计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2018.20180126
55(9):1843-1852,2018

面向时序数据异常检测的可视分析综述

韩东明　郭方舟　潘嘉铖　郑文庭　陈　为
(CAD & CG 国家重点实验室(浙江大学)　杭州　310058)
(dongminghan@zju.edu.cn)

Visual Analysis for Anomaly Detection in Time-Series: A Survey

Han Dongming, Guo Fangzhou, Pan Jiacheng, Zheng Wenting, and Chen Wei
(*State Key Laboratory of CAD & CG (Zhejiang University), Hangzhou* 310058)

Abstract　Anomaly detection for time-series denotes the detection and analysis of abnormal and unusual patterns, trends and features. Automatic methods sometimes fail to detect anomalies that are subtle, fuzzy or uncertain, while visual analysis can overcome this challenge by integrating the capability of human users and data mining approaches through visual representations of the data and visual interface. In this paper, we identify the challenges of anomaly detection, and describe the existing works of visual analysis along two categories: types of anomalies (attributes, topologies and hybrids), and anomaly detection means (direct projection, clustering and machine learning). We highlight future research directions.

Key words　anomaly detection; visual analysis; visualization; time-series data; data mining

摘　要　时序数据中的异常检测指的是在时序上去检测分析数据中异常的特征、趋势或模式.自动化的异常检测方法常会忽略细微的、模糊的、不确定的异常.可视分析通过对数据的可视表达和可视界面,集成用户和数据挖掘的能力.首先总结异常检测的挑战;然后从异常类型(属性、拓扑和混合)和异常检测方法(直接投影法、聚类方法和机器学习方法)2个角度对面向时序数据异常检测的可视分析工作进行分类和总结;最后阐述了未来的研究方向.

关键词　异常检测;可视分析;可视化;时序数据;数据挖掘

中图法分类号　TP391

时间相关的数据集无处不在,对其分析的关键是识别其中的模型、趋势和相关性[1].时序数据的可视分析广泛地应用在科学、工程和商业领域中[2-3].例如社交媒体[4]、城市数据[5]、电子交易[6]、时序排名[7].在不同领域中,发现时序数据中的特征和趋势的需求正日益增长,刺激了许多可视分析交互探索工具的发展[8]:Line Graph Explore[9],LiveRAC[2],SignalLens[10]和 Data Vases[11]等.

时序数据的可视分析任务,包括特征提取[12]、相关性分析和聚类[7]、模式识别[9]、异常检测[10]等.

收稿日期:2018-02-21;**修回日期**:2018-07-12

基金项目:国家重点研发计划项目(2018YFB0904503);国家"九七三"重点基础研究发展计划基金项目(2015CB352503);国家自然科学基金优秀青年科学基金项目(61422211);国家自然科学基金项目(61772456,61761136020)
This work was supported by the National Key Research and Development Program of China (2018YFB0904503), the National Basic Research Program of China (973 program) (2015CB352503), the National Natural Science Foundation of China for Excellent Young Scientists (61422211), and the National Natural Science Foundation of China (61772456, 61761136020).

通信作者:郑文庭(wtzheng@cad.zju.edu.cn)

而异常检测在不同的研究领域都是一个重要的问题.异常检测的目的是找到某些观察结果,它与其他的观察结果有很大的偏差,这样的偏差很有可能是由于不同的原因或机制所产生的[13-14].对应到不同的领域中,网络安全中的异常表示网络设备异常或者可疑的网络状态[15].情感分析中的异常表示一组数据中反常的观点、情绪模式或者产生这些模式的特殊时机[16].社交媒体中的异常可以是反常的用户,例如网络机器人[17].或是反常的信息传播过程,例如谣言的传播[18].这些异常信息或模式,例如电脑侵入、社交机器人、道路拥堵状况等都会极大地影响日常生活、社会稳定、科技发展.识别这些异常有助于及时认清实际状况,找出产生原因,进一步分析解决问题.

异常检测已经有许多成熟的方法,机器学习领域也已经提出了许多异常检测方法,包括有监督[19]和无监督的异常检测方法[20].自动化的学习算法通常假设有充足的训练数据可用,同时这些数据反映的是正常的行为;否则,很有可能因为新的观测数据是不常见的正常事件[21]而导致分类错误.机器学习中需要大量人工标注数据,这些数据的收集和标记都费时费力,同时标记过程十分依赖人的主观判断.这些因素对最后分析结果的质量会产生很大的偏差和影响[17].与此同时,如何在自然数据中定义其中正常或异常的行为也是十分困难的[22].此时人类的经验和知识在异常检测方面便显现出了优势,它可以被用来更新改进模型以及对异常检测过程进行实时控制.

如今的大数据时代,面对数据维度多、数据尺寸大的场景,数据可视分析可以帮助人来分析数据,理解其中的行为、模式等.交互式可视分析系统可以将异常检测方法和人的智能结合起来,帮助人们发现从未想到的异常,减少人的劳动,提高异常的检测识别能力.

1 挑 战

异常检测的挑战在 Chandola 等人[13]的综述中,已经进行了全面的总结.在检测数据中的异常前,首先要给出异常的定义,但正常和异常的界限往往是难以区分的.在一些设定阈值的异常检测方法中,往往很难判断在阈值附近的数据是否属于异常,需要其他的信息进行辅助判断.异常检测的 5 种挑战为:

1) 异常数据有可能模仿数据中的正常数据.有些异常的行为通常是人为恶意操控的,它会模仿现实中真实正常的行为,让异常的现象观测起来和正常现象一样,导致异常检测的任务变得十分困难.例如社交网络中机器人[17]回复,它会模仿真实人类的语气、时间频率等特征,以假乱真.

2) 异常的定义在不断变化.随着发展和进步,许多领域的正常行为也在与时俱进,其概念可能在未来会失效.而且很多数据集都是复杂和动态的,例如传感器数据[21,23-24]、网络安全数据[25]等,这些挑战在 FluxFlow[18]中都有提到.而在复杂多变时效性很高的场景中,需要人的监督和判断来进行异常检测.

3) 异常与领域高度相关.领域间的技术很难互相应用,不同领域间的实际情况不一样,有些异常的现象在其他领域可能就是正常的情况.

4) 带有标记的异常数据难以获取.用于训练确认异常的模型所使用的标记数据十分难以获取,人工的标注费时费力.

5) 异常和噪声具有一定的相似性.如何去区分和清洗两者也是面临的挑战之一.

2 异常分类

时序数据在不同的环境和应用领域中,会包含许多领域相关的信息及属性.例如网络数据中流量节点的类型、社交媒体上个人的信息[18]、注册时间等信息都可以视为时序数据中的属性.时序数据的异常情况便可以通过附属在时间维度的属性来进行分析检测.除此之外,数据中实体间蕴含的关系,即拓扑结构,也会揭示时序上的异常情况.例如社交网络中信息转发回复的会话网络[18]、动态图中网络演变[26]的数据及场景.

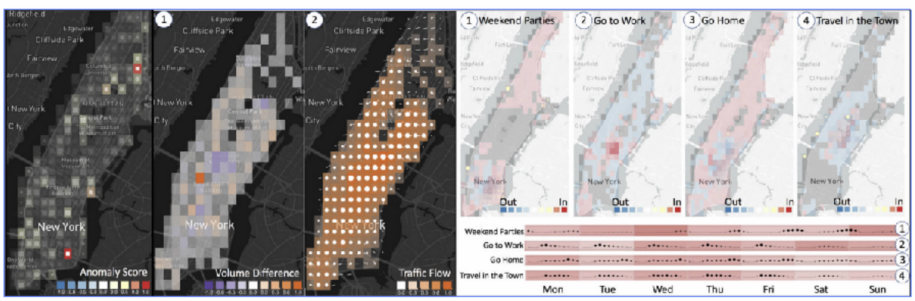
时序数据异常检测可视分析中,可基于拓扑结构或属性来对异常进行分析检测.下面将从 3 个角度去对已有的时序数据异常检测可视分析工作进行分类:1)属性上,例如时空数据中的地理信息;2)拓扑结构上,例如传感器网络的传输顺序;3)混合情况,例如网络数据中的节点属性和拓扑结构.属性和拓扑结构在时序上变化多端,规律难寻.而通过可视分析的方法来探究时序数据中的异常,与人的知识和经验相结合,将会有事半功倍的效果.

2.1 属性异常

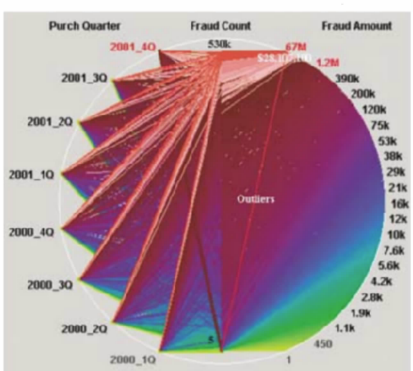
本节介绍基于时序数据中的属性进行异常检测

的相关可视分析工作,主要是对时序上属性变化的异常进行可视分析.

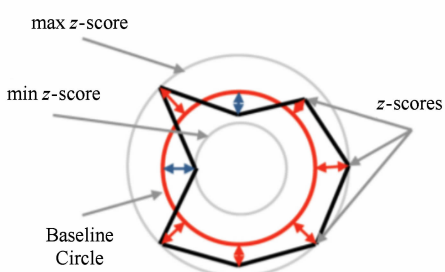
Thom 等人[27]的工作中根据信息的内容和发送的地理位置等属性进行聚类,形成标签云,用来可视分析时空数据中异常情况,例如地震、骚动等. Schreck 等人[28]的工作则是将实时数据和历史数据进行对比,包括频率、内容等多种属性,探寻时序上的属性异常. Onut 等人[29]针对网络流量异常进行检测,基于不同时间内的属性例如 DNS,HTTP 等请求数目或流量,检测相对应的异常情况,例如 DNS 攻击、探测攻击等. Hao 等人[30]对运营数据中的异常流程进行分析,如图 1(b)所示,通过在弦图中可视每两个实例之间的关系,设定阈值以及通过判定关系之间的交叉和时序上的变化,进行异常如诈骗的检测. Voila[31]系统提出了一种基于张量的异常分析算法用于转换时空数据,如出租车行驶数据中时序变化的交通流量、位置等信息,如图 1(a)所示,得到的张量时间序列再去结合历史数据进行期望模式分析,之后对张量分解后的上下文进行异常检测,同时可以根据用户的交互对异常的模式进行排序. Riveiro 等人[21]基于传感器数据来检测船只的异常状况,例如船的速度值异常、船舶靠近海岸线等历史数据中匹配不到的异常行为等.



(a) Visual analysis for anomaly detection in streaming spatiotemporal data



(b) Visual analysis for anomaly detecion in business data



(c) Visual analysis for anomaly detection in multivariate data

Fig. 1 Visual analysis for anomaly detection in time-series on attribute

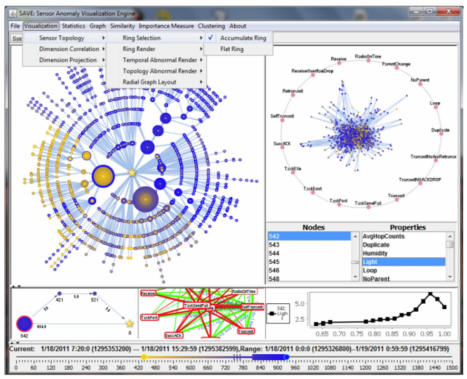
图 1 检测时序数据属性异常的可视分析工作.

2.2 拓扑异常

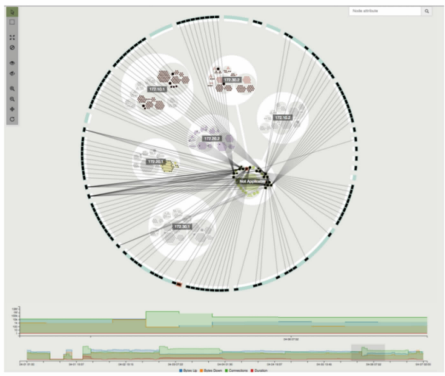
本节介绍基于时序数据中的拓扑结构进行异常检测的相关可视分析工作,即原始数据中存在拓扑结构,或者把数据抽象成拓扑结构,进而对拓扑结构在时序上的异常进行可视分析.

SAVE[23]是一个用于检测传感器数据异常变化的时序拓展模型(TEM),在拓扑视图中如图 2(a)所示,把传感器节点置于环形布局上,用不同颜色表示时间的先后顺序,可以观测传感器时序上拓扑结构的变化,进而发现异常. PolicyVis[32]系统用于防火墙安全策略的可视监测,巧妙地把安全策略规则命令的逻辑顺序转化成拓扑结构,不同的异常对应不同的拓扑结构,按照时间和流量等维度绘制矩形去代表相应策略及规则,矩形间的重合所带来的阴影则代表了不同的情况,其中就包含如策略冲突等异常情况.

Ocelot[33]应用了一种新型的层次结构和节点连接相结合的可视分析方法,如图 2(b)地所示. 系统可以用来检测网络流量中异常,并灵活地创建相关维度属性之间的响应.



(a) Visual analysis for anomaly detection in sensor data



(b) Visual analysis for anomaly detection in network data

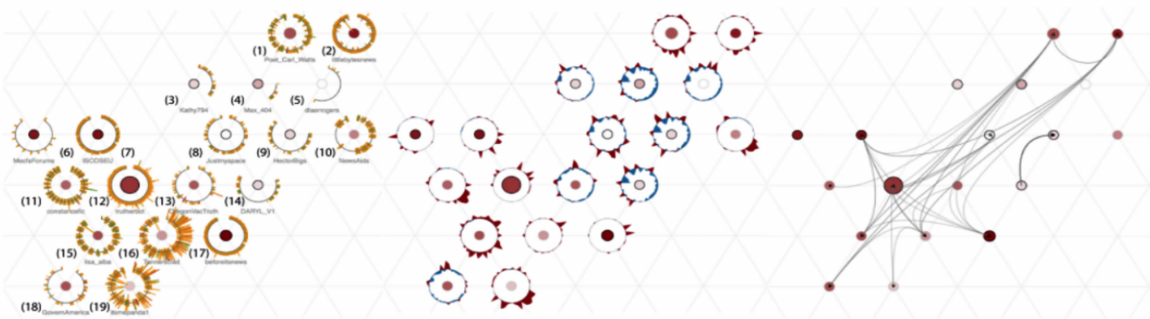
Fig. 2 Visual analysis for anomaly detection in time-series on topology

图 2 检测时序数据拓扑异常的可视分析工作

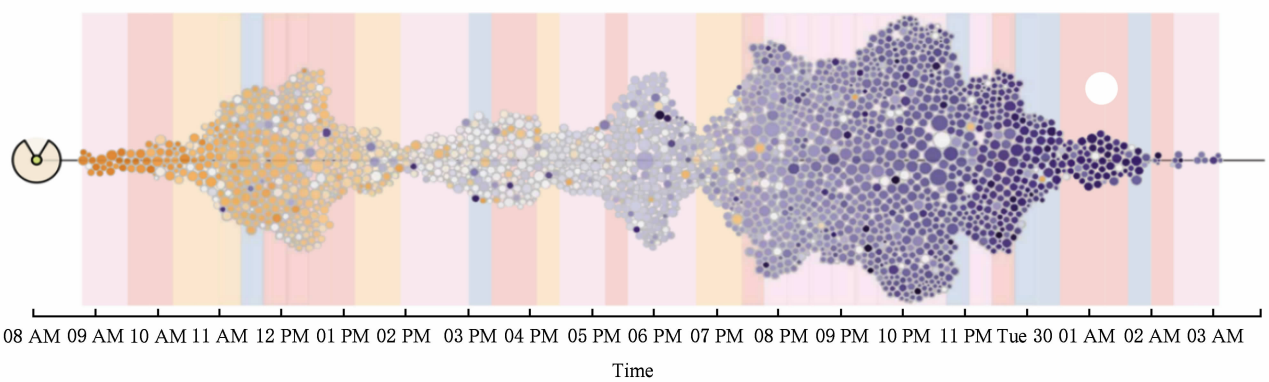
2.3 混合异常

本节介绍结合时序数据中的拓扑结构和属性，进行异常检测的相关可视分析工作，

Twitter 上消息的回复转发构成了会话网络，FluxFlow[18] 系统用于对会话传播过程进行异常检测、探索、解释. 它的分析模块结合了多种机器学习算法组来探索异常转发的特征. 例如用户注册的时间，朋友数量等属性上的特征. 图 3(a)设计了一个时序上的包裹圆，用于展示原始信息随着时间的推移在用户之间传播的可视分析视图，并通过 OCCRF[34] 模型来计算其中每一个圆(用户)的异常分数，其中圆的大小编码了用户的重要性. 结合拓扑结构和属性上的特征进行异常检测. 如图 3(b)所示，TargetVue[17] 用于检测社交网络中具有异常行为的用户，系统最初为每个用户提取一组行为特征，并使用异常检测算法来发现可疑的用户. 最可疑的用户将会用 2 种类型的图标进行可视分析编码：1)他们的通信行为(即发布消息，转发消息)；2)与之相应的行为的特征.



(a) Visual analysis for anomaly detection in user behaviors



(b) Visual analysis for anomaly detection in social media

Fig. 3 Visual analysis for anomaly detection in time-series on hybridize

图 3 检测时序数据混合异常的可视分析工作.

Liao 等人[35]的工作中研究管理网络数据中的动态异常问题,从多个层面和角度去关注节点、关系和社团在时序上的异常变化.

3 异常检测方法分类

本节根据异常检测的方法,对已有的时序数据异常检测可视分析的工作进行分类.异常检测方法在[13,36-38]中被介绍了许多,针对不同的领域和不同的类型数据,异常检测方法也不尽相同,同一种异常检测方法在面对不同实际情况时,效果可能会大打折扣.结合异常检测算法,加上可视分析方法和人类的知识经验,可以更加精确,多角度地检测出以前难以通过自动化算法检测出的异常.下面将时序数据异常检测的可视分析工作分成 3 个大类:直接投影法,聚类方法和机器学习方法

3.1 直接投影法

本节介绍通过直接投影法进行异常检测的相关时序数据可视分析工作.

投影在数学中定义为图形的影子投到一个面或一条线上,这里表示对数据的数值特征直接进行可视化,可视分析流程较短,例如盒须图、直方图等可视化方法都可以直接看出数据中的异常信息.

Celenk 等人[39]基于短期的网络特征和平均时间熵的观测结果,针对异常网络流量设计了 FLD 图,用于对网络异常中统计得到的特征进行可视分析.Z-Glyph[22]被设计用于检测多元数据中的离群点,如图 1(c)所示,可以配合多个统一视图(small multiples)进行时序上异常检测.同样可以用来检测离群点的方法还包括平行坐标轴[40]的方法、盒须图[41]等.Sun 等人[42]设计了与时空属性相结合、基于仿真数据的盒须图,用于对例如环境降雨、全球环流、海面温度等数据进行异常检测.

Onut 等人[29]针对网络流量进行异常检测.对不同时间内的不同类型流量进行可视分析,投影到六边形视图上.每种异常都会有特定的特征可视在六边形视图上,例如 DoS 攻击、探测攻击等异常.

2004 年 PCA[43]第 1 次用于流量异常的检测,之后 Brauckhoff 等人[44]提出了一个解决方案用于处理数据中时间相关性的问题,让 PCA 可以更好地应用于异常检测.此外还有基于直方图[45]、最大熵估计[46]等方法被用于此项分析任务.PCA 在时空数据的异常检测中经常被用到,但是数据中如果包含张量结构,PCA 方法可能会隐藏一些有特殊意义的异常.Gama 等人[47]介绍了基于张量来检测识别异常的新技术手段.Dereszynski 等人[48]用三阶张量来表示动态变化的社交网络,并从中检测异常的演变.Voila[31]用张量来表示时空数据,根据动态的张量网络来结合多个统一视图和图标(glyph)来可视分析车流交通的异常情况,并通过张量分解从数据中推测出异常的特征和模式.

直接投影法应用于时序数据的异常检测,任务往往比较简单,用于结果展示、结论推断,直接投影法往往要与其他异常检测方法和多种可视化方法相结合,才可以完成复杂的异常检测分析任务,因为方法多样,针对解决的问题不同,所以如何对一个特定的场景或问题去选择效果最好的直接投影方法是非常困难的.

3.2 聚类方法

本节介绍通过聚类的方法进行异常检测的相关时序数据可视分析工作.

聚类表示把集合中的相似对象分成多个类.正常的数据会被分为一个或多个类,而异常的数据距离这些类比较远.正常数据的聚类中通常是多而密集的,而异常数据的聚类是小而稀疏的.

Thom 等人[27]提出了针对于连续数据流,例如微博帖子的增强 Lloyd 聚类方法.设计了一个系统,其中包括用于探索捕获重要事件及异常的分析工作台,和针对时空数据的异常检测模块.系统可以在地图视图上根据位置去布局聚类之后的标签.

Kim 等人[49]使用了结合图可视分析的最小生成树聚类方法,在一些时序数据的线性回归中进行异常检测,并提出多种用于分类异常的图形特征.Malinao 等人[50]用 X-means 聚类算法,与可视分析相结合,对时空交通数据的流量变化进行包括异常检测在内的多种任务的分析.Imoto 等人[51]提出一种基于折线图的 3 维时变可视分析技术,用于提取、显示相似的值,并应用 SAX(symbolic aggregate approximation)来检测其中频繁或异常的模式.Lin 等人[52]也是通过聚类一组时序变化的值来检测其中的异常值.Hao 等人[30]使用多个统一视图以及弦图对商业管理数据多个维度的影响因子进行可视分析,并运用相关性分析,局部匹配以及聚类的技术来提取其中重要的影响因素.

聚类方法应用于时序数据的异常检测,适用于正常数据偏多、并与异常数据差距较大数据集的分析任务中.聚类也经常要与其他异常检测方法和多

种可视化方法相结合. 而对于特定的场景或问题,如何选择聚类方法和如何调节聚类方法中参数也是非常困难的.

3.3 机器学习方法

本节介绍通过机器学习的方法进行异常检测的相关时序数据可视分析工作.

机器学习的方法可以从大量的数据中提取到相关的特征,进而可以从数据中区分出异常的数据. 机器学习方法通常用于预处理阶段.

SOM[53]可以看做是一个基于神经网络的聚类算法[54],高维数据投影到 2D 空间后可以识别定义不同聚类间的边界[55],许多工作都会加入一个高维影视图用于数据集分类的总览. 然而对于很难明确的数据分类情况时,SOM 并不能提供一个效果很好的解决方案时,GMM 便可以应用于 SOM 上. Riveiro 等人[21]根据历史的观察数据建立正常的模型,之后对新生的实时数据进行异常行为检测. 如果计算的累计概率值高于设定阈值便通知用户进行判断,对模型进行迭代,其中的训练器和检查器是基于 GMM 和 SOM 来完成的. 系统中包括一个用于过滤设定参数的控制视图,例如传感器的选取、覆盖范围过滤、累计的概率值阈值调整等. 同时设计了一个概览地图用于标注船舶的类型、速度等属性. 当异常发生时会在地图上进行异常标注,人为判断其是否为异常情况,从而对模型进行迭代.

文献[17-18]中基于 OCCRF 模型来对 Twitter 数据中的异常转发等用户行为进行异常检测,结合多种机器学习算法探索社交网络转发过程的重要特征.

机器学习方法应用于时序数据的异常检测,可以对历史数据抽取特征,建立模型,特别适用于数据量大、维度多的分析任务中. 但是需要数据足够多,才可以进行更准确的分类. 而监督学习模型中需要有人工标注好的数据,这些都限制了许多场景下的机器学习方法应用.

3.4 其他方法

还有一些工作运用一些例如图标法、地理视图、矩阵图等直接可视分析的方法对时序数据进行异常检测. 例如通过颜色的差异[32]、地理视图上[28]的高亮、多种统计图中相关的方法来显示时序数据中的异常. Schreck 等人[28]将实时数据和历史数据的内容及频率等属性进行对比,在地理视图上对异常的属性进行高亮.

4 总结和未来发展

本文对于时序数据的异常检测可视分析工作进行了综述,并针对现有的工作对时序数据的异常类型如表 1 所示和异常检测方法如表 2 所示分别进行了总结. 表 3 对相关工作进行了分类总结. 近年来的

Table 1 Classification of Anomaly and Relevant Visual Analysis Tasks in Visualization Forms

表 1 异常分类和相关的可视分析任务以及可视表达

Type	References	Data	Visualization Forms
Attribute	Ref [21]	Sensor Data	3D, physical map, parallel coordinates, scatter plot, histogram
	Ref [22,27-28]	Social Media	physical map, tag cloud, glyph
	Ref [25,29,39,45-46,48]	Network Traffic	3D, physical map, histogram, matrix, glyph, small multiples, Line chart, stacked chart
	Ref [30]	Business Data	Chord diagram
	Ref [31,50]	Traffic Data	Glyph, small multiples, volume map, line chart, Histogram, scatter plot
	Ref [42]	Weather Data	Boxplots
	Ref [49,52]		Line chart, tree map
	Ref [51]	Temperature Data	Line chart, volume map
Topology	Ref [23]	Sensor Data	Node-link diagram, line chart
	Ref [26,33]	Network Traffic	Node-link diagram, circle packing, line chart
	Ref [32]	Firewall Policy	
Hybridize	Ref [17-18]	Social Media	Glpyh, flow, node-link diagram, heatmap
	Ref [35]	Network Management	Node-link diagram, line chart
	Ref [56]	Traffic Data	Histogram, volume map, line chart

Table 2 Classification of Anomaly Detection Methods and Relevant Visual Analysis Tasks in Visualization Forms

表 2 异常检测方法分类和相关的可视分析任务以及可视表达

Type	References	Data	Visualization forms
Direct Projection	Ref [22,48]	Social Media	Glyph
	Ref [26,29,31,33,39,45-46]	Network Traffic	Glyph, small multiples,3D, line chart, stacked chart, Node-link diagram, circle packing
	Ref [42]	Weather Data	Boxplots
	Ref [56]	Traffic Data	Physical map, line chart, volume map, histogram
Clustering	Ref [23]	Sensor Data	Node-link diagram, line chart
	Ref [27]	Social Media	Physical map, tag cloud
	Ref [30]	Business Data	Chord diagram
	Ref [35]	Network Management	Node-link diagram, line chart
	Ref [49,52]		Line chart, tree map
	Ref [50]	Traffic Data	Scatter plot
	Ref [51]	Temperature Data	Line chart, volume map
Machine Learning	Ref [17-18]	Social Media	Glyph, flow map, heatmap, node-link diagram
	Ref [21]	Sensor Data	3D, physical map, parallel coordinates, scatter plot, histogram
Others	Ref [25]	Network Traffic	Physical map, histogram,matrix
	Ref [28]	Social Media	Physical map
	Ref [32]	Firewall Policy	

Table 3 Classification of Relevant Papers

表 3 相关工作分类

Type	Direct Projection	Clustering	Machine Learning	Others
Attribute	Ref [22,29,31,42,45-46,48]	Ref [27,30,49-52]	Ref [21]	Ref [25-28]
Topology	Ref [26-33]	Ref [23]		Ref [32]
Hybridize	Ref [56]	Ref [35]	Ref [17-18]	

纯拓扑或拓扑相关的工作很少,更多的是属性上的异常检测可视化工作,而应用机器学习方法的相关工作也较少.拓扑的异常经常要通过属性来进行辅助判断,同时也很少有大规模异构数据在拓扑上的异常检测可视化工作,许多工作的针对性较强,很难适用于其他的场景或数据集上的异常检测任务.

异常检测方法中,直接投影法适用于简单的分析任务,或者是做结果展示,但是对于大规模多维异构数据,很难找到合适的手段去应对,需要背景知识和专家经验.聚类方法适用正常数据偏多、与异常数据差距较大的数据分析任务中,而不同聚类方法的选择和调参没有一个很好的指导方法.机器学习算法适合于大规模数据且有标注数据的异常检测任务,但是人工数据标注引入的误差、时间和人力成本带来了许多局限性.

近年来,随着可视分析方法在时序数据异常检测上的不断应用,逐步展现出可视分析的巨大优势.例如在社交领域[57]中,因为现有的自动化方法的内在局限性,通过可视分析来检测异常的用户行为将会是一个很有前途的方向.但是时序数据中的异常检测仍然存在一些挑战,对于使用直接投影方法的挑战,例如如何自动化地针对数据的特点去推荐直接投影方法完成异常检测分析任务、如何在大规模多维异构数据集中使用适合的直接投影方法等.对于使用聚类方法的挑战,例如如何自动化推荐合适的聚类算法、聚类参数去应对不同数据集下的异常检测分析任务;如何与其他异常检测方法间的结合应用等;对于使用机器学习方面的挑战,例如如何使用少量标注数据去完成异常检测任务中的模型等.总之不同场景,数据集中异常检测方法的使用面向大众的异常检测可视分析系统的开发、大规模多维异构数据的多种可视分析任务、随时代变化的异常判定标准和类型等问题和挑战都是未来可以继续深入研究的方向.

时序数据异常检测的任务特别需要人的经验和背景知识进行分析判断.可视分析则可以更好地对数据进行抽象、表达、发现自动算法不能判别的异常情况.自动化的异常检测方法、人的知识经验和交互式可视分析方法,三者相互结合,可以更准确智能地检测异常.将会给人们的生活、国家的进步、社会的发展带来了巨大的安全保障.

参 考 文 献

[1] Keim D A, Nietzschmann T, Schelwies N, et al. A spectral visualization system for analyzing financial time series data [C] //Proc of Eurographics/IEEE TCVG Symp on Visualization. Aire-la-Ville, Switzerland: Eurographics Associatio, 2006: 195-202

[2] McLachlan P, Munzner T, Koutsofios E, et al. LiveRAC: Interactive visual exploration of system management time-series data [C] //Proc of the ACM SIGCHI Conf on Human Factors in Computing Systems. New York: ACM, 2008: 1483-1492

[3] Weber M, Alexa M, Müller W. Visualizing time-series on spirals [C] //Proc of the 2001 IEEE Symp on Information Visualization. Los Alamitos, CA: IEEE Computer Society, 2001: 7-14

[4] Kumar P, Sinha A. Real-time analysis and visualization of online social media dynamics [C] //Proc of the 2nd Int Conf on Next Generation Computing Technologies. Piscataway, NJ: IEEE, 2016: 362-367

[5] Chen Wei, Huang Zhaosong, Wu Feiran, et al. VAUD: A visual analysis approach for exploring spatio-temporal urban data [J]. IEEE Trans on Visualization and Computer Graphics, 2018, 24(9): 2636-2648

[6] Xie Cong, Chen Wei, Huang Xinxin, et al. Vaet: A visual analytics approach for e-transactions time-series [J]. IEEE Trans on Visualization and Computer Graphics, 2014, 20 (12): 1743-1752

[7] Xia Jing, Hou Yumeng, Chen Y V, et al. Visualizing rank time series of Wikipedia top-viewed pages [J]. IEEE Computer Graphics and Applications, 2017, 37(2): 42-53

[8] Cho M, Kim B, Bae H J, et al. Stroscope: Multi-scale visualization of irregularly measured time-series data [J]. IEEE Trans on Visualization and Computer Graphics, 2014, 20(5): 808-821

[9] Kincaid R, Lam H. Line graph explorer: Scalable display of line graphs using focus+ context [C] //Proc of the Working Conf on Advanced Visual Interfaces. New York: ACM, 2006: 404-411

[10] Kincaid R. Signallens: Focus+ context applied to electronic time series [J]. IEEE Trans on Visualization and Computer Graphics, 2010, 16(6): 900-907

[11] Thakur S, Rhyne T M. Data vases: 2d and 3d plots for visualizing multiple time series [C] //Proc of 2009 Int Symp on Visual Computing. Berlin: Springer, 2009: 929-938

[12] Alonso O, Khandelwal K. Kondenzer: Exploration and visualization of archived social media [C] //Proc of the 30th IEEE Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2014: 1202-1205

[13] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey [J]. ACM Computing Surveys, 2009, 41(3): 15:1-15: 58

[14] Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying density-based local outliers [C] //Proc of the ACM SIGMOD Record. New York: ACM, 2000: 93-104

[15] Pearlman J, Rheingans P. Visualizing network security events using compound glyphs from a service-oriented perspective [C] //Proc of the Workshop on Visualization for Computer Security. Berlin: Springer, 2007: 131-146

[16] Wang Zhaoxia, Joo V, Tong Chuan, et al. Anomaly detection through enhanced sentiment analysis on social media data [C] //Proc of the 6th IEEE Int Conf on Cloud Computing Technology and Science. Piscataway, NJ: IEEE, 2014: 917-922

[17] Cao Nan, Shi Conglei, Lin S, et al. TargetVue: Visual analysis of anomalous user behaviors in online communication systems [J]. IEEE Trans on Visualization and Computer Graphics, 2016, 22(1): 280-289

[18] Zhao Jian, Cao Nan, Wen Zhen, et al. # FluxFlow: Visual analysis of anomalous information spreading on social media [J]. IEEE Trans on Visualization and Computer Graphics, 2014, 20(12): 1773-1782

[19] Steinwart I, Hush D, Scovel C. A classification framework for anomaly detection [J]. Journal of Machine Learning Research, 2005, 6: 211-232

[20] Eskin E, Arnold A, Prerau M, et al. A geometric framework for unsupervised anomaly detection [M]. Berlin: Springer, 2002: 77-102

[21] Riveiro M, Falkman G, Ziemke T. Improving maritime anomaly detection and situation awareness through interactive visualization [C] //Proc of the 11th IEEE Int Conf on Information Fusion. Piscataway, NJ: IEEE, 2008: 1-8

[22] Cao Nan, Lin Yuru, Gotz D, et al. Z-Glyph: Visualizing outliers in multivariate data [J]. Information Visualization, 2017, 17(1): 22-40

[23] Shi Lei, Liao Qi, He Yuan, et al. SAVE: Sensor anomaly visualization engine [C] //Proc of 2011 IEEE Conf on Visual Analytics Science and Technology. Piscataway, NJ: IEEE, 2011: 201-210

[24] Wang X R, Lizier J, Obst O, et al. Spatiotemporal anomaly detection in gas monitoring sensor networks [M]. Berlin: Springer, 2008: 90-105

[25] McKenna S, Staheli D, Fulcher C, et al. Bubblenet: A cyber security dashboard for visualizing patterns [J]. Computer Graphics Forum, 2016, 35(3): 281-290

[26] Zhang Haocheng, Wu Xiaojie, Tang Xiang, et al. System detecting network anomaly with visualization techniques [J]. Chinese Journal of Network and Information Security, 2018, 4(2): 40-54 (in Chinese)
(张浩城, 吴晓洁, 唐翔, 等. 基于可视分析的网络异常检测系统[J]. 网络与信息安全学报, 2018, 4(2): 40-54)

[27] Thom D, Bosch H, Koch S, et al. Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages [C] //Proc of 2012 Pacific Visualization Symp. Piscataway, NJ: IEEE, 2012: 41-48

[28] Schreck T, Keim D. Visual analysis of social media data [J]. Computer, 2013, 46(5): 68-75

[29] Onut I V, Zhu Bin, Ghorbani A A. A novel visualization technique for network anomaly detection [C/OL] //Proc of the 2nd Annual Conf on Privacy, Security and Trust, 2004: 167-174. [2018-02-01]. http://citeseerx. ist. psu. edu/viewdoc/download?doi=10. 1. 1. 83. 5029&rep=rep1&type=pdf

[30] Hao M C, Keim D A, Dayal U, et al. Business process impact visualization and anomaly detection [J]. Information Visualization, 2006, 5(1): 15-27

[31] Cao Nan, Lin Chaoguang, Zhu Qiuhan, et al. Voila: Visual anomaly detection and monitoring with streaming spatiotemporal data [J]. IEEE Trans on Visualization and Computer Graphics, 2018, 24(1): 23-33

[32] Tran T, Al-Shaer E S, Boutaba R. PolicyVis: Firewall security policy visualization and inspection [C] //Proc of the 21th Large Installation System Administration Conf. Berkeley, CA: USENIX Association, 2007: 1-16

[33] Arendt D L, Burtner R, Best D M, et al. Ocelot: User-centered design of a decision support visualization for network quarantine [C] //Proc of 2015 IEEE Symp on Visualization for Cyber Security (VizSec). Piscataway, NJ: IEEE, 2015: 1-8

[34] Song Yale, Wen Zhen, Lin C Y, et al. One-class conditional random fields for sequential anomaly detection [C] //Proc of the 23rd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2013: 1685-1691

[35] Liao Qi, Striegel A. Intelligent network management using graph differential anomaly visualization [C] //Proc of 2012 Network Operations and Management Symp. Piscataway, NJ: IEEE, 2012: 1008-1014

[36] Hodge V, Austin J. A survey of outlier detection methodologies [J]. Artificial Intelligence Review, 2004, 22(2): 85-126

[37] Zhang Tianye, Wang Xumeng, Li Zongzhuang, et al. A survey of network anomaly visualization [J]. Science China Information Sciences, 2017, 60(12): 121101

[38] Zhao Ying, Fan Xiaoping, Zhou Fangfang, et al. A survey on network security data visualization [J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(5): 687-697 (in Chinese)
(赵颖, 樊晓平, 周芳芳, 等. 网络安全数据可视化综述[J]. 计算机辅助设计与图形学学报, 2014, 26(5): 687-697)

[39] Celenk M, Conley T, Willis J, et al. Predictive network anomaly detection and visualization [J]. IEEE Trans on Information Forensics and Security, 2010, 5(2): 288-299

[40] Novotny M, Hauser H. Outlier-preserving focus+context visualization in parallel coordinates [J]. IEEE Trans on Visualization and Computer Graphics, 2006, 12(5): 893-900

[41] Williamson D F, Parker R A, Kendrick J S. The box plot: A simple visual method to interpret data [J]. Annals of Internal Medicine, 1989, 110(11): 916-921

[42] Sun Ying, Genton M G. Adjusted functional boxplots for spatio-temporal data visualization and outlier detection [J]. Environmetrics, 2012, 23(1): 54-64

[43] Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies [C] //Proc of the 2004 ACM SIGCOMM Computer Communication Review. New York: ACM, 2004: 219-230

[44] Brauckhoff D, Salamatian K, May M. Applying PCA for traffic anomaly detection: Problems and solutions [C] //Proc of INFOCOM 2009. Piscataway, NJ: IEEE, 2009: 2866-2870

[45] Kind A, Stoecklin M P, Dimitropoulos X. Histogram-based traffic anomaly detection [J]. IEEE Trans on Network and Service Management, 2009, 6(2): 110-121

[46] Gu Yu, McCallum A, Towsley D. Detecting anomalies in network traffic using maximum entropy estimation [C] //Proc of the 5th ACM SIGCOMM Conf on Internet Measurement. Berkeley, CA: USENIX Association, 2005: 32-32

[47] Fanaee-T H, Gama J. Tensor-based anomaly detection: An interdisciplinary survey [J]. Knowledge-Based Systems, 2016, 98: 130-147

[48] Dereszynski E W, Dietterich T G. Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns [J]. ACM Trans on Sensor Networks, 2011, 8(1): 3:1-3:36

[49] Kim S S, Krzanowski W J. Detecting multiple outliers in linear regression using a cluster method combined with graphical visualization [J]. Computational Statistics, 2007, 22(1): 109-119

[50] Malinao J A, Juayong R A B, Becerral J G, et al. Patterns and outlier analysis of traffic flow using data signatures via IDIRBrG method and vector fusion visualization [C] //Proc of the 3rd Int Conf on Human-Centric Computing. Piscataway, NJ: IEEE, 2010: 1-6

[51] Imoto M, Itoh T. A 3D visualization technique for large scale time-varying data [C] //Proc of the 14th Int Conf on Information Visualisation. Piscataway, NJ: IEEE, 2010: 17-22

[52] Lin J, Keogh E, Lonardi S. Visualizing and discovering non-trivial patterns in large time series databases [J]. Information visualization, 2005, 4(2): 61-82

[53] Munoz A, Muruzábal J. Self-organizing maps for outlier detection [J]. Neurocomputing, 1998, 18(1): 33-60

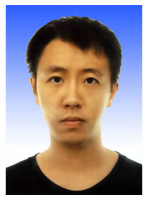
[54] Kohonen T. The self-organizing map [J]. Proceedings of the IEEE, 1990, 78(9): 1464-1480

[55] Kraiman J B, Arouh S L, Webb M L. Automated anomaly detection processor [J]. Proceedings of SPIE: Enabling Technologies for Simulation Science VI, 2002, 4716: 128-137

[56] Shekhar S, Lu C T, Liu R, et al. CubeView: A system for traffic data visualization [C] //Proc of the 5th IEEE Int Conf on Intelligent Transportation Systems. Piscataway, NJ: IEEE, 2002: 674-678

[57] Wu Yingcai, Cao Nan, Gotz D, et al. A survey on visual analytics of social media data [J]. IEEE Trans on Multimedia, 2016, 18(11): 2135-2148

Han Dongming, born in 1995. PhD candidate. His main research interests include visualization and visual analytics.



Guo Fangzhou, born in 1991. PhD candidate. His main research interests include visualization and visual analytics.

Pan Jiacheng, born in 1995. Master candidate. His main research interests include visualization and visual analytics.

Zheng Wenting, born in 1974. Associate professor. His main research interests include computer graphics and visualization.

Chen Wei, born in 1976. Professor. His main research interests include visualization and visual analysis.