

文章编号: 2095-2163(2023)12-0001-08

中图分类号: TP181

文献标志码: A

智能物联网时序数据分析关键技术研究综述

梁志宇, 王宏志

(哈尔滨工业大学 计算学部, 哈尔滨 150001)

摘要: 智能物联网是人工智能和物联网的结合,通过人工智能技术处理物联网产生的海量数据,提供智慧化的分析和决策,从而提升数据的实用价值。智能物联网被广泛应用于智慧城市、智慧医疗、智能家居、无人驾驶等多个领域。时间序列数据是智能物联网中最重要的数据类型之一。时序数据泛指一切随时间有序变化的数据集合。智能物联网各种应用场景产生的海量监测数据多以时序数据的形式存在。智能化的时序数据分析技术,包括时间序列分类、聚类和异常检测等,是支撑智能物联网应用的重要基础。本文对智能物联网时序数据分析中重要程度高、需求迫切的关键技术问题进行总结和分析,并进一步探讨未来的研究方向。

关键词: 智能物联网; 时序数据分析; 时间序列分类; 联邦学习; 自监督表示学习

Survey on key techniques of AIoT time series analysis

LIANG Zhiyu, WANG Hongzhi

(Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Artificial Intelligence of Things (AIoT) is the combination of Artificial Intelligence (AI) and Internet of Things (IoT), which provides intelligent analysis and decision making by processing the massive data generated by IoT using AI techniques, so that to enhance the practical value of the data. AIoT is widely used in many fields such as smart city, smart healthcare, smart home, autonomous driving and so on. Time series is one of the most important data type in AIoT. Time series refers to the data collections ordered over time. The massive monitoring data generated from various AIoT scenarios mostly exist in the form of time series. Intelligent time-series data analysis techniques, including time series classification, clustering and anomaly detection, are the basic foundations for AIoT applications. In this paper, we conduct a survey on the key techniques with high importance and urgent needs, and discuss the future research directions in the area of AIoT time series analysis.

Key words: AIoT; time series analysis; time series classification; federated learning; self-supervised representation learning

0 引言

物联网是继计算机、互联网后的世界信息产业第三大浪潮,是第四次工业革命的核心。移动通信技术的发展使得物联网扩展到人、车、家用和公共设施等泛在物体,实现了万物互联。根据互联网数据中心(Internet Data Center, IDC)预测,2021到2027年,为数字孪生建模的实物资产和流程数量将从5%增加到50%;到2025年,30%的城市将通过物联网等技术,将物理和数字相结合,以改善关键基础设施和数字服务的远程管理。

万物互联在给社会提供更为便捷的生产生活条件的同时,也引发了新的技术问题。其中最重要的

问题之一,是如何有效分析和利用万物互联制造的海量数据。为此,智能物联网(Artificial Intelligence of Things, AIoT)的概念应运而生。智能物联网是人工智能和物联网的融合,通过人工智能技术处理物联网产生的数据,提供智慧化的分析和决策,从而提升物联网的实用价值。智能物联网被广泛应用于智慧城市、智慧医疗、智能制造、智能家居、无人驾驶等多个领域。

时间序列泛指一切随时间有序变化的数据集合,时间序列数据是智能物联网中最具代表性的数据类型之一。智能物联网各种应用场景产生的海量监测数据,如:智能穿戴设备上的健康状态监测数据;智慧城市中的智能电网、智能通信基站、智能交

基金项目: 国家电网有限公司科技项目(5700-202119176A-0-0-00)。

作者简介: 梁志宇(1994-),男,博士研究生,主要研究方向:智能时序数据分析、联邦学习、工业大数据分析等。

通讯作者: 王宏志(1978-),男,博士,教授,主要研究方向:大数据管理与分析、智能化数据管理。Email: wangzh@hit.edu.cn

收稿日期: 2023-04-02

哈尔滨工业大学主办 ◆ 学术研究与应用

通信号灯的状态监控数据;以及工业传感器上的设备运行状态监控数据等,多以时间序列数据的形式存在。智能化的时间序列数据分析技术,包括时间序列分类、聚类、异常检测等,是实现人体健康状态监控、城市大脑、设备故障诊断等智慧应用的关键基础。例如:通过对智能手机、智能手表/手环等设备监测的人体运动数据进行准确分类,可以判断出携带者的行、走、坐、卧、跌倒等运动姿态,从而提供个性化的服务或对可能存在的风险意外进行及时告警;对城市电网中各个关键节点实时监控的电压、电流等信号进行异常检测,可以帮助电网管理人员及早对故障做出预警和定位,从而提高管理和维护的效率。因此,研究智能物联网时序数据的分析技术意义深远。

然而,智能物联网时间序列数据的特点和智能物联网应用,对数据分析技术的独特需求,为研究带来诸多挑战,主要体现在如下方面:

1) 数据异构

由于智能物联网应用场景广泛而多样,所产生的时间序列数据在采样频率、样本数量、序列长度、类别数量、维度、数据分布及数据所代表的物理规律和自然属性等方面均是高度异构的。例如:对于城市环境的监测中,空气污染指数和气温等,往往以小时为单位。而对于生产线上高速运转的机械加工设备,其监测数据的采样周期达到秒级甚至毫秒级。因此,智能物联网时序数据的分析技术应对海量异构数据具有泛化能力。

2) 标注稀少

在智能物联网实际应用中,为采集的时间序列数据提供准确充分的标注是相当困难的。例如:对于使用传感器监测生产线上运行设备的小型制造企业来说,由于特定工况相关的时间序列片段(如仪器的潜在故障等)通常位于整个监测时间序列中的未知区域且较为罕见,识别这些片段并标注类别往往需要丰富的专家经验,使得标注的获取代价十分昂贵。而没有足够的标记数据,则难以通过机器学习等人工智能技术来有效对数据建模和分析。

3) 响应迅速

智能物联网应用需要对监测对象源源不断产生的时间序列数据进行分析,要求分析方法具有良好的效率,从而保证能够对于不断产生的新数据作出快速响应,以满足实际应用的需求。

4) 决策可解释

不同于主要关注虚拟世界的传统互联网,物联

网应用与现实世界紧密相连。而诸多现实应用需要智能决策具有可解释性,从而提升人机物交互的可靠性,并满足法律、道德和伦理的要求。例如:在自动驾驶中,算法根据车辆的实时监控数据来规避故障并预测接下来的动作,要求算法的决策“有理可依”、“有迹可循”,从而对潜在的事故作出准确把握和判断,以此提高自动驾驶系统的安全性,并在发生交通事故时,能够准确对事故原因进行溯源,以明确法律责任。然而众所周知,一些现有的人工智能技术,如深度神经网络,具有非常弱的可解释性,因而难以很好地支撑智能物联网的时间序列数据。

面对智能物联网应用对时序数据分析的迫切需求和上述重要挑战,本文对智能物联网时序数据分析的关键技术问题的研究进展进行梳理和分析,主要包括时间序列分类、联邦学习和自监督表示学习。

1 时间序列分类

时间序列分类是智能物联网时间序列分析中最重要的问题之一。过去的十年中,有大量的时间序列分类算法相继提出。这些方法可以归纳为六大类,其中包括基于全局相似性的方法、基于区间特征的方法、基于词典的方法、基于元形状的方法、基于深度学习的方法和混合方法。

1.1 基于全局相似性方法

该类方法通过度量测试时间序列样本与训练数据的整体相似性,用最近邻分类器来判断序列的类别。其中,最基础的距离度量方法是欧氏距离。使用欧式距离度量的最近邻分类器(1NN-ED)在多种时间序列数据集上表现稳健^[1]。然而,欧氏距离无法很好地度量在时间维度上存在对齐差异的时间序列^[2-3]。因此,一些研究使用弹性的距离度量准则来替代欧式距离,其中最具代表性的弹性度量是动态时间规整(Dynamic Time Wrapping, DTW)^[4]。

1.2 基于区间特征方法

在时间序列中,分类特征通常位于整个序列的某些区间,而区间外的区域往往包含与类别无关的噪声。受此启发,一些研究从固定时间区间的序列中提取特征(如:平均值、傅里叶系数等),并用这些特征来训练分类器。由于可能的区间数量非常庞大,而对所有区间进行暴力枚举是不切实际的。因此,时间序列森林(Time Series Forest, TSF)^[5]提出随机采样不同长度的区间来提取特征,并使用集成的树模型来选择特征并构建分类器,以此提高模型的泛化性能。在此基础上,时间序列特征袋(Time

Series Bag of Features, TSBF)^[6]先利用随机采样区间上的特征构建模型,对每个区间的类别分布概率进行估计,并将所有采样区间的类别分布概率估计值离散化,最后以每个样本对应的所有离散化值的统计直方图,作为该样本的表示向量来训练分类器。学习模式相似性(Learned Pattern Similarity, LPS)^[7]利用多个回归树来建模不同区间子序列间的相关性,并以每个回归树叶节点子序列的数目,作为每个样本的特征向量。两个改进框架 TSBF 和 LPS 流程复杂,且分类性能相较于原始的 TSF 框架没有显著提升。因此,后续的研究主要以 TSF 框架作为基础,并通过改善特征种类和区间的采样方式来提高分类性能。随机区间谱集成(Random Interval Spectral Ensemble, RISE)^[8]提出对集成中每个树模型采样一个随机区间,并提取区间上的谱特征来构建树。典型区间森林(Canonical Interval Forest, CIF)^[9]提出使用精选的时间序列分类特征集^[10]来提取区间中序列的特征,从而大幅提高了分类的准确性。多样表示典型区间森林(Diverse representation Canonical Interval Forest, DrCIF)^[11]在 CIF 的基础上额外增加了 9 种特征类型,从而进一步提高了分类性能。

1.3 基于词典方法

受文本分类的启发,一些研究试图为时间序列设计特征词典,并根据词典中每个特征的计数,将时间序列实例转化为高维稀疏的特征向量。模式袋(Bag of Patterns, BOP)^[12]直接实现了这个想法。其通过符号聚合近似(Symbolic Aggregate approximation, SAX)^[13],将原始时间序列在滑动窗口内的子序列转换为单词,从而构成特征词典。文献[14]采用了一种著名的信息检索技术,即向量空间模型,来描述 SAX 单词的特征,从而为每个类别生成一个词袋集合来代表这个类别的特征,进一步改进了模式袋。文献[15]结合 SAX 表示和序列学习进行时间序列分类。该研究可以看作是字典方法的一个特例,每个 SAX 单词的词袋记录的是一个序列学习模型学习到的分类权重,而不是词频。除了基于 SAX 的方法外,文献[16]采用符号傅里叶近似(Symbolic Fourier Approximation, SFA)来获得时间序列在频域上的特征符号表示,从而将原始序列转换为词频向量,并使用定制的方法来构建分类模型,在分类准确率和运行效率上都表现出很大的优势。

1.4 基于元形状方法

元形状(Shapelet)是时间序列上具有类别区分能力的子序列^[17]。在早期研究中,元形状的挖掘过程与分类决策树训练过程耦合在一起,因而无法针对不同分类场景选择不同的分类模型,缺乏灵活性。元形状变换^[18]方法采用两阶段的方式实现基于元形状的分类。首先从原始时间序列的子序列中搜索得到最优的若干元形状,并通过原始数据与所有发现的元形状之间的距离,将时间序列映射为表示向量后,任何传统的分类模型都可以在这些向量上进行训练。由于元形状的搜索过程较为耗时,一些研究针对元形状搜索过程的特点提出了相应的加速技术,如:元形状距离计算剪枝^[17]、信息增益剪枝^[19]和计算缓存^[20]。

文献[21]提出了一种基于 SAX 表示和随机投影的快速元形状搜索方法,相较于直接在子序列上搜索更为高效。然而,由于采用了近似策略,这种方法会降低分类的精度。不同于从原始时间序列的子序列中选择元形状的方法,元形状学习(Learning shapelets, LS)^[22]提出将元形状作为分类模型的可学习参数,通过模型训练来学习优化的元形状。继承这一思想,后续的研究提出了许多改进方案。文献[23]使用更高效的学习算法来提高学习的效率。文献[24]通过学习 3 种不同类型的元形状来提高模型的性能。文献[25]对基于元形状的时间序列分类算法进行了综述。

1.5 基于深度学习方法

由于深度学习在计算机视觉和自然语言处理领域的巨大成功,一些研究适配或设计深度神经网络,来实现端到端的时间序列分类。文献[26]对几类代表性的深度时间序列分类方法进行综述和实验评估,其中包含多层感知机、卷积神经网络和回升状态网络。结果表明,采用卷积神经网络架构的深度残差网络,能够实现当前最优的分类性能,但由于众所周知的黑箱效应,基于深度学习的方法具有很弱的可解释性。文献[27]探索了基于长短时记忆网络的深度时间序列分类模型。文献[28]提出集成 5 个具有多样性架构的残差网络来降低模型的泛化误差。文献[29]提出基于多尺度残差的全卷积神经网络结构,融合不同网络来学习更丰富的分类表征。文献[30]提出带有注意力机制的原型网络来实现时间序列分类。文献[31]提出了基于转换器(Transformer)架构的时间序列分类方法。

1.6 混合方法

混合方法通过将多种不同类型的分类特征/模型组合在一起,进一步提高分类的性能。弹性集成(Elastic Ensemble, EE)^[32] 11个基于不同距离度量的最近邻分类器。接近森林(Proximity Forest, PF)^[33]将距离度量引入到随机森林训练中,以时间序列样本与一些参考样本间的距离作为节点划分的依据。异构集成嵌入森林时间序列组合(Time Series Combination of Heterogeneous and Integrated Embedding Forest, TS-CHIEF)^[34]进一步扩展了PF,其将不同类型的距离度量、序列的区间特征和字典特征均作为节点划分函数的候选,并选择最佳的划分函数来划分叶子节点。此外,有研究探索了一种元集成架构^[8],该架构以多个基于不同类型时间序列特征的模型作为基分类器,将每个基分类器输出的类别概率分布进行加权组合作为最终的输出。

综上所述,现有时间序列分类方法中,基于DTW距离的1最近邻具有简单易实现、性能稳定的特点,因而是最流行的基线方法。基于区间特征、字典特征、元形状和深度学习的方法在分类精度上的表现相当,但相较于深度模型,前3类方法在可解释性上具有显著优势。以元集成架构为代表的混合方法通过元集成组合多种分类模型,能够实现当前最优的分类性能。但其性能优劣主要取决于各个基分类器自身的性能。

2 联邦学习

联邦学习旨在以隐私保护的方式,联合利用多个参与方的数据构建机器学习模型。本节主要围绕3个有关方向对联邦学习的现有研究进行综述,其中包括联邦学习的基础模型、联邦学习的特定任务以及联邦学习的隐私保护。

2.1 联邦学习基础模型

联邦学习基础模型的研究主要关注线性模型、树模型和神经网络模型。

2.1.1 线性模型

作为最基础的机器学习模型,线性模型在联邦学习中持续受到关注。文献[35]使用差分隐私技术保护原始数据,以此实现明文上的联邦逻辑回归。文献[36]提出了基于同态加密的高效岭回归。文献[37]使用二阶多项式近似逻辑回归的代价函数,从而通过在加密数据上求解线性系统来学习逻辑回归模型。文献[38]研究了面向联邦学习的分

布式并行逻辑回归。

2.1.2 树模型

树模型具有良好的拟合能力和可解释性,因而在联邦学习中被广泛研究。文献[39]提出基于秘密共享的多方安全计算的决策树训练协议,能够在不泄露任何中间信息的前提下实现联邦学习。文献[40]提出了一种用于横向划分数据的梯度提升决策树训练协议。协议协调各方轮流使用本地数据训练回归树拟合残差,并将模型传递给下一方。由于每个树模型的学习过程不依赖加密计算技术和各方向通信,协议具有和本地计算相当的运行效率,但在各方向传递明文模型时会造成一定的隐私泄露。文献[41]提出一种用于纵向联邦学习的梯度提升决策树框架,利用纵向联邦决策树训练的特性,通过同态加密来实现隐私保护。以该框架为基础,文献[42]设计了更为高效的多方安全计算协议来提高联邦梯度提升决策树算法的效率;文献[43]使用分布式并行计算技术,设计和实现了工业级的高性能联邦学习系统。文献[44]研究了一种通用的树模型纵向联邦学习系统,联合使用同态加密、秘密共享和差分隐私等多种隐私保护技术实现安全高效的联邦学习。该系统支持基础的决策树模型及随机森林、梯度提升决策树等集成模型。

2.1.3 神经网络模型

虽然利用安全两方梯度下降算法实现了秘密共享数据上的神经网络联邦学习,但由于神经网络模型结构复杂、运算量大、过于依赖密态计算等因素,将大大降低联邦学习的效率。因此,近些年的研究主要探索在不暴露重要信息的前提下,充分利用明文数据进行联邦学习。文献[45]提出基于可选参数共享的神经网络联邦学习算法,通过在各方共享部分模型参数来实现构建模型。文献[46]提出一种模型平均联邦学习架构,借助一个受信任的第三方服务器聚合各方的模型参数,各方利用本地数据联合学习模型。以该框架为基础,文献[47]进一步对通信效率进行了优化。文献[48]利用差分隐私技术来降低参数共享时隐私泄露的风险。文献[49]提出了基于同态加密的安全聚合算法来保护各方私有的模型参数。考虑到加密和密文传输会降低联邦学习的效率,文献[50]提出将成批次的梯度值量化并编码到一个长整数,再对长整数加密和传输,从而减少加密运算次数和通信开销。文献[51]针对纵向划分数据,提出基于同态加密和秘密共享的反向传播计算协议,从而支持安全的神经网络模

型训练。

2.2 联邦学习特定任务

许多数据分析和机器学习任务,无法直接使用基础模型来完成,因而出现了一些面向特定任务的联邦学习研究。文献[52]探索了循环语言模型的联邦学习;文献[53]研究了联邦设定下的多任务学习;文献[54]提出了一种联邦迁移学习框架;文献[55]探索了联邦的排名学习问题;文献[56]针对分布式联邦学习各方数据非独立同分布问题,提出了个性化联邦学习的解决方案;文献[57]提出了一种用于联邦推荐系统训练的算法;文献[58]研究了无监督的联邦节点表示学习问题。此外,文献[59]提出了一种联邦特征选择方法;文献[60]探索了联邦学习中的数据调试问题;文献[61]针对联邦划分数据上的空间查询任务,给出了安全高效的解决方案;文献[62]针对联邦设定下的分类问题,提出了一种基于众包的隐私保护系统方案;文献[63]探究了基于联邦学习的服务质量(Quality of Service)优化方法;文献[64]面向金融预测领域,提出了一种改进的联邦决策树算法,有效提高了预测的准确率和效率。

2.3 联邦学习隐私保护

联邦学习最重要的基础之一是隐私保护技术。支撑联邦学习的隐私保护技术主要包括多方安全计算、同态加密和差分隐私等。安全多方计算是一种通用的隐私保护框架^[65],包含支撑许多常用运算操作的安全协议^[66]。在联邦学习通常考虑的半诚实模型下,这些安全计算协议不仅能够提供满足实际应用需求的性能,还可以通过零知识证明,将这些运算扩展到恶意模型上^[67]。同态加密通过对隐私数据加密并在密文上执行运算,来实现隐私保护。基于同态加密,可以很容易地实现安全加法运算,但其不支持除法和对比等较为复杂的操作;数据的加密和解密往往具有很高的计算开销^[43]。差分隐私技术通过在私密数据中添加噪声来实现隐私保护^[68],由于数据的值发生了改变,差分隐私会影响分析结果的准确性。

由于联邦学习的研究涉及诸多领域(如:机器学习、数据库、网络通信及密码学等),因此通过联邦学习的基础模型,可以为诸多实际应用提供通用的解决方案。而对于特定的联邦学习任务,考虑到安全性、准确性和效率等诸多因素,需要研究定制化的方法。在联邦学习中,隐私保护的实现主要基于多方安全计算、同态加密和秘密共享。多方安全计

算能够提供多种基本的运算操作,但其执行过程需要额外的计算和通信开销;同态加密能够方便地支持安全加法操作,但数据的加密和解密需要较高的计算代价;通过差分隐私技术,可以实现明文上的联邦计算,因而具有较高的计算效率。但噪声会影响联邦学习的准确性,因而基于差分隐私的方案常需要在隐私保护和准确性之间权衡。

3 自监督表示学习

3.1 概述

表示学习也称特征学习,是通过机器学习模型自动从数据中提取对下游分析任务有效的特征。自监督表示学习不使用数据标签,仅利用数据自身的结构特点来学习表示,自监督表示学习在计算机视觉和自然语言处理领域已被广泛研究。例如:在计算机视觉领域,文献[69]提出了一种简单有效的对比学习框架;文献[70]利用表示向量的原型(Prototype)作为参考,来提高对比学习的性能;文献[71]通过同时考虑不同场景和实例间的相关性,提出了一种更加通用的自监督图像表示框架。在自然语言处理领域,文献[72]是一个经典的自监督表示学习框架,其利用句子中相邻词之间的联系来学习词表征,文献[73]通过掩码的方式预训练通用的编码器,从而获得有益于下游任务的表示;文献[74]则通过对比学习范式来学习句子的表示。

3.2 面向时间序列的自监督表示学习

与计算机视觉和自然语言处理等领域不同,时间序列的自监督表示学习研究仍处于起步阶段。受自监督表示学习框架的启发,文献[75]将三元组损失扩展到时间序列领域来解决表示学习问题。类似地,由于转换器(Transformer)模型^[76]在自然语言建模中的成功,文献[31]探索了Transformer模型在自监督时间序列表示学习上的效果;文献[77]提出通过在潜在空间中预测未来时间步的表示来实现自监督学习;文献[78]在此基础上进行了扩展,通过联合利用预测和语义对比来提高表示的质量;文献[79]将时间戳级别的对比与上下文对比结合起来,以实现层次化的表示;文献[80]通过假设时间重叠序列之间的表示一致性,来建模随时间动态变化的潜在状态,而文献[81]则利用时域和频域之间的一致性来丰富表示的信息。

综上,自监督表示学习能够充分利用无标注的数据来学习有益于下游任务的表示(特征)。然而,对于时间序列的自监督表示学习研究方兴未艾。该

问题上的现有方法大多受到计算机视觉和自然语言处理等领域相关经验的启发,但时间序列在数据表现形式和语义上与图像和文本数据具有显著差异,故直接将上述领域的方法扩展到时间序列上,将无法充分发挥自监督表示学习范式的巨大潜力。未来的研究将更多地从时间序列数据的特性出发,从而学习到更高质量的表示。

4 结束语

本文对智能物联网时序数据分析的关键技术问题进行了综述,虽然现有的研究技术能够在一定程度上解决智能物联网时间序列分析的挑战问题,但仍然存在着诸多不足,主要结论如下:

(1)时间序列分类方法在准确性、可解释性和效率方面的综合表现仍有待提高。对智能物联网时间序列的分类通常需要确保对异构数据的高准确性和可解释性,同时具有良好的运行效率。现有研究中,基于元形状学习的时间序列分类方法在准确性和可解释性方面具有显著优势。然而,现有方法需要维护和更新大量类别共享的固定长度元形状,这导致算法的效率较低,同时也在一定程度上限制了分类准确性和可解释性。

(2)对联邦设定下可解释时间序列分类的研究呈现空白。时间序列分类方法通常依赖足够的标记数据。面对智能物联网缺少足够标注的问题,解决思路是通过联邦学习来安全地利用不同数据拥有者的标记数据共同构建分类模型。然而,现有的联邦学习研究主要关注通用的基础模型,但直接使用这些模型将无法很好地解决时间序列分类问题。具体来说,直接使用线性模型或树模型进行分类,会忽视时间序列的潜在特性,导致准确性较低,而神经网络模型具有很弱的可解释性。因此,需要研究定制化的联邦学习方法,从而快速构建准确、可解释的时间序列分类模型。

(3)自监督时间序列表示学习的研究刚刚起步,尚未充分考虑时间序列数据的特点。一个解决数据标注缺失的思路是利用自监督学习来获取高质量的表示,这样仅需要少量标注信息就可以获得有竞争力的分类性能。不仅如此,自监督学习的表示包含多种任务通用的信息,对于聚类、异常检测等重要时间序列分析任务同样有效,因而无需对不同任务从头训练模型,有利于提高响应速度。然而,现有的时间序列自监督表示学习方法大多由其它领域扩展而来,没有充分挖掘时间序列数据的特性。此外,

有些方法依赖于领域特定的假设,因而缺乏通用性。

针对现有研究存在的不足和空白,未来的研究方向应在现有分析算法的基础上持续优化性能,并探索面向可解释时序数据分析的联邦学习和自监督学习方案。

参考文献

- [1] BAGNALL A, LINES J, BOSTROM A, et al. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances [J]. *Data Mining and Knowledge Discovery*, 2017, 31(3): 606-660.
- [2] DING H, TRAJCEVSKI G, SCHEUERMANN P, et al. Querying and mining of time series data: experimental comparison of representations and distance measures [J]. *Proceedings of the VLDB Endowment*, 2008, 1(2): 1542-1552.
- [3] WANG X, MUEEN A, DING H, et al. Experimental comparison of representation methods and distance measures for time series data [J]. *Data Mining and Knowledge Discovery*, 2013, 26(2): 275-309.
- [4] BERNDT D J, CLIFFORD J. Using dynamic time warping to find patterns in time series [C]//*Proceedings of AAAI KDD Workshop*. Seattle: AAAI Press, 1994: 359-370.
- [5] DENG H, RUNGER G, TUV E, et al. A time series forest for classification and feature extraction [J]. *Information Sciences*, 2013, 239: 142-153.
- [6] BAYDOGAN M G, RUNGER G, TUV E. A bag-of-features framework to classify time series [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2796-2802.
- [7] BAYDOGAN M G, RUNGER G. Time series representation and similarity based on local autopatterns [J]. *Data Mining and Knowledge Discovery*, 2016, 30(2): 476-509.
- [8] LINES J, TAYLOR S, BAGNALL A. Time series classification with HIVE - COTE: The hierarchical vote collective of transformation - based ensembles [J]. *ACM Transactions on Knowledge Discovery from Data*, 2018, 12(5): 1.
- [9] MIDDLEHURST M, LARGE J, BAGNALL A. The canonical interval forest (CIF) classifier for time series classification [C]//*Proceedings of 2020 IEEE International Conference on Big Data*. 2020: 188-195.
- [10] LUBBA C H, SETHI S S, KNAUTE P, et al. Catch22: Canonical time - series characteristics - selected through highly comparative time - series analysis [J]. *Data Mining Knowledge Discovery*, 2019, 33(6): 1821-1852.
- [11] MIDDLEHURST M, LARGE J, FLYNN M, et al. HIVE-COTE 2.0: A new meta ensemble for time series classification [J]. *Machine Learning*, 2021, 110(11): 3211-3243.
- [12] AGGARWAL C C, ZHAI C X. A survey of text classification algorithms [C]//*Mining Text Data*. Springer. 2012: 163-222.
- [13] LIN J, KHADE R, LI Y. Rotation-invariant similarity in time series using bag-of-patterns representation [J]. *Journal of Intelligent Information Systems*, 2012, 39(2): 287-315.
- [14] SENIN P, MALINCHIK S. Sax-vsm: Interpretable time series classification using sax and vector space model [C]//*Proceedings of the 2013 IEEE 13th International Conference on Data Mining*. 2013: 1175-1180.

- [15] LE NGUYEN T, GSPONER S, IFRIM G. Time series classification by sequence learning in all-subsequence space[C]//Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE). 2017: 947-958.
- [16] SCHÄFER P. The BOSS is concerned with time series classification in the presence of noise[J]. *Data Mining and Knowledge Discovery*, 2015, 29(6): 1505-1530.
- [17] YE L, KEOGH E. Time series shapelets: a new primitive for data mining[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 947-956.
- [18] LINES J, DAVIS L M, HILLS J, et al. A shapelet transform for time series classification[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012: 289-297.
- [19] MUEEN A, KEOGH E, YOUNG N. Logical-shapelets: an expressive primitive for time series classification[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2011: 1154-1162.
- [20] 原继东, 王志海, 韩萌. 基于 Shapelet 剪枝和覆盖的时间序列分类算法[J]. *软件学报*, 2015, 26(9): 2311-2325.
- [21] RAKTHANMANON T, KEOGH E. Fast shapelets: A scalable algorithm for discovering time series shapelets[C]//Proceedings of the 2013 SIAM International Conference on Data Mining. 2013: 668-676.
- [22] GRABOCKA J, SCHILLING N, WISTUBA M, et al. Learning time-series shapelets[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 392-401.
- [23] HOU L, KWOK J T, ZURADA J M. Efficient learning of timeseries shapelets[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016: 1209-1215.
- [24] MA Q, ZHUANG W, COTTRELL G. Triple-shapelet networks for time series classification[C]//Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM). 2019: 1246-1251.
- [25] 闫汶和, 李桂玲. 基于 shapelet 的时间序列分类研究[J]. *计算机科学*, 2019, 46(1): 29-35.
- [26] FAWAZ H I, FORESTIER G, WEBER J, et al. Deep learning for time series classification: a review[J]. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917-963.
- [27] KARIM F, MAJUMDAR S, DARABI H, et al. Multivariate LSTM-FCNs for time series classification[J]. *Neural Networks*, 2019, 116: 237-245.
- [28] ISMAIL FAWAZ H, LUCAS B, FORESTIER G, et al. Inceptiontime: Finding alexnet for time series classification[J]. *Data Mining and Knowledge Discovery*, 2020, 34(6): 1936-1962.
- [29] 张雅雯, 王志海, 刘海洋, 等. 基于多尺度残差 FCN 的时间序列分类算法[J]. *软件学报*, 2022, 33(2): 555-570.
- [30] ZHANG X, GAO Y, LIN J, et al. TapNet: Multivariate time series classification with attentional prototypical network[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. AAAI. 2020: 6845-6852.
- [31] ZERVEAS G, JAYARAMAN S, PATEL D, et al. A transformer-based frame-work for multivariate time series representation learning[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 2114-2124.
- [32] LINES J, BAGNALL A. Time series classification with ensembles of elastic distance measures[J]. *Data Mining and Knowledge Discovery*, 2015, 29: 565-592.
- [33] LUCAS B, SHIFAZ A, PELLETIER C, et al. Proximity Forest: an effective and scalable distance-based classifier for time series[J]. *Data Mining and Knowledge Discovery*, 2019, 33(3): 607-635.
- [34] Shifaz A, Pelletier C, Petitjean F, et al. TS-CHIEF: a scalable and accurate forest algorithm for time series classification[J]. *Data Mining and Knowledge Discovery*, 2020, 34(3): 742-775.
- [35] CHAUDHURI K, MONTELEONI C. Privacy-preserving logistic regression[C]//Advances in Neural Information Processing Systems, 2008: 21.
- [36] NIKOLAENKO V, WEINSBERG U, IOANNIDIS S, et al. Privacy-preserving ridge regression on hundreds of millions of records[C]//Proceedings of the 2013 IEEE Symposium on Security and Privacy. 2013: 334-348.
- [37] AONO Y, HAYASHI T, TRIEU PHONG L, et al. Scalable and secure logistic regression via homomorphic encryption[C]//Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy. 2016: 142-144.
- [38] YANG S, REN B, ZHOU X, et al. Parallel distributed logistic regression for vertical federated learning without third-party coordinator[J]. arXiv preprint arXiv:1911.09824, 2019.
- [39] ABSPOEL M, ESCUDERO D, VOLGUSHEV N. Secure training of decision trees with continuous attributes[J]. *Cryptology ePrint Archive*, 2020: 167-187.
- [40] LI Q, WEN Z, HE B. Practical federated gradient boosting decision trees[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 4642-4649.
- [41] CHENG K, FAN T, JIN Y, et al. Secureboost: A lossless federated learning framework[J]. *IEEE Intelligent Systems*, 2021, 36(6): 87-98.
- [42] FANG W, ZHAO D, TAN J, et al. Large-scale secure XGB for vertical federated learning[C]//Proceedings of the 30th ACM International Conference on Information and Knowledge Management. 2021: 443-452.
- [43] FU F, SHAO Y, YU L, et al. VF2Boost: very fast vertical federated gradient boosting for cross-enterprise learning[C]//Proceedings of the 2021 International Conference on Management of Data. 2021: 563-576.
- [44] WU Y, CAI S, XIAO X, et al. Privacy preserving vertical federated learning for tree-based models[J]. arXiv preprint arXiv: 2008.06170, 2020.
- [45] SHOKRI R, SHMATIKOV V. Privacy-preserving deep learning[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. 2015: 1310-1321.
- [46] MCMAHAN H B, MOORE E, RAMAGE D, et al. Federated learning of deep networks using model averaging[J]. arXiv preprint arXiv:1602.05629, 2016.
- [47] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. 2017: 1273-1282.
- [48] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016: 308-318.
- [49] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical secure aggregation for privacy-preserving machine learning[C]//

- proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017: 1175–1191.
- [50] ZHANG C, LI S, XIA J, et al. {BatchCrypt}: Efficient homomorphic encryption for {Cross-Silo} federated learning [C]//Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC 20). 2020: 493–506.
- [51] FU F, XUE H, CHENG Y, et al. BlindFL: Vertical federated machine learning without peeking into your data[C]//Proceedings of the 2022 International Conference on Management of Data. 2022: 1316–1330.
- [52] MCMAHAN H B, RAMAGE D, TALWAR K, et al. Learning differentially private recurrent language models[J]. arXiv preprint arXiv:1710.06963, 2017.
- [53] CHEN J, ZHANG A. FedMSplit: Correlation-adaptive federated multi-task learning across multimodal split networks [C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 87–96.
- [54] LIU Y, KANG Y, XING C, et al. A secure federated transfer learning framework[J]. IEEE Intelligent Systems, 2020, 35(4): 70–82.
- [55] WANG Y, TONG Y, SHI D, et al. An efficient approach for cross-silo federated learning to rank[C]//Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE). 2021: 1128–1139.
- [56] HUANG Y, CHU L, ZHOU Z, et al. Personalized cross-silo federated learning on non-IID data[C]//Proceedings of the AAAI Conference. 2021: 7865–7873. arXiv:2007.03797
- [57] MUHAMMAD K, WANG Q, O'REILLY-MORGAN D, et al. Fedfast: going beyond average for faster training of federated recommender systems [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 1234–1242.
- [58] PAN Q, ZHU Y. FedWalk: Communication efficient federated unsupervised node embedding with differential privacy [C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022:1317–1326.
- [59] LI X, DOWSLEY R, DE COCK M. Privacy-preserving feature selection with secure multiparty computation[C]//Proceedings of the International Conference on Machine Learning. 2021: 6326–6336.
- [60] LIU Y, WU W, FLOKAS L, et al. Enabling SQL-based training data debugging for federated learning[J]. Proc. VLDB Endow, 2021, 15(3): 388–400.
- [61] TONG Y, PAN X, ZENG Y, et al. Hu-Fu: Efficient and secure spatial queries over data federation[J]. Proceedings of the VLDB Endowment, 2022, 15(6): 1159.
- [62] 金歌, 魏晓超, 魏森茂, 等. FPCBC: 基于众包聚合的联邦学习隐私保护分类系统[J]. 计算机研究与发展, 2022, 59(11): 2377–2394.
- [63] 张鹏程, 魏芯森. 移动边缘计算下基于联邦学习的动态 QoS 优化[J]. 计算机学报, 2021, 44(12): 2431–2446.
- [64] 郭艳卿, 王鑫磊, 付海燕, 等. 面向隐私安全的联邦决策树算法[J]. 计算机学报, 2021, 44(10): 2090–2103.
- [65] YAO A C. Protocols for secure computations[C]//Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), 1982: 160–164.
- [66] CHEN V, PASTRO V, RAYKOVA M. Secure computation for machine learning with SPDZ [J]. arXiv preprint arXiv:1901.00329, 2019.
- [67] GOLDREICH O, OREN Y. Definitions and properties of zero-knowledge proof systems [J]. Journal of Cryptology, 1994, 7(1): 1–32.
- [68] WEI K, LI J, DING M, et al. Federated learning with differential privacy: Algorithms and performance analysis [J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3454–3469.
- [69] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of Machine Learning Research, Vol 119: Proceedings of the 37th International Conference on Machine Learning, ICML. 2020: 1597–1607.
- [70] 汤凌韬, 陈左宁, 张鲁飞, 等. 联邦学习中的隐私问题研究进展 [J]. 软件学报, 2023, 34(1): 197–229.
- [71] LI Z, ZHU Y, YANG F, et al. Univip: A unified framework for self-supervised visual pre-training [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14627–14636.
- [72] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [C]//BENGIO Y, LECUN Y. Proceedings of the 1st International Conference on Learning Representations, ICLR. 2013. arXiv preprint arXiv:1301.3781.
- [73] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//BURSTEIN J, DORAN C, SOLORIO T. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2019: 4171–4186.
- [74] GAO T, YAO X, CHEN D. SimCSE: Simple contrastive learning of sentence embeddings [C]//MOENS M, HUANG X, SPECIA L, et al. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 6894–6910.
- [75] FRANCESCHI J Y, DIEULEVEUT A, JAGGI M. Unsupervised scalable representation learning for multivariate time series [C]//Advances in Neural Information Processing Systems, 2019, 32: 4652–4663.
- [76] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems. 2017: 5998–6008.
- [77] Den OORD A V, LI Y, VINYALS O, et al. Representation learning with contrastive predictive coding [J]. arXiv preprint arXiv:1807.03748, 2018.
- [78] ELDELE E, RAGAB M, CHEN Z, et al. Time-series representation learning via temporal and contextual contrasting [C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI. 2021: 2352–2359.
- [79] YUE Z, WANG Y, DUAN J, et al. Ts2vec: Towards universal representation of time series [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 8980–8987.
- [80] TONEKABONI S, EYTAN D, GOLDENBERG A. Unsupervised representation learning for time series with temporal neighborhood coding [C]//Proceedings of the 9th International Conference on Learning Representations, ICLR. 2021
- [81] YANG L, HONG S. Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion [C]//Proceedings of the International Conference on Machine Learning. 2022: 25038–25054.