

时序数据挖掘的预处理研究综述

张友浩, 赵 鸣, 徐梦瑶, 赵光晶
(上海工程技术大学 航空运输学院, 上海 201620)

摘要: 由于时序数据的海量、高维等特点, 对原始数据进行预处理是实现时序数据挖掘的重要步骤。有效的预处理不仅能去除噪声、冗余, 减低数据规模、提升计算效率, 还能通过提取数据特征, 揭示数据关系, 有助于实施时序数据挖掘。本文从数据缺失、数据噪声、离群点/异常值、特征提取和模式表示几个方面, 对已有的原理和方法进行分析归纳, 为后续时序数据挖掘研究提供参考。

关键词: 时间序列; 预处理; 数据清洗; 特征提取; 模式表示

Summary of research on preprocessing of time series data mining

ZHANG Youhao, ZHAO Ming, XU Mengyao, ZHAO Guangjing

(School of Air Transport, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 Due to the massive and high-dimensional characteristics of time series data, preprocessing the original data is an important step for time series data mining. Effective preprocessing can not only remove noise and redundancy, reduce the scale of data, and improve computing efficiency, but it can also reveal data relationships by extracting data features, which helps to implement time series data mining. This paper analyzes and summarizes the existing principles and methods from the aspects of data missing, data noise, outliers/anomalous value, feature extraction and pattern representation, and provides references for subsequent time series data mining research.

【Key words】 time series; preprocessing; data cleaning; feature extraction; pattern representation

0 引言

时间序列数据(Time Series Data)是指一系列依时间为序的观察值的集合。例如在设备运行状态监测、金融市场趋势预测、病人健康状态监测、交通流量监测, 网络流量监测等领域, 都会产生海量的时间序列数据。时变性是时序数据的重要特征, 不同的应用背景, 时序数据表现出的特性各不相同, 对其进行分类的标准相应地也有所不同。按照时序数据变量数目, 可分为单变量时间序列和多变量时间序列; 按其变量特征, 可分为平稳性时间序列和非平稳性时间序列; 按其连续性, 可分为连续时间序列和离散时间序列, 等等。

对时序数据分析的目的在于揭示时间序列规律, 并基于此规律预报未来事件, 再通过干预来控制未来事件。时序数据分析经历了描述性时序分析、统计性时序分析、频域分析、时域分析, 时间序列数据挖掘(Time Series Data Mining, TSDM)几个阶段。

由于时序数据具有高维度、高噪声、动态性及采集过程引起的数据缺失、数据冗余、离群点/异常值

等特点, 将直接导致后续处理分析耗时长, 无法提取有用信息, 甚至得出完全错误的结论。因此时序数据预处理是进行时序数据挖掘的重要步骤。

本文重点讨论时序数据预处理中涉及的数据缺失、数据噪声、离群点/异常值、特征提取与模式表示问题。

1 数据缺失

在数据采集过程中, 产生数据缺失的机制主要有完全随机缺失、随机缺失和非随机缺失三种情况。其中, 完全随机缺失完全由随机因素造成, 随机缺失只依赖于已观测到的变量值, 非随机缺失与缺失值自身有关。根据数据缺失机制与观测变量间的关系及不同的研究目标, 采取不同方法处理数据缺失。

1.1 基于统计学的填充方法

主要有固定值填充、常用值填充、均值填充、中值填充、上下文数据填充、插值填充^[1]等。例如, 文献[2]针对 GNSS 基准站坐标序列含缺失值问题, 根据小波系数与坐标序列数据的重构关系, 提出非插值的二进小波变换最小范数解法, 与传统的插值算

作者简介: 张友浩(1997-), 男, 硕士研究生, 主要研究方向: 交通通信与智能信息处理; 赵 鸣(1977-), 女, 副教授, 主要研究方向: 民航信息工程及控制、智能系统与数据挖掘、数据异常检测等。

通讯作者: 赵 鸣 Email: zhaoming@sues.edu.cn

收稿日期: 2020-09-23

哈尔滨工业大学主办 ◆ 系统开发与应用

法精度相当。文献[3]针对单变量时间序列缺失值问题,提出基于稀疏表示下的缺失值插补方法;针对多变量时间序列缺失值问题,提出SPCA-GKNNI算法,提高了数据集利用率和插补精度,避免了传统最近邻算法高维数据的维灾问题。

1.2 基于模型的填充方法

主要有自回归模型(Autoregressive Model, AR)^[4]、自回归差分移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)^[5]、马尔科夫链蒙特卡罗法等。此类方法的缺点是模型的误判容易导致估计值不够准确。文献[6]根据时空相关性建立填充缺失值模型,处理脏数据。文献[7]针对多元时间序列回归任务中的非随机缺失问题,提出一种称为正反向可变敏感的长短期记忆网络,由缺失指标、正向和反向缺失的两个时间间隔和每个变量的缺失率作为增强模型的输入,能较好处理大量信息缺失,提高预测精度。

1.3 基于机器学习的填充方法

主要有最近邻(K-Nearest Neighbour, KNN),递归神经网络(Recurrent Neural Networks, RNN),期望最大化算法(Expectation Maximization, EM)等。文献[8]针对KNN补全算法在缺失数据K近邻选取上的偏向性问题,引入空间坐标、阈值和影响权重进行改进,使得补全的数据更为合理。文献[9]针对多变量时间序列数据缺失问题,在RNN的长短时记忆单元引入衰减机制,能更好地获取长时间间隔的隐藏信息和实现缺失数据填充。文献[10]针对GPS坐标序列中的缺失值问题,提出基于数据驱动的RegEM插补算法,在大量数据缺失的情况下,插值效果均优于传统方法。文献[11]将统计学与机器学习结合,利用动态贝叶斯网络将测量数据通过隐藏状态传递给环境控制数据,用期望最大化算法来推断隐藏状态并学习模型的参数。与基于回归的分析模型相比,该模型能够以非常精确的方式拟合校准数据,对环境参数的变化具有更强的鲁棒性。

2 数据噪声

在数据采集过程中,噪声是不可避免的、不随原信号变化、无规律的额外信息,一般从时域和频域两个方面来进行分析。对此拟展开研究论述如下。

2.1 时域滤波方法

传统的时域滤波法是用一个滤波器来去除噪声,滤波器通常是一个滑动窗口,而去噪就是用窗口内数据的统计值(如均值、中值等)来代替当前值,

主要有滑动平均滤波、算术平均滤波、中位值滤波、自适应滤波等。时域滤波实际就是对数据进行平滑化处理,去除序列中的噪声,但计算量大且可能忽视时间序列局部细节变化。

2.2 频域滤波方法

传统的频域滤波法是将信号从时域变换到频域,通过设定阈值将不同频率的信号分开,主要有低通滤波、高通滤波、带通滤波等,但由于需要获取序列的先验知识且一定程度忽视了时间这一特征,无法准确展示序列细节变化,所以实际应用较少。随着小波理论的发展,其克服了傅里叶变换的缺点,在时序去噪方面主要有小波分解与重构法去噪^[12]、小波变换阈值去噪^[13]、小波变换模极大值去噪^[14]。文献[15]将信息熵引入小波包变换去噪中来表示残差序列小波系数的噪声含量,以获得有色噪声的小波系数,通过逆变换获得有色噪声的序列,并将其剔除。然而小波分析的结果受小波基的影响很大,确定的小波基可能是全局最优,但在局部并非最优。为克服基函数无法自适应的问题,经验模态分解(Empirical Mode Decomposition, EMD)将序列分解为若干个内涵模态分量(Intrinsic Mode Function, IMF)之和,各IMF分量包含了原序列不同时间尺度的局部特征,但EMD存在端点效应和模态混叠的缺陷。针对模态混叠问题,集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)^[16]通过向序列中添加白噪声的方式,使序列极值点分布均匀以减少模态混叠的影响。变分模态分解(Variational Mode Decomposition, VMD)^[17]通过迭代搜寻变分模态的最优解,不断更新各模态函数及中心频率,得到若干具有一定宽带的模态函数,有效地解决了传统EMD算法的端点效应和模态混叠问题。

3 离群点/异常值

数据中的异常值(也称为“离群点”)不同于噪声,可能蕴含着重要警示信息。在时间序列数据中,“异常”可能是在特定时间点的外部误差或外生变化,只影响此特定时刻的观测值;也可能是由于内部变化或噪声过程中的内源性作用引起,影响所有后续观测值^[18]。对此,文中将给出阐释分述如下。

3.1 基于统计学的方法

(1) 3σ 法则。该方法的缺点是需要先前获知序列的分布模型,而对于大多数序列而言难以描述其分布模型。

(2)回归分析。将时间作为自变量,序列值作为因变量,依据历史数据建立回归模型,如果预测值与观测值相差大于指定值,则认定为异常。主要有自回归(AR)模型、自回归移动平均(Auto Regressive Moving Average Model, ARMA)模型,但该类方法存在难以确定序列所属模型问题。

(3)相似性度量。主要以欧式距离、动态弯曲距离、编辑距离等作为测度函数,进行异常值检测,并可以借助于模式表示方法对原序列进行特征提取^[19],以减少干扰和计算复杂度,提高算法的效率,具有较好的鲁棒性。

(4)聚类分析。利用聚类算法对数据进行聚类,最终不能被聚类或者数据个数少的被视为“异常”,但需要满足数据集中大部分样本都是正常的或者异常占比例较小的前提条件,主要有DBSCAN^[20], K-means^[21]及其改进算法。

3.2 基于模式表示的方法

时间序列具有海量性的特点,直接对其进行异常检测所耗费的计算量和时间是十分巨大的,而模式表示可以保留原序列的形态趋势,去除干扰并进行一定程度的压缩,大大简化了计算复杂度,提高算法计算效率^[22]。目前主要有频率表示、奇异值表示(Singular Value Decomposition, SVD)、符号聚合近似(Symbolic Aggregate Approximation, SAX)、分段线性表示(Piecewise Linear Representation, PLR)、分段聚合近似(Piecewise Aggregate Approximation, PAA)等。

3.3 基于机器学习的方法

用于时间序列的机器学习方法主要有循环神经网络(Recurrent Neural Network, RNN)^[23]、卷积神经网络(Convolutional Neural Networks, CNN)^[24],其原理同传统的异常检测方法类似,将异常检测问题转换为分类或者预测问题。文献[25]针对时序数据不平衡问题,提出了基于深度卷积神经网络的偏斜类异常检测方法,采用抽样法增加少数类异常样本,缩小异常样本与正常样本间的差距,随后对数据进行尺度变换和时间切片分段作为分类器的输入。文献[26]在驾驶辅助领域将路面检测问题转换为时间序列分类问题,将小波卷积网络与多通道网络技术相结合,实现对路面状况的异常检测,提升驾驶安全。

4 特征提取和模型表示

由于时间序列数据的高维性,直接处理原数据

会占用大量资源,运算效率低下。而通过提取原数据中的主要特征并用合适的模型表示,不仅降低维度,节省运算成本,而且可以去除一定的噪声。目前主要有基于统计特征、基于预测模型、基于变换、基于分段函数的表示方式。这部分研究内容可做剖析阐述如下。

4.1 基于统计特征的时序表示

基于统计特征的模式表示方法主要涉及时域和频域两方面。时域包括最大值、最小值、峰值、均值、方差、均方值、均方根值等有量纲特征,以及峰值因子、脉冲因子、裕度因子、峭度因子、波形因子、偏度等无量纲特征。频域包括重心频率、均方频率、均方根频率、频率方差、频率标准差等特征。文献[27]针对滚动轴承故障提取轴承运行的时域信号特征,提高了故障识别准确率。文献[28]针对转子系统早期碰磨故障提出了一种基于时域和时频域联合特征提取和分析的方法。

4.2 基于预测模型的时序表示

基于预测模型的时序表示,假设时间序列是某个预测模型的观测值,确定该预测模型的参数值,构建合适的模型。一般分为模型假设、模型估计、参数估计、模型检验几个步骤,代表性的方法有自回归移动平均(ARMA)模型、马尔科夫链(Markov Chains, MCs)、隐马尔可夫模型(Hidden Markov Model, HMM)^[29]等。文献[30]针对滚动轴承故障识别问题,利用ARMA模型的自回归参数矩阵反映轴承工况的变化并结合表征故障程度的峭度值以充分提取轴承故障程度特征,为轴承故障诊断提供了一种有效的方法。文献[31]在时间序列异常检测方面,通过自编码器对分段后的序列进行特征提取并采用K-means处理生成符号化序列,进而构建隐马尔可夫模型(HMM),提升了表征原始时间序列的能力。

4.3 基于变换的时序表示

基于变换的表示方法将序列进行不同域中的映射变换,使得特征能在某一个维度上突显出来。常见的特征变换有傅里叶变换(Fourier Transformation, FT)、小波变换(Wavelet Transform, WT)、主成分分析(Principal Component Analysis, PCA)等。

傅里叶变换算法将序列分析从时域分析转换成频域分析,但傅里叶变换是一种整体性变换,无法描述序列的时频局部特性。针对这一缺陷, Gabor提出了短时傅里叶变换(Short-time Fourier Transform, SFTF),采用加窗的方法来描述序列局部特性。但

对于给定的窗函数而言,其分辨率是固定的。窄窗会导致时间分辨率高、频率分辨率低,宽窗会导致时间分辨率低、频率分辨率高。为解决这一缺陷,小波变换(WT)采用有限长会衰减的小波基,通过伸缩和平移实现对序列的多尺度细化分析。

主成分分析(PCA)将可能存在相关性的一组变量转换为线性不相关的变量,经过转换后的变量称为主成分,其本质是将方差最大的方向作为主要特征,从而剥离出对分类影响大的变量,实现降维。PCA很好地解除线性相关,但无法应用于非线性时序数据。因此,核主成分分析方法将在低维线性不可分的数据用非线性的变换映射到高维线性可分的数据空间,然后在高维的空间应用PCA提取特征进行降维。

文献[32]针对传统电路诊断方法诊断波形图与实际波形差异大,诊断结果准确率低的问题,提出基于小波变换和神经网络的诊断方法,利用小波变换来提取故障特征,通过DWT技术构建时频矩阵作为深度神经网络的输入,建立不同故障类型的映射关系实现故障诊断。文献[33]针对多元时间序列指标多、时序长,难以兼顾2个维度进行特征提取的问题,采用PCA提取指标维度特征结合长短时记忆神经网络(Long Short Term Memory Network, LSTM)降低时间维度,从而实现对数据的特征提取,有效地保留数据间的相互关系,降低维度和模型复杂度。

4.4 基于分段的时序表示

采用分段的思想可以提取原时间序列的整体特征,进行有效的数据压缩和噪声过滤,但计算量很大。代表性的方法有分段线性表示(Piece Linear Representation, PLR)^[34]、分段聚合近似(Piecewise Aggregate Approximation, PAA)^[35],符号聚合近似(Symbolic Aggregate Approximation, SAX)^[36]等。

PLR分为滑动窗口、自顶而下、自底而上三种类型^[37],其原理是利用首尾相连的线段将原序列的特征点连接起来,从而实现连续平滑的简化表示,但容易受到序列极值点和波动幅度的影响。PAA是将原序列进行分段,得到以分段均值组成的特征序列,但由于仅采用均值作为特征,而忽视了序列的形态趋势特征,可以通过增加特征来更为精确地表征原序列,例如斜率,标准差(或方差)等。SAX借助于PAA,对规范后(均值为0,标准差为1)的时序数据进行降维处理,将PAA所表示的均值离散映射到相应符号空间,应用领域扩展到文本处理、信息检索

等。对PAA的缺陷主要从分段特征和分段距离两个方面进行改进。在分段特征上使用统计特征,例如均值、最大值、最小值、斜率,拐点等。在分段距离方面就是对序列进行不等间隔的划分,例如使用全局关键点对序列进行分段,随后在分段上根据趋势特征进行分段或者利用随机小距离提高接近点分为同一段的概率。

文献[38]从时间序列形态趋势的角度,提出高、低滤波点和高、低滤波线概念,提取上、下滤波点作为线性分段点来描述时间序列,具有良好的逼近性。文献[39]针对分段线性表示受序列极值点和波动幅度影响问题,提出基于转折点和趋势段的时间序列趋势提取算法,克服使用斜率差作为评价指标的非线性缺陷,结合极值趋势段得到全局意义下的转折点,并针对分段点不足问题,定义差值转折点,在分段点序列中进行插值以满足最后的压缩率要求。

5 结束语

由于原始时序数据中存在数据缺失、数据噪声、离群点/异常值等问题,数据挖掘前对数据进行预处理至关重要。

面对时序数据海量性和高维度所导致的高度复杂性,在数据缺失方面,传统的统计学方法已经无法满足处理要求,借助于深度学习算法强大的学习能力,基于数据驱动学习变量间的相互关系和内在规律以克服人工特征的局限性,从而提高缺失值填充准确率。

在时序数据的去噪方面,小波变换和经验模态分解已得到了广泛应用。但对于特殊领域以及噪声分布的不均衡性问题,往往只能对特定的噪声进行去除,例如高频白噪声等。因此,如何最大程度地消除噪声的影响,根据噪声的特性自适应寻优去噪一直是研究的热点。

在异常检测方面,单变量时间序列异常检测的发展已经较为成熟,主要有基于聚类、相似性、模型、分类、预测等异常检测方法。鉴于实际应用中的时效性要求,未来研究将更加侧重于算法的在线实时检测能力。多变量时间序列异常检测目前有2种主流思路,一是将多变量时间序列转换为多个单变量时间序列进行异常检测,二是直接对多变量时间序列进行异常检测。

特征提取和模式表示是在充分表征原序列的基础上,最大程度地简化序列,以减少计算复杂度,提

高算法效率。多变量时间序列包含时间维和变量维,极大地增加了数据规模,同时又涉及变量维间相关性等问题,如何对现有特征提取和模式表示方法进行扩展以适应多变量时间序列的要求是未来研究的重点。

总的来说,现有的时序数据预处理方法主要针对单变量时间序列,而随着研究的深入,能够更加全面描述事物特性的多变量时间序列将成为主流。相较于单变量时间序列,多变量时间序列增加了变量维度,导致数据规模急剧扩大并引入变量维间相关性等问题,使得预处理更加复杂,因此在其预处理中应当着重考虑变量的维间相关性,去除冗余信息,在变量维和时间维进行特征提取和模式表示以简化序列,为后续的数据分析做铺垫。

参考文献

- [1] MA Jun, CHENG J C P, DING Yuexiong, et al. Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series[J]. *Advanced Engineering Information*, 2020, 44:101092.
- [2] 稽昆浦,沈云中. 含缺值 GNSS 基准站坐标序列的非插值小波分析与信号提取[J]. *测绘学报*,2020,49(5):537-546.
- [3] 程万伟. 时间序列缺失值插补方法研究[D]. 长沙:湖南大学,2018.
- [4] SRIDEVI S, RAJARAM S, PARTHIBAN C, et al. Imputation for the analysis of missing values and prediction of time series data [C]//2011 International Conference on Recent trends in information Technology(ICRITIT) . [S. l.] : IEEE,2011:1158-1163.
- [5] 邓方进,王绪本,李德伟. ARIMA 模型在 LMT 数据处理中的应用[J]. *物探化探计算技术*,2017,39(5):612-619.
- [6] 孙天成. 基于时空相关性的感知数据清洗研究[D]. 北京:北京建筑大学,2020.
- [7] FOULADGAR N, FRÄMLING K. A novel LSTM for multivariate time series with massive missingness [J]. *Sensors (Basel, Switzerland)*, 2020,20(10):2832.
- [8] 张孙力, 杨慧中. 基于改进的 K 近邻缺失数据补全[J]. *计算机与应用化学*, 2015, 32(12):1499-1502.
- [9] 孙晓丽, 郭艳, 李宁, 等. 基于改进 RNN 的多变量时间序列缺失数据填充算法[J]. *信息技术与网络安全*, 2019,38(11):47-53.
- [10] 王方超, 吕志平, 吕浩, 等. 基于数据驱动的 Reg EM 算法在 GPS 坐标时间序列插值中的应用[C]// 第十届中国卫星导航学术年会. 北京:中国卫星导航学术年会组委会,2019:29-34.
- [11] KOCIAN A, CARMASSI G, CELA F, et al. Bayesian Sigmoid-type time series forecasting with missing data for greenhouse crops [J]. *Sensors*, 2020, 20(11):3246.
- [12] 张鹏,王甜,王远明. 基于小波变换去噪的 ARMA 堤坝水平位移预测模型[J]. *人民长江*,2017,48(S2):259-261,287.
- [13] 牛宏侠,张肇鑫,宁正,等. 基于小波变换的阈值自适应寻优去噪方法[J]. *传感器与微系统*,2020,39(2):33-36.
- [14] 赵鸿图,刘云. 基于三次样条插值的小波模极大值去噪算法[J]. *计算机工程与设计*,2014,35(8):2965-2968,2975.
- [15] 马俊,曹成度,姜卫平,等. 利用小波包系数信息熵去除 GNSS 站坐标时间序列有色噪声[J/OL]. *武汉大学学报(信息科学版)*:1-11 [2020-06-30]. <http://doi.org/10.13203/j.whugis20190353>.
- [16] WU Zhaohua, HUANG N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method [J]. *Advances in Adaptive Data Analysis*,2009,1(1):1-41.
- [17] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. *IEEE Transactions on Signal Processing*,2004, 62(3):531-544.
- [18] 尚华. 两类时间序列模型的异常值检测研究[D]. 北京:首都经济贸易大学,2016.
- [19] 孙焱,林意. 基于相似性分析的时间序列异常检测方法[J]. *山东农业大学学报(自然科学版)*,2017,48(2):287-292.
- [20] 肖勇,郑楷洪,余忠忠,等. 基于三次指数平滑模型与 DBSCAN 聚类的电量数据异常检测[J]. *电网技术*,2020,44(3):1099-1104.
- [21] 张仁斌,许辅昊,刘飞,等. 基于 K-均值聚类的工业异常数据检测[J]. *计算机应用研究*,2018,35(7):2180-2184.
- [22] 詹艳艳,陈晓云,徐荣聪. 基于时间序列模式表示的异常检测算法[J]. *计算机应用研究*,2007,24(11):96-99.
- [23] 曾霄笑. 基于改进 RNN 及密度聚类的异常流量检测方法[D]. 北京:北京邮电大学,2019.
- [24] 胡姣姣. 基于深度学习的飞行器遥测时间序列数据异常检测与预测方法研究[D]. 西安:西安理工大学,2019.
- [25] 胡姣姣,王晓峰,张萌,等. 基于深度学习的时间序列数据异常检测方法[J]. *信息与控制*,2019,48(1):1-8.
- [26] 李博,张洪刚. 基于多通道小波卷积神经网络的路面异常检测算法[J]. *华中师范大学学报(自然科学版)*, 2019,53(2):200-206.
- [27] 马丽华,朱春梅,赵西伟,等. 基于思维进化优化极限学习机的滚动轴承故障的智能诊断[J]. *制造技术与机床*,2019(11):103-109.
- [28] 赵炳钢,冀大伟,袁奇,等. 基于时域与时频域联合特征提取和可分度分析的转子系统碰磨故障诊断[J/OL]. *西安交通大学学报*:1-10 [2019-10-17]. <http://kns.cnki.net/kcms/detail/61.1069.T.20191017.1340.004.html>.
- [29] 李海林,杨丽彬. 时间序列数据降维和特征表示方法[J]. *控制与决策*,2013,28(11):1718-1722.
- [30] 韩清鹏,李天成,李晨晨,等. 基于 ARMA 模型的滚动轴承故障诊断研究[J]. *机电工程技术*,2018,47(1):68-71.
- [31] 霍伟纲,王慧芳. 基于自编码器和隐马尔可夫模型的时间序列异常检测方法[J]. *计算机应用*,2020,40(5):1329-1334.
- [32] 何东中,贡丽霞,白艳萍. 小波变换和神经网络的电路故障诊断[J]. *现代电子技术*,2020,43(10):30-32,37.
- [33] 张可,崔乐. 基于 PCA-LSTM 模型的多元时间序列分类算法研究[J]. *统计与决策*,2020,36(15):44-49.
- [34] 喻高瞻,彭宏,胡劲松,等. 时间序列数据的分段线性表示[J]. *计算机应用与软件*,2007,24(12):17-18.
- [35] 李海林,郭崇慧,杨丽彬. 基于分段聚合时间弯曲距离的时间序列挖掘[J]. *山东大学学报(工学版)*,2011,41(5):57-62.
- [36] LIN J,KEOGH E,LONARDI S, et al. A symbolic representation of time series, with implications for streaming algorithms [C]// Proc of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, California, USA: ACM Press,2003:2-11.
- [37] 张苗苗. 心电监测时序数据识别算法研究[D]. 南京:南京航空航天大学,2018.
- [38] 林意,朱志静. 基于趋势的时间序列分段线性化算法[J]. *重庆大学学报*,2019,42(3):92-98.
- [39] 刘意杨,李俊朋,白洪飞,等. 基于转折点和趋势段的时间序列趋势特征提取[J]. *计算机应用*,2020,40(S1):92-97.