带有标签 y_S , 而目标域样本 x_T 没有标签. 本文的目标是设计一个预测模型, 使得这个模型可以从目标域预测给定的时间序列样本 x_T 的标签 \hat{y}_T . 本文涉及的符号及其含义如表 1 所示.

表 1 符号和对应描述

符号	含义
x, y	多维时间序列变量和对应的标签
M	多元时间序列的维度
z, z_a, z_b, f, U, H	特征向量
S, \mathcal{T}	源域, 目标域
seg	片段
T	矩阵的转置
e	相关度权重
α, β	变量内, 变量间的依赖矩阵
$L(\cdot)$	损失函数
$ \cdot $	绝对值
$avg(\cdot)$	求平均
\cdot	矩阵乘法 (叉乘)
$\ \cdot\ _n$	n 范式
$[\cdot:\cdot]$	拼接操作
\sum	加和操作

为了解决以上问题, 本文提出了基于路径签名的时间序列领域自适应方法 (path-signature-based time-series domain adaptation, PSDA). 如图 3 所示, 该模型分为 3 个组成部分: 1) 基于路径签名方法的特征提取器对每一个变量的特征提取; 2) 然后通过注意力机制进行稀疏依赖学习; 3) 通过对齐源域和目标域稀疏依赖进行知识迁移.

3.2 模型概览

本文提出基于路径签名的时间序列领域自适应方法 (PSDA), 该模型包含 3 个步骤如图 3 所示: 1) 首先, 本文使用基于路径签名方法的特征提取器, 通过对每一个变量进行特征提取, 排除不同领域对时序数据的影响, 提取到更好的时序特征. 2) 其次, 本文通过注意力机制进行稀疏依赖学习, 其中使用基于自注意力机制的特征聚合方法进行变量内的相关依赖学习、使用基于注意力机制的相关依赖学习方法进行变量间的相关依赖学习. 3) 最后, 通过以上依赖约束方法, 本文对齐源域和目标域的稀疏依赖结构, 从而进行知识从源域到目标域的迁移.

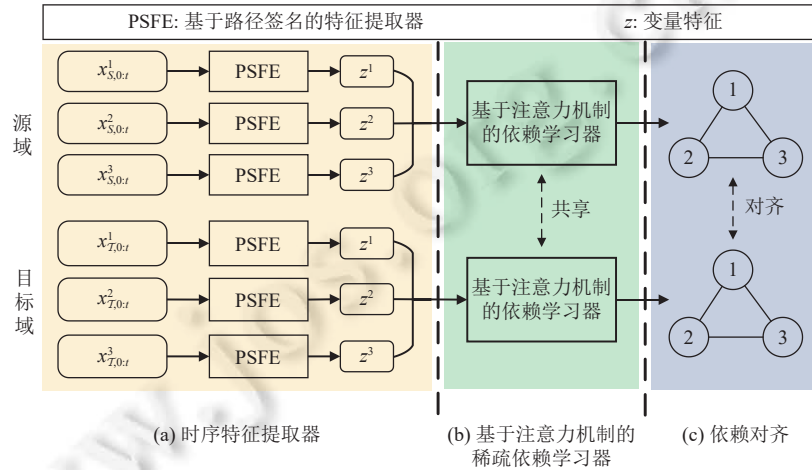


图 3 基于路径签名的时间序列领域自适应方法

3.3 基于路径签名的特征提取

以第 1 维度时间序列数据 $x_{0:t}^1$ 为例的基于路径签名的特征提取器如图 4 所示, 其中包括 4 个步骤, 分别是: (1) 序列分段; (2) 特征提取; (3) 滑动窗口切片; (4) 路径签名变换. 每一步具体实现如下.

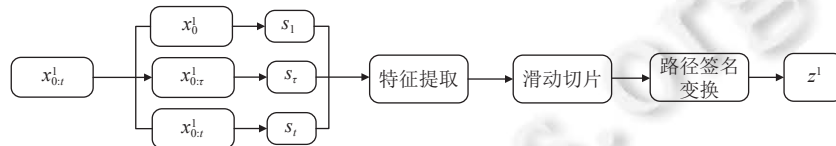


图 4 以 $x_{0:t}^1$ 为例, 基于路径签名的特征提取器示意图

(1) 序列分段: 为了提取出每一维时序数据的有效影响范围, 对于第 j 维, 时间从 $t-N+1$ 到 t (为方便表述, 本文缩写为 $t-N+1:t$), 长度为 N 的原始单变量序列 $x_{t-N+1:t}^j = \{x_{t-N+1}^j, \dots, x_{t-1}^j, x_t^j\}$, 本模型首先进行分段操作, 将时间序列数据分成多个片段 (Segment), 如公式 (7)、公式 (8) 所示:

$$seg_\tau^j = x_{t-\tau:t}^j \tag{7}$$

$$seg^j = \{s_{N-1}^j, \dots, s_1^j, s_0^j\} \tag{8}$$

(2) 特征提取: 对每一个片段进行特征提取, 以便更准确地提取时序信息, 本文采用卷积神经网络对每一个片段进行特征处理, 如公式 (9) 所示:

$$(z_a)_\tau^j = \text{CNN}(s_\tau^j, \theta_C^j) \tag{9}$$

其中, z_a 表示 CNN 提取后的特征, CNN 表示卷积神经网络, θ_c^i 表示卷积神经网络可训练参数, τ 表示该片段的段长度.

(3) 滑动切片: 为了更好地学习单变量序列的表征, 本模型在不同的子段上应用滑动窗口进行切片, 以便可以在序列的多个局部应用路径签名方法, 以提取特定时间间隔内的信息, 如公式 (10) 所示:

$$(z_b)_\tau^j = \left[\left((z_a)_{0:d}^j \right)^T, \left((z_a)_{1:d+1}^j \right)^T, \dots, \left((z_a)_{\tau-d:\tau}^j \right)^T \right] \quad (10)$$

其中, z_b 表示 z_a 的集合, d 为切片长度, T 为转置操作.

(4) 路径签名变换: 最后本文对子段的不同切片进行子段间依赖的路径签名变换特征提取, 即将切片并拼接后得到的表征进行路径签名变换, 得到段表征 z_τ^j . 形式化如公式 (11) 所示:

$$z_\tau^j = \text{Sig} \left((z_b)_\tau^j \right) \quad (11)$$

3.4 基于注意力机制的相关依赖学习

本文所提出的基于注意力机制的相关依赖学习分为两步, 分别是变量内的相关依赖学习和变量间的相关依赖学习, 在第 3.4.1 节中本文将介绍变量内的相关依赖学习, 在第 3.4.2 节中本文将介绍变量间的相关依赖学习.

3.4.1 变量内的相关依赖学习

为了学习变量内的相关依赖, 本文使用基于自注意力机制的特征聚合方法, 计算变量序列 x^j 的每个段表征 z_τ^j 的权重如公式 (12)、公式 (13) 所示:

$$Q^j = z^j \cdot W^Q, K^j = z^j \cdot W^K, V^j = z^j \cdot W^V \quad (12)$$

$$\alpha^j = \text{sparsemax} \left(\frac{1}{N} \sum_{k=1}^N \left(\frac{Q^j K^j T}{\sqrt{d_k}} \right) \right) \quad (13)$$

其中, Q^j, K^j, V^j 为第 j 个变量的段表征 z_τ^j 生成的用于自注意力机制中的查询 (query, Q)、键 (key, K)、值 (value, V)^[50], W^Q, W^K 为可训练的映射参数, $\sqrt{d_k}$ 为缩放因子, α^j 为序列 x^j 的每个段表征 z_τ^j 的权重矩阵, N 为序列的分段数, sparsemax 为稀疏操作^[51-54], 具体形式化为 $\text{sparsemax}(z) = \text{argmin}_{p \in \Delta^{K-1}} \|p - z\|_2$, 作用是将向量 $z \in \mathbb{R}^K$ 欧氏投影到概率单纯形 Δ^{K-1} 上. 由此可得到变量序列 x^j 的加权段表征如公式 (14):

$$f^j = \sum_{\tau=1}^N \alpha_\tau^j \cdot (z_\tau^j W^V) \quad (14)$$

其中, W^V 为可训练的映射参数, f^j 为基于变量内依赖的聚合表达, 表示变量 j 的特征, 将用于后续依赖结构学习及标签预测, z_τ^j 为第 j 维变量分段长度为 τ 的段表征, α_τ^j 为相应权重. 同时本文考虑给定变量段的时间长度随不同领域而变化的情况. 在这种情况下, 为了更精确地重构依赖结构, 本模型最小化源域与目标域的权重结构 α_S^m 和 α_T^m 之间的最大平均差异 (MMD)^[9]. 该对齐操作约束了不同领域同一个变量的内部依赖是一致的, 如公式 (15) 所示:

$$L_\alpha = \sum_{m=1}^M |\text{avg}(\alpha_S^m) - \text{avg}(\alpha_T^m)| \quad (15)$$

其中, α_S^m 和 α_T^m 表示从源域和目标域计算的每个变量内不同细粒段之间的权重, avg 表示平均值操作. $\alpha_S^m / \alpha_T^m \in \mathbb{R}^{N \times N}$ 表示是源域/目标域第 m 个变量内不同分段之间 self-attention 的权重. M 表示时间序列的维度. m 表示源域和目标域中存在同一变量, 是指变量一一对应.

3.4.2 变量间的相关依赖学习

基于第 3.4.1 节, 每个变量都获得分段序列表征的聚合表达, 对于变量间的依赖结构学习, 本文同样采用基于注意力机制的相关依赖学习方法. 现有基于注意力机制的方法通常直接计算变量间的相关度, 然后在领域间进行对齐. 给定 M 个变量 x^1, x^2, \dots, x^M , 由第 3.4.1 节可以获得的变量表示 f^1, f^2, \dots, f^M , 其中 M 为变量个数. 假设计算变量 i 和变量 j 的相关度的方法为余弦相似度, 如公式 (16) 所示:

$$e^{ij} = \cos(f^i, f^j) = \frac{f^i \cdot f^j}{|f^i| \cdot |f^j|} \quad (16)$$

其中, f^j 为第 j 维变量序列 x^j 的聚合表达, 但这种方法忽略了变量之间存在的时间步信息, 学习到的依赖结构可

能因此丢失一定程度的语义信息, 可能导致结构学习不准确, 不利于领域间的对齐. 为了考虑变量间的时间步信息, 给定第 i 个变量 f^i , 本文将其与其他 $M-1$ 个变量的所有段表征计算出稠密的相关结构. 需要学习变量 i 与变量 j 的不同段之间的相关度, 例如, 第 i 个变量和第 j 个变量之间的相关度可以通过公式 (17)、公式 (18) 计算得出:

$$e_{\tau}^{ij} = \frac{f^i \cdot z_{\tau}^j}{|f^i| \cdot |z_{\tau}^j|} \quad (17)$$

$$e^{ij} = \{e_1^{ij}, \dots, e_{\tau}^{ij}, e_N^{ij}\} \quad (18)$$

其中, e_{τ}^{ij} 表示变量 i 与变量 j 的长度为 τ 的段的相关度, f^i 为第 i 维变量序列 x^i 的聚合表达, z_{τ}^j 为第 j 维变量序列 x^j 分段长度为 τ 的段表征. 同样地, 考虑到公式 (17)、公式 (18) 学习到的相关结构过于稠密, 容易引入领域变化的相关结构, 因此进一步进行稀疏化操作^[51-54].

$$\beta^i = \text{sparsemax}(\{e^{i1}, \dots, e^{ij}, e^{iN}\}) \quad (19)$$

其中, β^i 表示变量 i 与其他变量的相关度, 本文将利用其进行领域间的对齐, sparsemax 为稀疏操作^[51-54], 作用是将向量 $\mathbf{z} \in \mathbb{R}^K$ 欧氏投影到概率单纯形 Δ^{K-1} 上. 在对齐操作中, 本文借鉴了领域混淆网络的思想, 应用了最大平均差异 (MMD)^[9] 的方法, 形式化如公式 (20):

$$L_{\beta} = \sum_{m=1}^M |\text{avg}(\beta_S^m) - \text{avg}(\beta_T^m)| \quad (20)$$

其中, avg 表示平均值操作, β_S^m 和 β_T^m 分别为源域和目标域第 m 维变量与其他 $M-1$ 维变量的相关度.

联合变量内 (基于自注意力机制的特征聚合方法) 和变量间 (基于注意力机制的相关依赖学习方法) 的依赖约束方法, 本文可以同时使用变量内依赖对齐约束 L_{α} 和变量间依赖对齐约束 L_{β} , 以实现源域和目标域中同一个变量的内部依赖的对齐、源域和目标域中相关依赖结构的对齐.

3.5 标签预测

本文利用变量表征 z 和对齐的关联结构 β 进行目标域的标志预测, 以期达到知识迁移的目的. 首先计算了变量 j 的表征 z^j 在相应关联结构 β^{ij} 加权后的表征, 如公式 (21)、公式 (22) 所示:

$$U^{ij} = \sum_{\tau=1}^N \beta_{\tau}^{ij} \cdot z_{\tau}^j \quad (21)$$

$$U^i = \sum_{m=1, m \neq i}^M U^{im} \quad (22)$$

其中, U^i 为变量 i 与其他 $M-1$ 个变量进行加权求和后的关联结构表征. 然后连接加权后的段表征 f 和关联结构表征 U 得到最终的表征如公式 (23)、公式 (24):

$$H^i = [f^i; U^i] \quad (23)$$

$$H = [H^1; H^2; \dots; H^M] \quad (24)$$

其中, H 表示最终表征, 即标签预测器 G_y 的输入. 定义标签预测器 G_y 的损失函数为 L_y . 对于分类问题, 本文使用交叉熵 (cross entropy loss, CE) 作为标签损失, 如公式 (26) 所示. 对于回归问题, 本文使用均方根误差 (root mean square error, RMSE) 作为标签损失, 如公式 (27) 所示.

$$\hat{y} = G_y(H, \sigma) \quad (25)$$

$$L_y = CE(\hat{y}, y) \quad (26)$$

$$L_y = RMSE(\hat{y}, y) \quad (27)$$

3.6 目标函数

基于上述模型介绍, 本文提出的时间序列领域自适应的结构对齐模型的总损失公式为:

$$L(\theta, W^Q, W^K, W^V, \sigma) = L_y + \varepsilon(L_\alpha + L_\beta) \quad (28)$$

其中, ε 为超参数, 实验中设置为 1, L_y 为标签损失, L_α 、 L_β 分别为不同领域间变量内不同细粒段之间的关联权重损失、不同领域间变量间的稀疏关联结构权重损失. 在上述目标函数下, 本文的模型在源域和目标域上使用随机梯度下降的优化算法 (SGD) 进行训练:

$$(\theta, W^Q, W^K, W^V, \sigma) = \operatorname{argmin}_{\theta, W^Q, W^K, W^V, \sigma} L(\theta, W^Q, W^K, W^V, \sigma) \quad (29)$$

总的来说, 本模型首先利用基于路径签名的特征提取器将变量分段提取特征, 解决了时间序列数据中可能因领域不同导致时序数据受影响的问题, 获得每个变量特有的表征, 然后分别在变量内和变量间利用基于注意力机制的依赖学习器学习同一变量的不同段、变量间的底层关联性, 并在领域间进行对齐, 同时进一步提取领域表征, 用于预测任务.

4 理论分析

在之前的工作中, 领域自适应的泛化性主要取决于源域与目标域之间分布的距离. 在本文中, 基于时间序列数据的生成过程, 本节定义时间序列结构距离并推导其泛化边界, 泛化风险不仅取决于源域数据的风险, 还取决于源域和目标域之间的时间序列结构距离. 本文将在第 4.1 节中定义时间序列结构距离并证明其可行性, 在第 4.2 节中推导其泛化边界.

4.1 基于时间序列数据生成过程的结构距离

本节将定义时间序列结构距离并证明其可行性. 基于时间序列数据的数据生成过程如公式 (30) 所示:

$$x_i^t = F_i(pa(x_i^t), N_i) \quad (30)$$

其中, $pa(x_i^t)$ 表示 x_i^t 的父项, N_i 表示独立的噪声项; F_i 表示第 i 个时间序列数据 x_i^t 的任意灵活的函数类型, 指代数据由父项生成子项过程中所根据的某种规律. 由公式 (30) 可以发现, x_i^t 的分布取决于其父项 $pa(x_i^t)$ 的分布以及时间序列生成机制 F .

受上述序列生成机制的启发, 本文定义一个时间序列数据的距离, 如定义 1 所示.

定义 1. 时间序列数据的结构距离.

设 x 为多元时间序列数据, x_0 为 x 的第 1 个时间步的值. 本文假设 x 仅由 x_0 受依赖结构 A 的影响而生成, 在给定的源域数据 x_S 和目标域数据 x_T 时, 源域与目标域之间的结构距离可形式化如公式 (31) 所示:

$$dis_{SD}^{(S,T)}(x_S, x_T) = dist(P_S(x_0), P_T(x_0)) + \|A_S - A_T\|_1 \quad (31)$$

其中, $dist(\cdot, \cdot)$ 为自定义分布的距离度量函数, A_S 和 A_T 是源域和目标域的依赖结构对应的稀疏关联结构. 本文认为时间序列分布的距离由两个因素影响.

(1) 第 1 项 $dist(P_S(x_0), P_T(x_0))$ 表示时间序列初始时间步 x_0 的分布距离, 一个可理解的例子是, 时间序列数据的值范围会影响其分布.

(2) 第 2 项 $\|A_S - A_T\|_1$ 表示稀疏关联结构从源域到目标域之间的距离.

理论 1 将证明时间序列数据的结构距离满足度量的 3 个公理.

理论 1. 结构距离 (SD) 需满足的一般度量的 3 个公理.

(1) 距离度量函数满足非负性.

$$dis_{SD}^{(S,T)} \geq 0 \text{ and } dis_{SD}^{(S,T)} = 0 \text{ (if and only if } S = T) \quad (32)$$

(2) 距离度量函数满足交换律.

$$dis_{SD}^{(S,T)}(x_S, x_T) = dis_{SD}^{(S,T)}(x_T, x_S) \quad (33)$$

(3) 距离度量函数满足三角不等式.

$$dis_{SD}^{(S,T)} \leq dis_{SD}^{(S,D)} + dis_{SD}^{(D,T)} \quad (34)$$

其中, D 为任意分布. 上述一般度量的 3 个公理将在以下证明中被证明. 由理论 1, 本文利用时间序列结构上的生

成距离来度量时间序列数据分布之间的距离。

证明: 本文在定义 1 中定义了源域与目标域之间的结构距离如公式 (31) 所示, 进一步地, 本文定义时间序列初始时间步 x_0 的分布距离 $dist(P_S(x_0), P_T(x_0))$ 由 Wasserstein 距离组成, Wasserstein 距离^[55,56]也叫做推土机距离 (earth mover's distance), 是评估由一个分布转换成另一分布所需要的最小代价 (移动的平均距离的最小值)。对于分布 P_S 和 P_T 的初始时间步的分布 $P_S(x_0)$ 和 $P_T(x_0)$, 其 Wasserstein 距离可表示为:

$$W(P_S(x_0) \| P_T(x_0)) = \inf \left\{ \int |x_{S,0} - x_{T,0}|^p d\gamma, \gamma \in \prod (P_S(x_0), P_T(x_0)) \right\}^{1/p} \quad (35)$$

其中, $x_{S,0}, x_{T,0}$ 分别为从分布 $P_S(x_0), P_T(x_0)$ 中采样的样本, γ 表示 $P_S(x_0), P_T(x_0)$ 的联合分布, p 表示范数, 取决于所比较的对象的不同维度和形状, \inf 表示下界。

(1) 证明时间序列数据的结构距离满足第 1 个公理: 距离度量函数满足非负性。

由公式 (34) 易知, $|x_{S,0} - x_{T,0}|$ 恒大于等于 0, 故 $W(P_S(x_0) \| P_T(x_0))$ 恒大于等于 0, 又 $\|A_S - A_T\|_1$ 恒大于等于 0, 故证序列数据的结构距离满足非负性。

(2) 证明时间序列数据的结构距离满足第 1 个公理: 距离度量函数满足交换律。

由公式 (35) 可推导出 $W(P_T(x_0) \| P_S(x_0))$ 可表示为公式 (36):

$$W(P_T(x_0) \| P_S(x_0)) = \inf \left\{ \int |x_{T,0} - x_{S,0}|^p d\gamma \right\}^{1/p} = \inf \left\{ \int |x_{S,0} - x_{T,0}|^p d\gamma \right\}^{1/p} = W(P_S(x_0) \| P_T(x_0)) \quad (36)$$

另外易证 $\|A_S - A_T\|_1 = \|A_T - A_S\|_1$. 故证序列数据的结构距离满足交换律。

(3) 证明时间序列数据的结构距离满足第 1 个公理: 距离度量函数满足三角不等式。

Santamorogio 等人^[57]证明了 Wasserstein 距离满足度量定义的三角不等式, 又易证对于 $\|A_S - A_T\|_1$, 若存在另一分布 D 的结构 A_D , 存在 $\|A_S - A_T\|_1 \leq \|A_S - A_D\|_1 + \|A_D - A_T\|_1$. 故证序列数据的结构距离满足三角不等式。

4.2 泛化性理论分析

本节将推导时间序列无监督领域自适应的泛化边界^[58]. 假设 X 是时间序列的一个实例集, Y 是相应的标签集。假设 H 是模型的参数空间, $h \in H$ 表示模型的一组参数, $h(x)$ 表示使用参数 h 将样本 x 映射到标签集 Y 所在空间; 假设 η 为数据生成角度的标记函数, $\eta(x)$ 表示样本 x 的真实标签。根据分布 P_S , 概率函数定义如公式 (37):

$$\epsilon_S(h, f) = E_{x \sim P_S} [|h(x) - \eta_S(x)|] \quad (37)$$

在后面的描述中, 本文将使用 $\epsilon_S(h)$ 来简化表示 $\epsilon_S(h, \eta_S)$, $\epsilon_T(h)$ 同理。基于上述定义, 可以作出假设 1。

假设 1. 时间序列分布的边界假设。

假设 P_S 和 P_T 是源分布, 目标分布, A_S 和 A_T 是稀疏相关结构, 则存在一个正值 K , 使不等式 (38) 成立:

$$|P_S(x) - P_T(x)| \leq K(|P_S(x_0) - P_T(x_0)| + \|A_S - A_T\|_1) = Kdis_{SD}^{(S,T)}(x_S, x_T) \quad (38)$$

当 $dist(\cdot, \cdot)$ 为 $|\cdot|$ 时等号成立。基于上述定义和假设, 本文提出了所提出的基于路径签名的时间序列领域自适应方法的泛化界, 如理论 2 所示。

理论 2. 时间序列无监督域自适应的泛化约束。

$$\epsilon_T(h) \leq \epsilon_S(h) + K(|P_S(x_0) - P_T(x_0)| + \|A_S - A_T\|_1) + \lambda = \epsilon_S(h) + Kdis_{SD}^{(S,T)}(x_S, x_T) + \lambda \quad (39)$$

其中, K, λ 是常数。

证明: 理论 2 中时间序列无监督域自适应的泛化性约束计算过程如公式 (40) 所示:

$$\begin{aligned} \epsilon_T(h) &= \epsilon_T(h) + \epsilon_S(h) - \epsilon_S(h) + \epsilon_S(h, \eta_T) - \epsilon_S(h, \eta_T) \leq \epsilon_S(h) - \epsilon_S(h) + \epsilon_S(h, \eta_T) + |\epsilon_T(h) - \epsilon_S(h, \eta_T)| \\ &\leq \epsilon_S(h) + \epsilon_S(\eta_S, \eta_T) + |\epsilon_T(h) - \epsilon_S(h, \eta_T)| \leq \epsilon_S(h) + \epsilon_S(\eta_S, \eta_T) + \int |P_S(x) - P_T(x)| |h(x) - \eta_T(x)| dx \\ &\leq \epsilon_S(h) + \epsilon_S(\eta_S, \eta_T) + \int |P_S(x) - P_T(x)| dx \leq \epsilon_S(h) + K(|P_S(x_0) - P_T(x_0)| + \|A_S - A_T\|_1) \\ &\quad + \epsilon(\eta_S, \eta_T) = \epsilon_S(h) + Kdis_{SD}^{(S,T)}(x_S, x_T) + \lambda \end{aligned} \quad (40)$$

其中, $\lambda = \epsilon(\eta_S, \eta_T)$, 是一个常数.

根据这个泛化边界, 目标风险 $\epsilon_T(h)$ 不仅由 $\epsilon_S(h)$ 决定, 还由关联结构之间的距离、源域和目标域之间的结构生成距离决定, 它启发本文去提取领域不变的结构.

5 实验设计及结果分析

本文在 5 个公开基准数据集进行实验验证. 对于 UCI 人类活动识别 (UCI human activity recognition, UCIHAR)^[59] 数据集、人体行为姿势识别 (wireless sensor data mining, WISDM)^[60] 数据集、机会数据集 (OPPORTUNITY)^[61] 以及南加州大学人类活动数据集 (USC-HAD)^[62], 本文进行实验验证并报告了其 $F1$ 分数; 对于心率估计 (PPG-DaLiA, PPG)^[63] 数据集, 本文进行实验验证并报告其均方根误差 (RMSE). 所有实验均采用 3 个不同的随机种子重复多次实验, 并报告均值标准差. 所有数据集的参数设置如表 2 所示.

表 2 参数及设置范围

参数	描述	取值范围		
		HAR	WISDM	PPG
Optimizer	优化器	SGD	SGD	SGD
Training step	训练轮次	5 000	5 000	10 000
Batch size	训练批次大小	[32:128]	[32:128]	[1024:2048]
Sequence len	时间序列长度	128	128	6
Segments num	分段序列个数	6	6	6
Learning rate	学习率	[1E-3:1E-2]	[1E-3:1E-2]	[1E-3:1E-2]
Coeff	结构对齐权重	[1, 10, 100]	[1, 10, 100]	[1, 10, 100]
Dropout	Dropout 概率	[0.1, 0.3, 0.5]	[0.1, 0.3, 0.5]	[0.1, 0.3, 0.5]
Hidden dim	隐藏层大小	[10, 15, 20]	[10, 15, 20]	15
Threshold	稀疏化阈值	[0.08:0.35]	[0.08:0.35]	[0.08:0.35]
Seed	随机种子	[1, 2, 3]	[1, 2, 3]	[1, 2, 3]
Window size	滑动窗口大小	[4, 8, 12]	[4, 8, 12]	[4, 8, 12]
Sliding step	滑动窗口步长	1	1	1
Depth	路径签名深度	3	3	3

本次实验基于 Python 环境下 1.12 版本的 PyTorch 框架. 图形加速器的型号为 NVIDIA GeForce RTX 2080 Ti, 操作系统为 Linux (Ubuntu 版本为 16.04), CUDA 版本为 10.1, 程序在 PyTorch 1.6.0 框架下构建和训练, 采用的编程语言为 Python.

此外, 为了验证本文提出的基于路径签名的时间序列领域自适应方法中的每一个模块的有效性, 本文做了一系列消融实验. 本方法去掉基于注意力机制的结构学习模块, 得到模型变种 PSDA- α ; 本方法去掉基于路径签名的特征提取模块, 得到模型变种 PSDA- β .

本文第 5.1 节介绍数据集, 第 5.2 节简要介绍对比方法, 第 5.3–5.7 节展示并分析实验结果, 第 5.8 节将本文模型在不同数据集上各模块的消融实验进行可视化, 第 5.9 节展示与其他去除冗余信息的数据预处理方法的对比, 第 5.10 节将本文模型在不同数据集上对齐的稀疏关联结构^[50–53]进行可视化.

5.1 数据集

5.1.1 UCIHAR 数据集

数据集 UCIHAR 来自 UCI Machine Learning Repository, 是评估时间序列数据性能的最广泛使用的数据集之一. 它包含 3 个不同的传感器, 即加速度计、陀螺仪和身体传感器. 这些传感器被用于从 30 个志愿者处收集数据. 每个受试者都进行了 6 项活动, 即走着、上楼、下楼、站着、坐着和躺下. 使用其嵌入式加速度计和陀螺仪, 实验以 50 Hz 的恒定速率捕获 3 轴线性加速度和 3 轴角速度. 由于上述受试者之间的可变性, 本文将每个受试者视为一个单独的领域, 整个数据集共 30 个领域. 不同的活动会有不同的运动过程, 所以通过对受试者活动过程的时间

序列进行建模,产生有判别力的表征,可以对受试者所进行的活动进行分类.所以本文选择使用这个数据集来评估所提出的基于路径签名的时间序列领域自适应方法模型的性能.

5.1.2 WISDM 数据集

在 WISDM 数据集中,应用加速度计传感器收集了 36 名执行与 UCIHAR 数据集相同活动的受试者的数据.然而,由于不同主题数据中的分类平衡问题,该数据集可能更具有挑战性.具体来说,来自一些主题的数据可能只包含来自总体类的一个子集的样本.数据不平衡现象在分类问题中十分常见,该数据集可以验证模型对于少类样本的特征学习能力,所以本文选择使用这个数据集来评估所提出的基于路径签名的时间序列领域自适应方法模型的性能.

5.1.3 PPG 数据集

数据集 PPG 全称为 PPG-DaLiA,是公开可用于基于 PPG 的心率估计的数据集.这个多模态数据集以生理和运动数据为特征,在 15 名志愿者进行广泛的活动时,用手腕和胸部佩戴的设备记录下来.原始传感器数据记录了两个设备:一个腕带设备和一个腕带设备.胸部佩戴的装置提供心电图、呼吸和三轴加速.所有信号都在 700 Hz 进行采样.该腕带设备提供以下传感器数据:血容量脉搏、皮肤电活动、体温和三轴加速度.虽然不同的活动会导致心率的变化,但身体信号对心率的影响通常遵循相同的因果机制,所以本文可以使用这个数据集来评估所提出的基于路径签名的时间序列领域自适应方法模型的性能.

5.1.4 OPPORTUNITY 数据集

数据集 OPPORTUNITY 是对人类活动识别算法进行基准测试的数据集.数据集一共使用了 3 大类传感器:惯性、物体传感器和环境传感器,记录了 4 个受试者做出的 6 种大类的将近 100 种行为,每一个行为主题包含 5 个日常活动实例和 1 个演练实例(其中包含脚本化的活动序列实例).由于数据中含有缺失值,我们使用就近填充法进行补全,并使用归一化对数据进行预处理.在本文中,我们选取数据集中的低级动作分类作为实验对象.

5.1.5 USC-HAD 数据集

数据集 USC-HAD 是来自南加州大学的人类活动数据集,其中包含了 14 名志愿者的行为数据.每位志愿者对 12 种身体活动行为进行 5 次实验,每次实验时长约 24 s,数据由佩戴在受试者腰间的 MotionNode 仪器的三轴加速度计与三轴陀螺仪记录,传感器采样频率为 100 Hz.由于上述受试者之间行为习惯的可变性,本文将每个受试者视为一个单独的领域,整个数据集共 14 个领域.相比于 HAR 数据集,该数据集的行为种类更广泛,数据集复杂度更高,具有挑战性.所以本文选择使用这个数据集来评估所提出的基于路径签名的时间序列领域自适应方法模型的性能.

5.2 对比方法

在实验中,本文采用的对比方法包括非时间序列领域自适应方法:DDC^[22]:通过最小化源域与目标域的最大平均差异(MMD)^[9]距离进行领域自适应;Deep-Coral^[64]:通过对齐源分布和目标分布的二阶统计量来最小化领域偏移量;HoMM^[24]:通过匹配源域和目标域的高阶矩,最小化不同领域之间的差异;MMDA^[65]:通过条件熵,结合了MMD和CORAL最小化来解决领域知识的转移;DSAN^[32]:通过对齐相关子领域分布的局部最大平均差异(LMMD)来最小化源域和目标域之间的差异;DANN^[25]:设计了梯度反转层对特征提取器网络的领域分类器进行反向训练;CDAN^[66]:在领域对齐步骤中将任务知识与特征相结合,实现条件对抗对齐;DIRT-T^[67]:采用虚拟对抗训练、条件熵和教师模型来对齐源域和目标域,和时间序列领域自适应方法:AdvSKM^[17]:利用对抗性光谱核匹配来解决时间序列数据中的非平稳性和非单调性问题;DAF^[38]:使用基于注意力的共享模块和领域鉴别器分别学习领域不变和领域特定的表征.SASA^[34]:利用因果关系的稳定性探索数据的领域不变结构,并在领域间进行对齐.本文进一步考虑了EC-HAR^[68]:提出了通道均衡模块用于替代卷积层,并以类似的计算成本对基线模型实现更高的识别性能,是信道均衡化是首次应用于多模态HAR场景;filter activation^[69]:提出了一个用于HAR任务的CNN,通过激活CNN中无效的过滤器得到了比简单删除无效过滤器更好的性能.

5.3 UCIHAR 实验结果

本节展示并分析了本文所提出的基于路径签名的时间序列领域自适应方法在UCIHAR数据集下的实验结果.

本文从大量的跨域组合中随机选择了 5 组跨域场景进行实验, 分别为 12→16, 9→18, 3→16, 21→5, 28→9. 其中“→”两边的数字分别代表源域和目标域, 每项实验结果报告其 F1 分数, 并在括号内展示相应均值标准差.

表 3 展示了 UCIHAR 数据集的实验结果, 通过实验结果, 本文可得出以下结论: 1) 本方法实验结果相比较于其他对比方法均有较大提升, 证明了本方法的有效性. 2) 基于傅里叶变换进行特征提取的方法如 AdvSKM 的效果总体优于基于 CNN 进行特征提取的方法如 DDC, 这是因为 HAR 数据集的时间序列具有一定的周期性, 傅里叶变换可充分提取其中频域信息. 3) 在对比方法中, 基于对抗的对齐方法如 DANN, DIRT-T 相比于使用特征进行 MMD 对齐的方法如 DDC, Deep-Coral 等, 总体效果更好. 这是由于对抗学习设有领域分类器和梯度反转层, 相比于直接最小化源域和目标域的特征, 更有利于学习领域不变表征.

表 3 UCIHAR 实验结果及其对比实验

模型	12→16	9→18	3→16	21→5	28→9
DDC	59.65 (4.11)	45.42 (5.89)	70.59 (6.16)	66.19 (2.47)	70.34 (5.32)
Deep-Coral	62.13 (7.74)	47.77 (2.84)	66.35 (7.38)	67.47 (5.44)	68.83 (5.43)
HoMM	59.84 (1.43)	60.02 (11.83)	72.30 (2.04)	76.83 (2.07)	71.65 (3.35)
MMDA	62.64 (10.42)	58.43 (2.55)	70.65 (2.35)	81.95 (11.90)	84.17 (5.03)
DSAN	61.71 (1.75)	67.10 (4.61)	78.78 (6.79)	39.53 (3.18)	58.48 (7.58)
DANN	63.26 (2.49)	57.49 (7.77)	66.86 (3.73)	88.57 (3.98)	79.73 (9.58)
CDAN	54.66 (2.91)	40.94 (3.18)	52.73 (3.95)	86.57 (11.42)	84.30 (5.53)
DIRT-T	58.47 (2.98)	65.89 (13.25)	80.49 (3.39)	77.14 (1.56)	78.35 (2.25)
AdvSKM	60.09 (1.40)	53.70 (4.61)	75.94 (0.08)	72.18 (3.44)	63.61 (5.27)
DAF	60.86 (0.55)	54.72 (1.06)	54.02 (5.24)	59.48 (0.57)	55.18 (2.95)
CE-HAR	84.95 (1.19)	71.50 (7.22)	86.28 (0.64)	72.37 (4.39)	76.69 (0.52)
Filter activation	73.47 (2.42)	61.21 (5.51)	80.47 (4.32)	79.79 (2.45)	74.87 (1.05)
SASA	85.71 (8.26)	73.07 (7.22)	85.48 (3.95)	76.10 (2.63)	76.01 (1.89)
PSDA	82.40 (9.35)	75.33 (2.19)	87.92 (2.48)	79.91 (8.78)	78.77 (1.39)

5.4 WISDM 实验结果

本节展示并分析了本文所提出的基于路径签名的时间序列领域自适应方法在 WISDM 数据集下的实验结果, 如表 4 所示. 与 UCIHAR 数据集类似, 本文将每个受试者视为一个单独的领域, 并随机选择 5 组跨域场景, 分别为 35→31, 7→18, 20→30, 6→19, 18→23, 其中每项实验结果报告其 F1 分数, 并在括号内展示相应均值标准差.

表 4 WISDM 实验结果及其对比实验

模型	35→31	7→18	20→30	6→19	18→23
DDC	48.85 (16.15)	45.17 (6.65)	70.04 (10.28)	57.51 (6.89)	42.40 (8.53)
Deep-Coral	42.36 (11.34)	47.07 (7.25)	67.16 (4.86)	65.06 (3.76)	47.57 (8.73)
HoMM	66.29 (0.84)	48.67 (6.31)	65.30 (2.45)	63.78 (4.35)	62.11 (7.57)
MMDA	60.34 (7.52)	41.58 (8.79)	64.39 (4.28)	55.74 (3.88)	64.47 (10.75)
DSAN	57.25 (6.07)	52.77 (2.23)	63.40 (0.70)	53.35 (5.37)	55.76 (1.46)
DANN	52.21 (1.09)	41.16 (6.62)	71.96 (10.10)	59.09 (3.57)	48.00 (0.90)
CDAN	49.02 (4.20)	57.65 (0.18)	65.50 (0.61)	44.03 (0.81)	50.16 (0.44)
DIRT-T	46.75 (3.54)	57.89 (0.15)	65.49 (0.62)	45.16 (0.00)	50.90 (0.40)
AdvSKM	61.91 (6.95)	49.84 (5.31)	69.35 (1.38)	54.89 (4.14)	51.30 (10.33)
DAF	68.09 (0.36)	50.09 (4.53)	50.55 (1.48)	36.64 (3.05)	47.97 (1.75)
CE-HAR	65.64 (5.62)	62.42 (5.59)	71.38 (0.91)	65.52 (0.23)	58.04 (2.07)
Filter activation	62.43 (2.42)	58.75 (7.05)	74.55 (4.87)	65.00 (2.29)	63.39 (3.09)
SASA	77.72 (3.72)	63.58 (2.47)	77.48 (4.84)	76.27 (4.23)	60.49 (7.30)
PSDA	80.66 (4.89)	65.13 (2.35)	81.20 (6.46)	79.10 (7.24)	65.68 (5.00)

表 4 展示了 WISDM 数据集的实验结果, 通过实验结果, 本文可得出以下结论: 1) 本方法实验结果在所有对比方法中呈现最优结果, 证明了本方法的有效性. 2) 一些领域自适应方向的实验结果普遍较差, 如 35→31, 原因是类别

的极度不均衡(领域 35 中缺少某些类的数据)导致模型学习不到缺失类的表征,降低目标域的分类准确度. 3) 在某些方向中,如 6→19、20→30,时间序列的领域自适应方法如 SASA 相比于非时间序列领域自适应方法如 Deep-Coral、HoMM 效果更好,这是由于 SASA 利用了适应于时间序列数据的特征提取器,可以有效地捕获时间序列信息.

5.5 PPG 实验结果

本节展示并分析了本文所提出的基于路径签名的时间序列领域自适应方法在 PPG 数据集下的实验结果. 本文将每个活动视为 1 个单独的领域,整个数据集共 8 个领域. 本文随机选择了 4 个领域进行实验,分别为“Cycling”(C),“Sitting”(S),“Working”(W),“Drive”(D),其中每项实验结果报告其均方误差 (RMSE),并在括号内展示相应均值标准差.

表 5 展示了 PPG 数据集的实验结果,通过实验结果,本文可得出以下结论: 1) 在回归数据集 (PPG 数据集) 上本方法的有效性也得到了证明. 2) 基于对抗训练的领域自适应方法 DIRT-T 相比于基于 MMD 对齐的领域自适应方法如 DDC、HoMM 等效果更好,因为回归任务对误差的要求比分类任务高,所以学习的用于预测的表征需要更有判别力. 对抗训练能更好地学习具有判别力的表征. 3) 时间序列的领域自适应方法如 PSDA 相比于非时间序列领域自适应方法如 DIRT-T 效果更好,证明了将非时间序列领域自适应方法应用至时间序列数据中效果的局限性.

表 5 PPG 实验结果及其对比实验

模型	C→D	S→D	W→D	C→W	S→W	D→W
DDC	0.0091 (8.4E-5)	0.0103 (2.4E-4)	0.0092 (1.3E-4)	0.0100 (1.0E-4)	0.0088 (1.0E-4)	0.0066 (1.6E-4)
Deep-Coral	0.0085 (2.0E-4)	0.0107 (3.0E-4)	0.0087 (1.4E-4)	0.0094 (1.4E-4)	0.0089 (2.0E-4)	0.0060 (1.0E-4)
HoMM	0.0088 (1.1E-4)	0.0091 (3.9E-4)	0.0080 (1.3E-4)	0.0095 (1.7E-4)	0.0085 (1.2E-4)	0.0086 (1.1E-4)
MMDA	0.0087 (1.4E-4)	0.0093 (4.0E-4)	0.0080 (1.4E-5)	0.0098 (1.2E-4)	0.0094 (3.9E-4)	0.0085 (9.9E-5)
DANN	0.0089 (1.0E-4)	0.0090 (4.1E-4)	0.0078 (2.4E-4)	0.0097 (1.5E-4)	0.0085 (1.0E-4)	0.0064 (1.2E-4)
CDAN	0.0090 (4.9E-5)	0.0108 (3.3E-4)	0.0081 (2.7E-4)	0.0097 (1.1E-4)	0.0102 (5.4E-4)	0.0065 (2.4E-4)
DIRT-T	0.0087 (2.8E-4)	0.0091 (4.5E-4)	0.0087 (5.3E-4)	0.0094 (5.3E-4)	0.0090 (6.8E-4)	0.0070 (3.6E-4)
AdvSKM	0.0089 (1.1E-4)	0.0091 (4.1E-4)	0.0090 (9.4E-5)	0.0097 (1.6E-4)	0.0085 (1.0E-4)	0.0064 (1.1E-4)
DAF	0.0184 (1.4E-3)	0.0193 (3.2E-3)	0.0162 (6.1E-4)	0.0178 (1.1E-3)	0.0292 (7.6E-3)	0.0191 (2.7E-3)
PSDA	0.0051 (1.0E-4)	0.0055 (5.3E-4)	0.0056 (1.0E-3)	0.0078 (1.4E-3)	0.0056 (4.3E-4)	0.0037 (7.3E-5)
模型	C→S	W→S	D→S	S→C	W→C	D→C
DDC	0.0051 (1.9E-4)	0.0068 (7.1E-4)	0.0041 (1.1E-4)	0.0105 (6.2E-4)	0.0096 (3.1E-3)	0.0058 (2.2E-4)
Deep-Coral	0.0056 (3.6E-4)	0.0049 (4.3E-4)	0.0036 (1.1E-4)	0.0128 (5.2E-4)	0.0085 (5.5E-4)	0.0069 (3.1E-4)
HoMM	0.0058 (5.4E-4)	0.0053 (6.8E-4)	0.0052 (2.0E-4)	0.0157 (4.6E-3)	0.0091 (6.1E-4)	0.0058 (2.2E-4)
MMDA	0.0051 (1.7E-4)	0.0050 (2.2E-4)	0.0037 (1.0E-4)	0.0150 (7.0E-4)	0.0088 (5.6E-4)	0.0057 (2.8E-4)
DANN	0.0049 (2.1E-4)	0.0053 (6.5E-4)	0.0039 (1.5E-4)	0.0104 (5.5E-4)	0.0089 (4.8E-4)	0.0056 (3.2E-4)
CDAN	0.0054 (1.7E-4)	0.0068 (6.7E-4)	0.0042 (1.6E-4)	0.0105 (6.1E-4)	0.0096 (3.3E-4)	0.0058 (2.1E-4)
DIRT-T	0.0049 (3.9E-4)	0.0061 (7.1E-4)	0.0041 (3.9E-4)	0.0089 (4.1E-4)	0.0080 (6.8E-4)	0.0056 (2.5E-4)
AdvSKM	0.0058 (5.8E-4)	0.0067 (6.8E-4)	0.0039 (1.6E-4)	0.0111 (1.0E-4)	0.0088 (5.0E-4)	0.0056 (3.3E-4)
DAF	0.0084 (1.8E-3)	0.0081 (1.6E-3)	0.0068 (1.1E-3)	0.0188 (6.9E-3)	0.0740 (5.3E-2)	0.0113 (2.0E-3)
PSDA	0.0029 (2.4E-4)	0.0037 (1.1E-3)	0.0016 (5.2E-5)	0.0083 (1.3E-3)	0.0066 (2.1E-3)	0.0023 (2.0E-4)

5.6 USC-HAD 实验结果

本节展示并分析了本文所提出的基于路径签名的时间序列领域自适应方法在 USC-HAD 数据集下的实验结果. 与 UCIHAR 数据集类似, 本文将每个受试者视为一个单独的领域, 并随机选择 5 组跨域场景, 分别为 1→8、2→1、3→9、4→12、5→7. 其中每项实验结果报告其 F1 分数, 并在括号内展示相应均值标准差.

表 6 展示了 USC-HAD 数据集的实验结果, 通过实验结果, 本文可得出以下结论: 1) 从性能上看, 本文提出的方法超过了大部分主流的领域自适应方法, 如 CDAN, Deep-Coral, 证明了本方法的有效性. 此外, 和时间序列领域自适应方法 DAF 相比, 本方法也取得了较好的性能. 2) 所有领域自适应方法在某些迁移方向上取得不太理想的效果, 如 1→8, 原因是这些方向的源域和目标域的分布差异较大, 领域自适应工作的难度较高. 3) EC-HAR、filter

activation 方法针对此类人体行为识别工作, 设计了基于卷积神经网络的人体行为识别模型, 在源领域训练, 目标领域测试的实验下, 在某些较为相似的领域自适应过程中体现出了较好的泛化性.

表 6 USC-HAD 实验结果及其对比实验

模型	1→8	2→1	3→9	4→12	5→7	Average
DDC	31.21 (2.90)	57.31 (1.69)	48.05 (2.15)	45.44 (0.48)	68.30 (4.10)	50.07
Deep-Coral	38.49 (4.13)	56.05 (2.00)	45.62 (4.60)	45.02 (3.71)	71.68 (1.19)	51.38
HoMM	34.25 (2.96)	57.33 (7.48)	40.53 (0.99)	36.64 (2.06)	66.99 (3.98)	47.15
MMDA	42.44 (2.74)	55.17 (4.30)	40.21 (2.77)	44.11 (1.19)	70.31 (4.70)	50.45
DANN	43.09 (3.47)	66.46 (2.21)	46.21 (2.18)	46.53 (4.13)	72.44 (1.17)	54.94
CDAN	32.11 (2.55)	60.77 (2.41)	39.63 (7.32)	40.72 (8.39)	66.13 (9.34)	47.88
DIRT-T	35.05 (2.96)	74.85 (3.31)	38.09 (1.55)	42.21 (1.18)	55.02 (2.97)	49.05
AdvSKM	33.35 (2.00)	59.62 (0.85)	51.18 (0.95)	42.46 (3.75)	67.28 (1.56)	50.78
DAF	27.33 (2.58)	44.95 (1.66)	25.18 (4.81)	33.08 (3.77)	43.58 (6.46)	34.83
CE-HAR	42.98 (0.84)	62.86 (0.65)	44.87 (1.04)	45.14 (0.16)	79.27 (2.36)	55.03
Filter activation	41.47 (0.56)	62.87 (0.81)	46.93 (1.47)	48.48 (0.68)	76.30 (1.55)	55.22
SASA	46.91 (1.45)	62.64 (1.63)	46.52 (2.59)	49.61 (0.73)	74.05 (3.10)	55.95
PSDA	44.21 (0.56)	65.40 (4.27)	50.65 (6.67)	49.36 (6.63)	74.18 (2.06)	56.77

5.7 OPPORTUNITY 实验结果

本节展示并分析了本文所提出的基于路径签名的时间序列领域自适应方法在 OPPORTUNITY 数据集下的实验结果. 与 UCIHAR 数据集类似, 本文将每个受试者视为一个单独的领域, 并随机选择 2 组跨域场景, 分别为 1→2, 2→1. 其中每项实验结果报告其 $F1$ 分数, 并在括号内展示相应均值标准差.

表 7 展示了 OPPORTUNITY 数据集的实验结果, 通过实验结果, 本文可得出以下结论: 1) 本研究的方法在性能方面胜过了许多主要领域自适应方法, 包括 DANN 等, 从而验证了本方法的有效性. 2) 大部分对比方法如 DDC、Deep-Coral 等方法在本数据集上的效果不太理想, 原因可能是该数据集的空缺值较多, 本文简单使用就近填充法补充空缺值, 并不能模拟数据本该有的时间动态. 3) 本文所提出的方法较 EC-HAR, filter activation 等基于卷积神经网络提取特征的方法效果略低, 原因可能是本文所提出方法需学习不同变量间的依赖结构, 而该数据集的变量较多, 学习依赖结构较为困难.

表 7 OPPORTUNITY 实验结果及其对比实验 (%)

模型	DDC	Deep-Coral	HoMM	DANN	AdvSKM	CE-HAR	Filter activation	SASA	PSDA
1→2	49.65 (0.82)	50.76 (0.97)	47.76 (1.39)	50.80 (1.14)	54.29 (1.28)	81.04 (1.39)	81.40 (1.59)	75.37 (5.28)	80.98 (1.66)
2→1	46.41 (2.62)	43.72 (1.11)	38.31 (0.73)	52.18 (0.39)	47.90 (2.06)	66.55 (1.70)	65.90 (0.16)	72.34 (2.85)	69.23 (1.64)
Average	48.03	47.24	43.04	51.49	51.10	73.79	73.65	73.85	75.11

5.8 消融实验结果可视化

为了更直观地体现本方法各模块的有效性, 本文将各消融实验的结果以柱状图形式进行可视化, 其中图 5(a) 为 HAR 数据集消融实验结果可视化, 图 5(b) 为 WISDM 数据集消融实验结果可视化, 图 5(c) 为 PPG 数据集消融实验结果可视化.

图 5 中, PSDA 为本文所提出基于路径签名的时间序列领域自适应模型. PSDA- α 为本方法去掉基于注意力机制的结构学习模块的模型变种; PSDA- β 为本方法去掉基于路径签名的特征提取模块的模型变种. 由图可得出以下结论: 1) 本文所提出方法优于各消融模型, 证明了本方法基于路径签名的特征提取模块的模型与基于注意力机制的结构学习模块结合的有效性. 2) 在某些方向上 (如 HAR 数据集中 12→16, WISDM 数据集中 20→30 等) 各模型变种的效果接近, 因为在此类领域自适应任务中源域与目标域分布接近, 也证明了模型各变种的有效性.

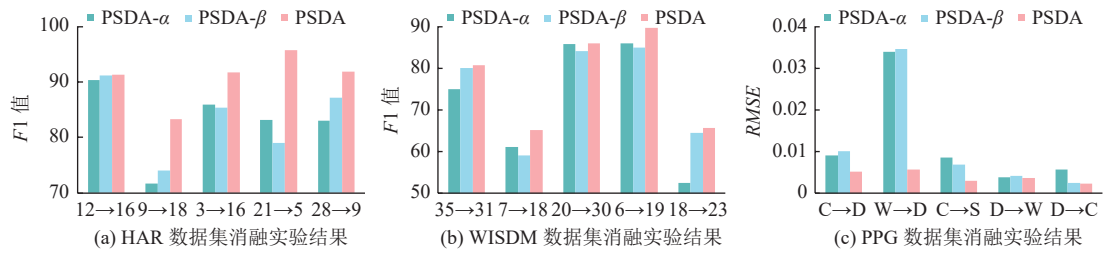


图5 HAR数据集及WISDM数据集消融实验结果可视化

5.9 与其他去除冗余信息的预处理方法的对比

考虑到预处理也可以去除冗余信息, 本文进一步对比所提方法在数据预处理阶段处理冗余数据的效果. 在此本文考虑了滚动平均值预处理方法, 并在PSDA、DAF、CDAN、AdvSKM方法上做对比, 随机选取了HAR数据集12→16, 9→18, 3→16这3个跨域场景. 实验结果可视化如图6所示. 由实验结果, 本文可以得出以下结论: 1) 对于本文所提出的方法, 滚动平均值预处理并不能使模型性能提高, 可能是因为在进行该预处理时, 破坏了时序数据中变量之间的时序依赖关系, 导致提取的特征不能充分解决领域之间因时滞、偏移带来的问题, 不利于依赖结构的学习. 2) 对于基于对抗的方法CDAN、AdvSKM, 滚动平均值预处理可使模型性能提高, 因为这些方法直接利用特征以对抗学习的方式对齐, 滚动平均值丰富了每个时间步蕴含的信息. 3) 对于DAF, 滚动平均值预处理也不能提高模型的性能, 与PSDA相似的, DAF也使用了注意力机制进行特征提取, 而预处理操作削弱了数据原本的时间信息.

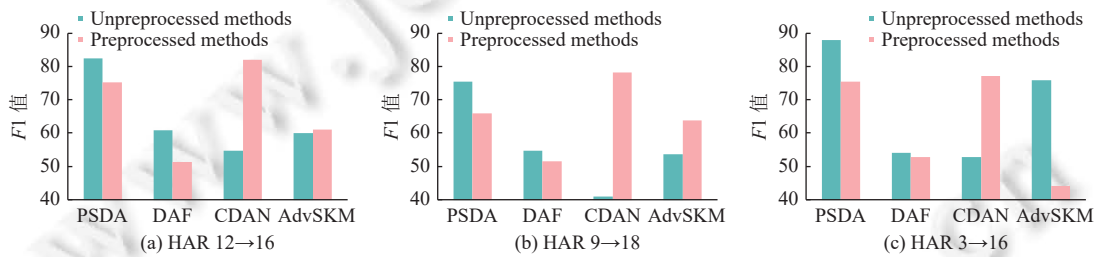


图6 预处理实验结果可视化

5.10 对齐结构可视化

为了在以后的工作中进一步对本文的结构对齐方法进行研究, 本节展示了在实验数据集中的稀疏关联结构的可视化, 其中图7为HAR数据集和WISDM数据集的稀疏关联结构可视化.

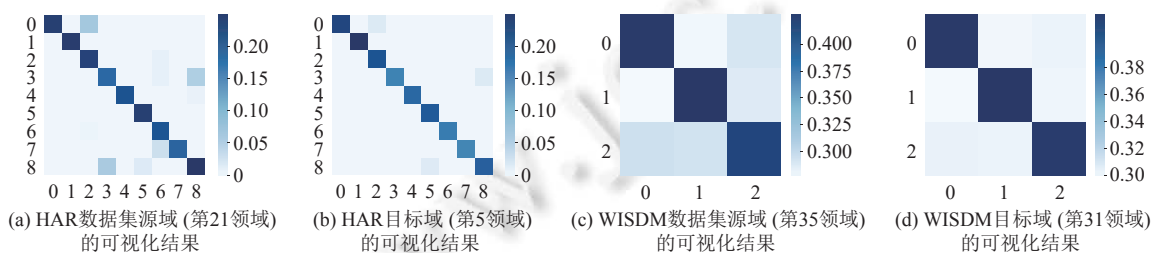


图7 稀疏关联结构可视化

图7中结构经稀疏化操作^[51-54], 重点体现关系较强的关联结构, 图中颜色越深, 表示关系越强. 可以发现:

- 1) 不同领域之间存在许多共享的关联关系, 意味着可以利用领域间共有的底层关联结构进行领域自适应工作.
- 2) 高维时间序列数据的底层关联结构的学习较低维数据更有挑战性.
- 3) 本文所采用的稀疏关联结构可充分捕获

不同变量之间较强的关联并舍弃较弱的关联, 避免了因结构冗余造成的对齐困难.

在本文实验中, 除了对比一些主流的和最近的领域自适应学习方法如 DANN, Deep-Coral 之外, 基于时间序列的领域自适应对比实验有 AdvSKM, SASA. 所有算法在同样的领域组合下评估性能. 综合 3 个数据集的实验结果, 基于路径签名^[70]的时间序列领域自适应方法取得了良好的效果.

6 结束语

本文提出了一种有效的时序领域自适应方法, 对于领域内每个变量分别进行基于路径签名的特征提取, 可以排除冗余的依赖关系并且充分探索有分辨力且可解释的变量表征. 其次通过基于注意力机制的结构学习方法对不同的变量进行相关性度量. 最后将不同领域内的稀疏依赖关系进行对齐, 并将变量表征与结构进行加权用于分类. 本文不但理论上分析了本方法的有效性, 而且在 UCIHAR, WISDM, PPG 和 OPPORTUNITY, 以及 USC-HAD 基准数据集上进行的广泛实验表明了架构的有效性.

References:

- [1] Li JJ, Meng LC, Zhang K, Lu K, Shen HT. Review of studies on domain adaptation. *Computer Engineering*, 2021, 47(6): 1–13 (in Chinese with English abstract). [doi: 10.19678/j.issn.1000-3428.0060659]
- [2] Zhuang FZ, Luo P, He Q, Shi ZZ. Survey on transfer learning research. *Ruan Jian Xue Bao/Journal of Software*, 2015, 26(1): 26–39 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [3] Gu X. A study on cross-domain classification and its application [Ph.D. Thesis]. Wuxi: Jiangnan University, 2014 (in Chinese with English abstract).
- [4] Qin JW. Research of transfer learning and its application in classifying cross-domain data [Ph.D. Thesis]. Guangzhou: South China University of Technology, 2011 (in Chinese with English abstract).
- [5] Xu M. A study on domain adaptation algorithm and its application [Ph.D. Thesis]. Wuxi: Jiangnan University, 2014 (in Chinese with English abstract).
- [6] Tang S, Ye M, Li XD. Domain adaptation object recognition. *ZTE Technologies*, 2017, 23(4): 25–31 (in Chinese with English abstract). [doi: 10.3969/j.issn.1009-6868.2017.04.005]
- [7] Li X. Research on cross domain image classification via transfer learning [Ph.D. Thesis]. Xi'an: Xidian University, 2017 (in Chinese with English abstract).
- [8] Fu JH. On the deep transfer learning and its applications [MS. Thesis]. Nanjing: Nanjing University of Posts and Telecommunications, 2020 (in Chinese with English abstract). [doi: 10.27251/d.cnki.gnjdc.2020.000222]
- [9] Gretton A, Borgwardt KM, Rasch MJ, Schölkopf BH, Smola A. A kernel two-sample test. *The Journal of Machine Learning Research*, 2012, 13(1): 723–773.
- [10] Cai RC, Li JH, Hao ZF. Unsupervised domain adaptive algorithm with intra-class maximum mean discrepancy. *Application Research of Computers*, 2020, 37(8): 2371–2375 (in Chinese with English abstract). [doi: 10.19734/j.issn.1001-3695.2019.03.0042]
- [11] Sun Q, Ling WX. Source domain selection in transfer learning based on domain similarity rank. *Science, Technology and Engineering*, 2020, 20(20): 8245–8251 (in Chinese with English abstract).
- [12] Zang WH. Research on transfer learning based on generative adversarial networks [MS. Thesis]. Chengdu: University of Electronic Science and Technology, 2018 (in Chinese with English abstract).
- [13] Jiang LL, Luo JJ, Zhu YP, Zhou DQ. A survey on adversarial attacks against deep learning. *Aerospace Electronic Warfare*, 2023, 39(1): 10–18, 50 (in Chinese with English abstract). [doi: 10.16328/j.htdz8511.2023.01.011]
- [14] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- [15] Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur, S. Recurrent neural network based language model. In: Proc. of the 11th Annual Conf. of the Int'l Speech Communication Association. Makuhari, 2010. 1045–1048. [doi: 10.21437/Interspeech.2010-343]
- [16] Da Costa PRDO, Akçay A, Zhang YQ, Kaymak U. Remaining useful lifetime prediction via deep domain adaptation. *Reliability Engineering & System Safety*, 2020, 195: 106682. [doi: 10.1016/j.res.2019.106682]
- [17] Liu Q, Xue H. Adversarial spectral kernel matching for unsupervised time series domain adaptation. In: Proc. of the 30th Int'l Joint Conf. on Artificial Intelligence. Montreal: IJCAI.org, 2021. 2744–2750. [doi: 10.24963/ijcai.2021/378]
- [18] He H, Queen O, Koker T, Cuevas C, Tsiligkaridis T, Zitnik M. Domain adaptation for time series under feature and label shifts. In: Proc.

- of the 40th Int'l Conf. on Machine Learning. Honolulu: PMLR, 2023. 12746–12774.
- [19] Wu HX, Xu JH, Wang JM. Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. In: Proc. of the 35th Conf. on Neural Information Processing Systems. OpenReview.net, 2021.
- [20] Li WD, Yang X, Liu WQ, Xia YC, Bian J. DDG-DA: Data distribution generation for predictable concept drift adaptation. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2022. 4092–4100. [doi: [10.1609/aaai.v36i4.20327](https://doi.org/10.1609/aaai.v36i4.20327)]
- [21] Zhang K, Schölkopf B, Muandet K, Wang ZK. Domain adaptation under target and conditional shift. In: Proc. of the 30th Int'l Conf. on Int'l Conf. on Machine Learning. Atlanta: JMLR.org, 2013. III-819–III-827.
- [22] Tzeng E, Hoffman J, Zhang N, Saenko K, Darrell T. Deep domain confusion: Maximizing for domain invariance. arXiv:1412.3474, 2014.
- [23] Sun BC, Feng JS, Saenko K. Correlation alignment for unsupervised domain adaptation. In: Csurka G, ed. Domain Adaptation in Computer Vision Applications. Cham: Springer, 2017. 153–171. [doi: [10.1007/978-3-319-58347-1_8](https://doi.org/10.1007/978-3-319-58347-1_8)]
- [24] Chen C, Fu ZH, Chen ZH, Jin S, Cheng ZW, Jin XY, Hua XS. HoMM: Higher-order moment matching for unsupervised domain adaptation. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2020. 3422–3429. [doi: [10.1609/aaai.v34i04.5745](https://doi.org/10.1609/aaai.v34i04.5745)]
- [25] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 2016, 17(1): 2096–2030.
- [26] Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial discriminative domain adaptation. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 2962–2971. [doi: [10.1109/CVPR.2017.316](https://doi.org/10.1109/CVPR.2017.316)]
- [27] Zhao H, Zhang SH, Wu GH, Costeira JP, Moura JMF, Gordon GJ. Adversarial multiple source domain adaptation. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 8559–8570.
- [28] Alam F, Joty S, Imran M. Domain adaptation with adversarial training and graph embeddings. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Melbourne: ACL, 2018. 1077–1087. [doi: [10.18653/v1/P18-1099](https://doi.org/10.18653/v1/P18-1099)]
- [29] Wright D, Augenstein I. Transformer based multi-source domain adaptation. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2020. 7963–7974. [doi: [10.18653/v1/2020.emnlp-main.639](https://doi.org/10.18653/v1/2020.emnlp-main.639)]
- [30] Xie SA, Zheng, ZB, Chen, L, Chen C. Learning semantic representations for unsupervised domain adaptation. In: Proc. of the 35th Int'l Conf. on Machine Learning. Stockholm: PMLR, 2018. 5419–5428.
- [31] Chen XY, Wang SN, Wang JM. Representation subspace distance for domain adaptation regression. In: Proc. of the 38th Int'l Conf. on Machine Learning. PMLR, 2021. 1749–1759.
- [32] Zhu YC, Zhuang FZ, Wang JD, Ke GL, Chen JW, Bian J, Xiong H, He W. Deep subdomain adaptation network for image classification. IEEE Trans. on Neural Networks and Learning Systems, 2021, 32(4): 1713–1722. [doi: [10.1109/TNNLS.2020.2988928](https://doi.org/10.1109/TNNLS.2020.2988928)]
- [33] Cai RC, Li ZJ, Wei PF, Qiao J, Zhang K, Hao ZF. Learning disentangled semantic representation for domain adaptation. In: Proc. of the 28th Int'l Joint Conf. on Artificial Intelligence. Macao: AAAI Press, 2019. 2060–2066.
- [34] Cai RC, Chen JW, Li ZJ, Chen W, Zhang KL, Ye JJ, Li ZZ, Yang XY, Zhang ZJ. Time series domain adaptation via sparse associative structure alignment. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2021. 6859–6867. [doi: [10.1609/aaai.v35i8.16846](https://doi.org/10.1609/aaai.v35i8.16846)]
- [35] Li ZH, Cai RC, Ng HW, Winslett M, Fu TZ, Xu BY, Yang XY, Zhang ZJ. Causal mechanism transfer network for time series domain adaptation in mechanical systems. ACM Trans. on Intelligent Systems and Technology, 2021, 12(2): 23. [doi: [10.1145/3445033](https://doi.org/10.1145/3445033)]
- [36] Xu XD. A visible digital watermarking classification method based on transfer learning. Technology Innovation and Application, 2023, 13(8): 139–142 (in Chinese with English abstract). [doi: [10.19981/j.CN23-1581/G3.2023.08.032](https://doi.org/10.19981/j.CN23-1581/G3.2023.08.032)]
- [37] Wang YY, Sun GW, Zhao GX, Xue H. Unsupervised new-set domain adaptation with self-supervised knowledge. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1170–1182 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6478.htm> [doi: [10.13328/j.cnki.jos.006478](https://doi.org/10.13328/j.cnki.jos.006478)]
- [38] Jin XY, Park Y, Maddix D, Wang H, Wang YY. Domain adaptation for time series forecasting via attention sharing. In: Proc. of the 39th Int'l Conf. on Machine Learning. PMLR, 2022. 10280–10297.
- [39] Ragab M, Eldele E, Chen ZH, Wu M, Kwok CK, Li XL. Self-supervised autoregressive domain adaptation for time series data. IEEE Trans. on Neural Networks and Learning Systems, 2024, 35(1): 1341–1351. [doi: [10.1109/TNNLS.2022.3183252](https://doi.org/10.1109/TNNLS.2022.3183252)]
- [40] Li ZJ, Cai RC, Fu TZJ, Hao ZF, Zhang K. Transferable time-series forecasting under causal conditional shift. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2023: 1–18. [doi: [10.1109/TPAMI.2023.3304354](https://doi.org/10.1109/TPAMI.2023.3304354)]
- [41] Chevyrev I, Kormilitzin A. A primer on the signature method in machine learning. arXiv:1603.03788, 2016.
- [42] Yang WX, Lyons T, Ni H, Schmid C, Jin LW. Developing the path signature methodology and its application to landmark-based human action recognition. In: Yin G, Zariphopoulou T, eds. Stochastic Analysis, Filtering, and Stochastic Optimization. Cham: Springer, 2022.

- 431–464. [doi: [10.1007/978-3-030-98519-6_18](https://doi.org/10.1007/978-3-030-98519-6_18)]
- [43] Glad W, Woolf T. Path signature area-based causal discovery in coupled time series. In: Proc. of the 2021 Causal Analysis Workshop Series. Minneapolis: PMLR, 2021. 21–38.
- [44] Morrill J, Fermanian A, Kidger P, Lyons T. A generalised signature method for multivariate time series feature extraction. arXiv:2006.00873, 2020.
- [45] Bonnier P, Kidger P, Arribas IP, Salvi C, Lyons TJ. Deep signature transforms. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 279.
- [46] De Curtò J, De Zarzà I, Yan H, Calafate CT. Learning with signatures. arXiv:2204.07953, 2022.
- [47] Zhao Y. Signature spatial improved temporal graph convolutional network. Computer Engineering and Science, 2022, 44(12): 2213–2219 (in Chinese with English abstract). [doi: [10.3969/j.issn.1007-130X.2022.12.015](https://doi.org/10.3969/j.issn.1007-130X.2022.12.015)]
- [48] Lyons TJ, Caruana M, Lévy T. Differential Equations Driven by Rough Paths. Berlin: Springer, 2007. [doi: [10.1007/978-3-540-71285-5](https://doi.org/10.1007/978-3-540-71285-5)]
- [49] Friz PK, Victoir NB. Multidimensional Stochastic Processes as Rough Paths: Theory and Applications. Cambridge: Cambridge University Press, 2010. [doi: [10.1017/CBO9780511845079](https://doi.org/10.1017/CBO9780511845079)]
- [50] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- [51] Duff IS. A survey of sparse matrix research. Proc. of the IEEE, 1977, 65(4): 500–535. [doi: [10.1109/PROC.1977.10514](https://doi.org/10.1109/PROC.1977.10514)]
- [52] Satuluri V, Parthasarathy S, Ruan YY. Local graph sparsification for scalable clustering. In: Proc. of the 2011 ACM SIGMOD Int'l Conf. on Management of Data. Athens: ACM, 2011. 721–732. [doi: [10.1145/1989323.1989399](https://doi.org/10.1145/1989323.1989399)]
- [53] Child R, Gray S, Radford A, Sutskever I. Generating long sequences with sparse Transformers. arXiv:1904.10509, 2019.
- [54] Martins AFT, Astudillo RF. From softmax to sparsemax: A sparse model of attention and multi-label classification. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: JMLR.org, 2016. 1614–1623.
- [55] Rüschemdorf L. The wasserstein distance and approximation theorems. Probability Theory and Related Fields, 1985, 70(1): 117–129. [doi: [10.1007/BF00532240](https://doi.org/10.1007/BF00532240)]
- [56] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein GANs. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 5767–5777.
- [57] Santambrogio F. Optimal Transport for Applied Mathematicians. Cham: Birkhäuser, 2015. [doi: [10.1007/978-3-319-20828-2](https://doi.org/10.1007/978-3-319-20828-2)]
- [58] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer, 2000. [doi: [10.1007/978-1-4757-3264-1](https://doi.org/10.1007/978-1-4757-3264-1)]
- [59] Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A public domain dataset for human activity recognition using smartphones. In: Proc. of the 21st European Symp. on Artificial Neural Networks. Bruges: ESANN, 2013. 437–442.
- [60] Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsletter, 2011, 12(2): 74–82. [doi: [10.1145/1964897.1964918](https://doi.org/10.1145/1964897.1964918)]
- [61] Roggen D, Calatroni A, Rossi M, *et al.* Collecting complex activity datasets in highly rich networked sensor environments. In: Proc. of the 7th Int'l Conf. on Networked Sensing Systems (INSS). Kassel: IEEE, 2010. 233–240. [doi: [10.1109/INSS.2010.5573462](https://doi.org/10.1109/INSS.2010.5573462)]
- [62] Zhang M, Sawchuk AA. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proc. of the 2012 ACM Conf. on Ubiquitous Computing. Pittsburgh: ACM, 2012. 1036–1043. [doi: [10.1145/2370216.2370438](https://doi.org/10.1145/2370216.2370438)]
- [63] PPG-DaLiA. 2019. <https://archive.ics.uci.edu/ml/datasets/PPG-DaLiA>
- [64] Rahman MM, Fookes C, Baktashmottlagh M, Sridharan S. On minimum discrepancy estimation for deep domain adaptation. In: Singh R, Vatsa M, Patel VM, Ratha N, eds. Domain Adaptation for Visual Understanding. Cham: Springer, 2020. 81–94. [doi: [10.1007/978-3-030-30671-7](https://doi.org/10.1007/978-3-030-30671-7)]
- [65] Long MS, Cao ZJ, Wang JM, Jordan MI. Conditional adversarial domain adaptation. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 1647–1657.
- [66] Shu R, Bui H, Narui H, Ermon S. A DIRT-T approach to unsupervised domain adaptation. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.
- [67] Huang WB, Zhang L, Wu H, Min FH, Song AG. Channel-Equalization-HAR: A light-weight convolutional neural network for wearable sensor based human activity recognition. IEEE Trans. on Mobile Computing, 2023, 22(9): 5064–5077. [doi: [10.1109/TMC.2022.3174816](https://doi.org/10.1109/TMC.2022.3174816)]
- [68] Huang WB, Zhang L, Wang SY, Wu H, Song AG. Deep ensemble learning for human activity recognition using wearable sensors via filter activation. ACM Trans. on Embedded Computing Systems, 2022, 22(1): 15. [doi: [10.1145/3551486](https://doi.org/10.1145/3551486)]
- [69] Kidger P, Lyons T. Signatory: Differentiable computations of the signature and logsignature transforms, on both CPU and GPU. In: Proc. of the 2019 Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2019.

- [70] Wilson G, Doppa JR, Cook DJ. Multi-source deep domain adaptation with weak supervision for time-series sensor data. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. ACM, 2020. 1768–1778. [doi: 10.1145/3394486.3403228]

附中文参考文献:

- [1] 李晶晶, 孟利超, 张可, 鲁珂, 申恒涛. 领域自适应研究综述. 计算机工程, 2021, 47(6): 1–13. [doi: 10.19678/j.issn.1000-3428.0060659]
- [2] 庄福振, 罗平, 何清, 史忠植. 迁移学习研究进展. 软件学报, 2015, 26(1): 26–39. <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [3] 顾鑫. 跨领域分类学习方法及应用研究 [博士学位论文]. 无锡: 江南大学, 2014.
- [4] 覃姜维. 迁移学习方法研究及其在跨领域数据分类中的应用 [博士学位论文]. 广州: 华南理工大学, 2011.
- [5] 许敏. 领域自适应学习算法及其应用研究 [博士学位论文]. 无锡: 江南大学, 2014.
- [6] 唐宋, 叶茂, 李旭冬. 领域自适应目标识别综述. 中兴通讯技术, 2017, 23(4): 25–31. [doi: 10.3969/j.issn.1009-6868.2017.04.005]
- [7] 李晓. 基于迁移学习的跨域图像分类方法研究 [博士学位论文]. 西安: 西安电子科技大学, 2017.
- [8] 付家慧. 深度迁移学习算法及其应用研究 [硕士学位论文]. 南京: 南京邮电大学, 2020. [doi: 10.27251/d.cnki.gnjdc.2020.000222]
- [10] 蔡瑞初, 李嘉豪, 郝志峰. 基于类内最大均值差异的无监督领域自适应算法. 计算机应用研究, 2020, 37(8): 2371–2375. [doi: 10.19734/j.issn.1001-3695.2019.03.0042]
- [11] 孙俏, 凌卫新. 基于域间相似度序数的迁移学习源领域的选择. 科学技术与工程, 2020, 20(20): 8245–8251.
- [12] 臧文华. 基于生成对抗网络的迁移学习算法研究 [硕士学位论文]. 成都: 电子科技大学, 2018.
- [13] 蒋玲玲, 罗娟娟, 朱玉鹏, 周东青. 基于深度学习的对抗攻击技术综述. 航天电子对抗, 2023, 39(1): 10–18, 50. [doi: 10.16328/j.htdz8511.2023.01.011]
- [36] 许宪东. 基于迁移学习的一种可见数字水印分类方法. 科技创新与应用, 2023, 13(8): 139–142. [doi: 10.19981/j.CN23-1581/G3.2023.08.032]
- [37] 汪云云, 孙顾威, 赵国祥, 薛晖. 基于自监督知识的无监督新集域适应学习. 软件学报, 2022, 33(4): 1170–1182. <http://www.jos.org.cn/1000-9825/6478.htm> [doi: 10.13328/j.cnki.jos.006478]
- [47] 赵艺. 基于路径签名的改进时空图卷积网络. 计算机工程与科学, 2022, 44(12): 2213–2219. [doi: 10.3969/j.issn.1007-130X.2022.12.015]



蔡瑞初(1983—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为因果发现, 图神经网络, 领域自适应学习, 自然语言处理.



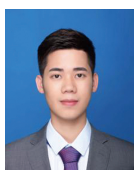
李梓健(1994—), 男, 博士生, 主要研究领域为迁移学习.



颜嘉文(1998—), 女, 硕士, 主要研究领域为时间序列领域自适应学习.



郝志峰(1968—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为因果发现, 图神经网络, 领域自适应学习, 自然语言处理.



陈道鑫(1996—), 男, 硕士, 主要研究领域为领域自适应学习.