

时间数据可视化综述

题 目 时间数据可视化综述

学 院 计算机学院

专 业 计算机技术

姓 名 李涵, 麻淑婕, 何龙, 白植权, 王文龙¹

学 号 222050239, 222050243, 221050008

学 号 221050020, 221050026

指导教师 余正生

¹ 5 位作者皆为共同作者，对此篇文章同等贡献.

具体如下:

李涵: 摘要, 介绍

麻淑婕: 国内研究现状

何龙: 国外研究现状

研究方法: 白植权

总结: 王文龙□í

摘要

本综述旨在探讨时间数据可视化的最新进展和应用。时间数据可视化是一种利用图形、图表等方式展示数据随时间变化的特征和规律的方法。时间数据可视化可以帮助人们更直观、高效地分析和理解数据，从而发现数据背后的信息、知识和智慧，它帮助人们从时间序列数据中发现趋势、模式和关联。本综述涵盖了近年来在国内外在时间数据可视化领域取得的重要成果，包括常见领域的时序数据的处理和自动化技术。并介绍了一些常用的可视化技术和工具。

介绍

时序数据基本概念

在众多数据类型中，时序数据是常见的数据类型之一。这类数据按照时间顺序统计，可以对历史数据进行有效的预测管理。它们是按照时间顺序数据的集合，观测数据与时间之间的数学关系如下[2]

$$Q = ((q_1, t_1), (q_2, t_2), \dots, (q_i, t_i), \dots, (q_n, t_n)) (t_1 < t_2 < \dots < t_n)$$

其中:在时刻 t 观测数据为 q_i ，时间序列的长度为 n 。

由此可以看出时间序列数据就是相等时间间隔内（例如每小时）内变量（例如降雨量）值的有序序列。例如，带有时间戳的数据（例如数据库运行监测数据）都可以视为时间序列[2]。时序数据每隔一段时间便会大量产生[1]。

时间数据可视化

时间数据可视化的主要目标是通过图表、图形和动画等可视化手段，直观地展示时间数据的特征和变化趋势。

将时间事件序列数据可视分析任务分为以下 4 类。

模式发现与探索

时间事件序列数据可视化的一个重要任务就是发现数据集中的频繁模式，并且对这些模式进行聚合、分类[4,5]。这些模式反映了常见序列模式的特征及其演变(如软件 A 最常见的使用模式可能随着时间而变化)。此外，对感兴趣模式的探索分析可以帮助分析特殊事件或事件序列的前因后果。如 Fan 等[6]采用多视图交互的方

式探索手机用户 对手机的使用模式, 并分析导致特定事件发生的 原因(如文中发现打电话事件通常会导致蓝牙启动 事件的发生).

可视化查询

在时间事件序列数据分析中, 精确地识别和 探索与任务相关的序列或记录子集对后续精准分 析非常重要(如寻找并分析与目标学生相似的往届 学生学业事件序列, 为目标学生制定精准的学业 和职业计划[7]). 但是在许多情况下, 很多容易直 观描述的模式很难用特定的查询语言来表达, 并 且制定复杂的查询非常困难. 研究新型的可视化 查询模型可以快速响应复杂事件序列的查询, 同 时也为用户提供了一种直观的方法来验证指定的 查询是否与他们的预期相符。

对比分析

在时间事件序列数据可视分析中, 比较 2 组或 多组数据记录是一个常见的分析任务, 它可以从 简单的统计比较到大量的假设性检验[8]. 对比分 析中所需的数据群组可以根据不同的标准(如事件 的多维属性、时间周期、结果事件或某些特定的事件等)拆分得到. 如 Malik 等[8]将患者分为低成本 和高成本患者群体, 比较分析 2 个群体的治疗效 果, 探索高血压药物依从性是否会影响患者的 花费。

结果事件分析

在时间事件序列数据的可视分析中, 分析不 同的事件序列如何导致不同的结果(结果事件), 可 以帮助分析人员生成关于因果关系的假设, 为精 确查询(T2)、有效的学术研究项目、医疗计划或营 销活动提供个性化建议. 如 Franklin 等[9]分析了与 药物治疗结果(有效或无效)、治疗副作用(有无副作 用)相关的事件序列, 并总结了疾病的有效护理方 案, 帮助医务人员以及患者做出正确决策。

中国时序数据可视化的现状

据《2019 年中国大数据产业市场前景研究报告》显示, 随着各行业的数字化转型和数据应用的普及, 时序数据可视化在金融、交通、医疗等领域的应用也越来越广泛。例如[14]:

金融领域：时间序列图表被广泛用于股票走势预测、经济指标分析、投资组合管理等方面。例如，银行可以利用时序数据可视化技术实时监控客户账户活动，识别异常交易并及时做出反应。同时，在保险领域中，时序数据可视化在理赔管理、风险评估等方面也有重要应用。

交通领域：时序数据可视化可以帮助交通管理部门监测和掌握城市的交通流量变化，调整道路规划和交通信号灯设置，提高城市交通的运行效率。此外，时序数据可视化还可以用于研究车辆行驶路径、公共交通站点分布等问题。

医疗领域：时序数据可视化可以帮助医生和研究人员更好地了解患者的病情和治疗效果。例如，在慢性疾病管理中，医生可以利用时序数据可视化技术监测患者的生命体征、药物使用情况等，并根据数据分析结果调整治疗方案。此外，时序数据可视化还可以用于研究流行病学问题，如传染病爆发时的病例增长趋势。

近年来，中国在时序数据可视化领域也取得了一定的进展。国内的一些高校和科研机构都在积极开展时序数据可视化的相关研究，许多企业也开始关注并应用时序数据可视化技术。

1. 基于交互式的时序数据可视化技术

交互式是时序数据可视化技术中重要的一环。国内有许多研究人员致力于开发和改进基于交互式的时序数据可视化技术。该方法可以帮助用户深入挖掘时序数据中的规律和趋势，其核心思想是将探索数据的过程转化为一个交互式的过程，让用户可以随时调整自己的视角和偏好来获取更全面、准确的信息。这种基于交互式探索的时序序列可视化方法通过以下步骤实现：数据预处理、可视化设计、交互式操作、数据分析。

主题河流是由 HAVRS 等[15]提出的一种广泛应用于时序数据的交互可视化方式。对网上课程、研讨会、会议等在线视频采用主题河流进行可视化，提取视频的关键词作为主题以及在视频中该关键词的表现强度作为河流的宽度，用户可以根据自己的兴趣点跳跃观看视频，避免了费时观看整个视频或盲目快进导致的信息遗漏。但由于主题河流之间相互堆叠，单个主题河流的形状和波动走向会受到下方主题河流的干扰。陈馨葳等[1]在改善主题河流的读图歧义方面，采用交互的方式，为用户标识出歧义点，歧义点的筛选由用户动态选择，用户获得歧义点后可根据标识信息对河流局部进行重排序，消除局部歧义，实现用户个性化读图。

2. 时序数据可视化的自动化技术

韩福波[11]提出一种根据时序数据的变化趋势自动生成图表类型的可视化方法，通过计算数据本身的局部多项式回归拟合曲线与散点图和折线图之间的一致性，选择一致性最高的最有图表类型。

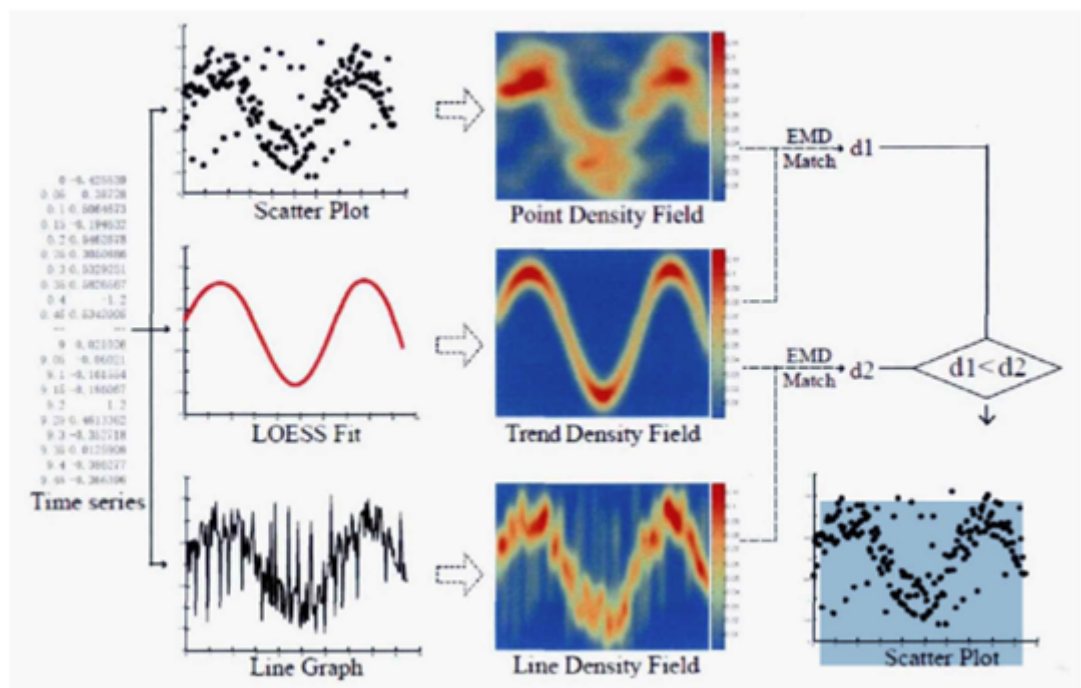


图 1 折线图和散点图选择算法流程图

3. 高维时序数据可视化技术

Line Chart 是出现较早的一个经典的时序数据可视化方法，这种方法主要将时序数据的每个数据点映射到一个二维坐标系中，并将所有的数据点用直线连接，X 轴表示时间上的变化，Y 轴表示数据点的其他属性变量。但是存在两种缺点：第一，当时序数据的数量规模较大的时候，折线之间将相互交错和叠加，造成视觉上的混乱感；第二，当数据的属性数量大于 2 时（多元时序数据），传统方法将无法对时序数据所有的属性进行展示。

在保留数据趋势特征的基础上，对于高维时序数据展示困难的问题，王子慧[12]提出了改进时间序列近似表示算法，综合考虑拟合误差和序列长度，提高了分段效率，准确高效地减少了时间维度。孙浩鑫[13]提出了基于动态时间弯曲距离的聚类算法对数据进行聚类，并将结果应用到主题河流可视化方法中，从而避免了传统可视化方法无法对大规模时序数据有较好的可视化的问题。

意义

揭示趋势和模式：时序数据可视化可以帮助我们发现趋势和模式。通过观察数据的走势，我们可以了解到某个现象的变化规律，从而对未来的趋势进行预测。

发现异常值：时序数据可视化可以帮助我们发现异常值。如果出现了不符合正常规律的数据点，那么在可视化图表中就会呈现出明显的异常情况，从而引起我们的注意。

对比不同时间段的数据：时序数据可视化还可以帮助我们对比不同时间段的数据。通过将不同时间段的数据放在一起，我们可以更加清晰地了解这些数据之间的差异和联系，并从中得出结论。

支持决策：时序数据可视化能够提供直观的数据展示和分析结果，因此可以支持决策。基于时序数据可视化的分析结果，我们可以做出更为准确和合理的决策。

国外时序数据可视化的现状

为了更好地利用时间数据，许多国外知名大学如麻省理工学院、哈佛大学等都将时间数据可视化作为研究课题，并且提供了丰富的方法和技术，用于有效地可视化和分析时间导向的数据。

Aigner 等[16]提出了一个层次化的模型，可以帮助用户对时间序列数据进行探索和分析。Heer 等[17]探讨了协作可视分析的设计考虑因素，并提供了一些关键的技术细节和原则，可以支持多人协作的数据分析和决策过程。Alsallakh 等[18]探讨了在高维空间中可视化集合交集的挑战，使用各种映射技术，例如多维缩放和降维方法，以将高维数据转换为二维或三维可视化。Brehmer 等[19]提出了抽象可视化任务的多层次分类方法，探讨了在时间数据可视化中涉及的不同抽象任务类型，并提供了对每个任务类型的详细描述和示例。Gruendl 等[20]利用平行坐标对时间序列进行可视化，并提供了一种交互方法。Wang 等[21]使用可视化时间序列相似性的方法，将时间序列数据转换为短语序列并使用可视化技术来展示相似性。Dong 和 Wei[22]使用平行坐标可视化多变量时间序列。首先，对时间序列进行分段聚类。然后，在平行坐标上绘制多个时间序列，并使用不同的颜色来区分时间序列片段的顺序。螺旋图也是一种可以用于时间序列可视化的图形。螺旋图由 Carlis[23]提出，适用于分析时间序列数据的周期性特征。Cheng 等[24]提出了一种螺旋坐标与平行坐标相结合的时间序列可视化方法。Perin 等[25]研发了一个基于时间数据可视化的足球分析工具，名为 SoccerStories，

它展示了如何使用可视化技术来分析和呈现足球比赛中的数据。Ordez 等[26]使用检测时间序列数据模式(通常称为 motif)的高效无监督算法, 创建一个个性化的多变量时间序列表示。

研究步骤、方法及措施

时间数据可视化是数据科学中的一个重要分支, 涉及到诸多领域, 如金融、气候科学、社会经济、医疗健康等。我们的研究步骤将围绕常见领域的时间序列数据处理, 以及自动化技术的应用展开。以下是具体的步骤和方法:

1. 数据收集和预处理

选择金融、气候变化、医疗健康和社会经济等领域的时间序列数据。在每一个研究领域, 我们选取具有代表性和足够复杂度的时间序列数据。以金融领域为例, 我们可以选择大盘指数的历史数据, 这种数据不仅包含了股票的开盘价、收盘价、最高价和最低价, 还包含了成交量等其他信息, 再如其他领域的全球温度变化数据、患者的健康监测数据和城市的人口增长数据。在数据预处理的过程中, 我们还会关注数据的季节性和趋势性, 以及可能存在的异常值和断点。我们会使用 Python 的 Pandas 库或者 R 的 dplyr 库进行数据处理。

2. 时间数据可视化方法选择

我们将使用包括但不限于以下的可视化工具和方法: 线图 (Line Plot)、面积图 (Area Plot)、热图 (Heatmap)、树状图 (TreeMap)、烛台图 (Candlestick Plot, 常用于金融数据)、雷达图 (Radar Plot)、以及更高级的动态时间序列图 (如使用 D3.js 创建的可交互的时间序列图)。这些工具和方法将使用 Python 或 R 的数据可视化库, 如 Matplotlib, Seaborn, ggplot2, Plotly 等, 以实现数据可视化的自动化。

3. 时间数据可视化实验

实验阶段, 我们将以金融数据为例。我们将使用 Matplotlib 和 Plotly 创造不同类型的图形来呈现股票的开盘价、收盘价、最高价和最低价, 包括线图、面积图和烛台图。我们将探索如何通过各种图形最好地展示股票的历史走势和交易量。此外, 我们还将尝试使用自动化的特征选择和维度缩减技术, 如 PCA 或 t-SNE, 以提取和可

视化数据中的重要特征和模式。此外，我们会开发自定义的可视化工具，例如创建使用 D3.js 的动态时间序列图，以实现更丰富的数据交互和展示。

4. 结果评估和优化

在评估和优化阶段，我们会详细分析各种可视化方法的优缺点，并尝试进行优化。我们会关注图形的可读性、直观性、美观性和信息传达能力。例如，我们会通过调整图形的颜色、大小、形状和布局，以提高图形的视觉效果和信息传达效率。我们还会尝试使用人工智能和机器学习方法来自动优化图形，例如自动选择最佳的图形类型和参数。

总结

时间序列图是最常用的时间数据可视化工具,它以时间为横轴,变量值或统计量为纵轴,通过线条或面积图直观显示变量随时间变化的趋势。这种图适用于分析变量在一段时间内的总体变动情况,深入理解时间序列数据的演变规律。

时间分布图以时间为横轴或纵轴,标记事件发生的时间点,常通过散点图或热力图展示。它侧重于展示离散事件在时间上的分布状况,适用于离散事件型时间数据的分析。这可以发现事件的集中期或稀疏期,理解事件频发的时间模式。

时间轮播图按时间顺序排列图像或图表,通过动画切换展示时间的流逝和数据的变化。它直观生动,适合展示数据量巨大的时间序列,但交互性较差,不易进行精细分析。时间轮播图宏观地反映变量在整个时间范围内的变化过程,有利于理解长期趋势或周期性变化。

时间条形图在普通条形图的基础上,以时间为横轴,每个条形表示同一时刻下的多个观测值。它同时显示时间序列数据在多个维度的差异,便于进行细致的比较和分析。我们可以发现不同变量在某一时期内的异同,理解变量间的相关性。

时间面板图将时间轴映射为网格的行或列,在同一个平面上展示多个时间序列的数据。它允许我们在一个视角内观察多个时间序列的变化,便于发现不同序列之间的内在关联。通过时间面板图,我们可以比较不同变量在时间变化上的灵敏度和周期性,理解多个时间序列数据之间的互动关系。综上,时间数据可视化方法种类繁多,研究者可以根据研究问题和数据属性选择适宜的可视化工具,从宏观上理解时间序列的总体变化,也可以深入分析数据在某一时间段的微观变化,发掘时间数据背后隐藏的知识与见解。这在时间序列数据的分析与展示中具有重要作用

引用

- [1] 张鹏. 时间序列数据快速索引与可视化[D]. 华中科技大学,2021.DOI:10.27157/d.cnki.ghzku.2021.005620.
- [2]Brillinger D R.Time series:data analysis and theory[M].Society for Industrial andApplied Mathematics,2001.
- [3]彭燕妮,樊晓平,赵颖等.时间事件序列数据可视化综述[J].计算机辅助设计与图形学学报,2019,31(10):1698-1710.
- [4]Li D Q, Mei H H, Shen Y, et al. ECharts: a declarative framework for rapid construction of web-based visualization[J]. Visual Informatics, 2018, 2(2): 136-146
- [5] Fulda J, Brehmer M, Munzner T. TimeLineCurator: interactive authoring of visual timelines from unstructured text[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 300-309
- [6]Fan X P, Peng Y N, Zhao Y, et al. A personal visual analytics on smartphone usage data[J]. Journal of Visual Languages & Computing, 2017, 41: 111-120
- [7]Du F, Plaisant C, Spring N, et al. EventAction: visual analytics for temporal event sequence recommendation[C] //Proceedings of the IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2016: 61-70
- [8]Malik S, Shneiderman B, Du F, et al. High-volume hypothesis testing: systematic exploration of event sequence comparisons[J]. ACM Transactions on Interactive Intelligent Systems, 2016, 6(1): 1-23
- [9]Franklin L, Plaisant C, Minhazur Rahman K, et al. TreatmentExplorer: an interactive decision aid for medical risk communication and treatment exploration[J]. Interacting with Computers, 2016, 28(3): 238-252
- [10]陈馨葳,李天瑞,杨欢欢.基于时间序列数据的交互式主题河流可视化[J].山东大学学报(工学版),2019,49(04):29-35+43.
- [11]韩福波. 面向二维时序数据可视化方法的研究[D].山东大学,2017.
- [12]王子慧. 时变数据可视化系统的研究与实现[D].北京邮电大学,2019.
- [13]孙浩鑫. 基于时序的山东省环境数据可视分析研究[D].山东大学,2016.
- [14]Liu, B., & Zhang, L. (2018). A Review of Data Visualization Applications in China's Public Services: Progress, Problems and Prospects. International Journal of Information Management, 38(1), 70-78.
- [15] Havre S,Hetzler B,Nowell L.ThemeRiver: Visualizing theme changes over time[C]// Information Visualization, 2000. InfoVis 2000. IEEE Symposium on. IEEE, 2000.

- [16] Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). Visualization of time-oriented data. Springer.
- [17] Heer, J., & Agrawala, M. (2012). Design considerations for collaborative visual analytics. *Information Visualization*, 11(4), 275-294.
- [18] Alsallakh, B., & Hauser, H. (2014). Visualizing set intersections in up to seven dimensions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2201-2210.
- [19] Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2376-2385.
- [20] H.Gruendl, P.Riehmman, Y.Pausch, B.Froehlich. Time series plots integrated in parallel coordinates displays *Comput. Gr. Forum*, 35 (3) (2016), pp. 321-330
- [21] Wang, Y., Ding, X., Jia, H., Chen, L., Yang, Y., & Shang, J. (2016). Visualizing time-series similarity with phrasograms. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 560-569.
- [22] Z.Dong, Y.Wei. Visualization method for multi variable time series data using parallel coordinates. *J. Chin. Comput. Syst.*, 36 (10) (2015), pp. 2408-2411
- [23] J.V. Carlis. Interactive visualization of serial periodic data. *Proceedings of the Symposium on User Interface Software and Technology* (1998), pp. 29-38
- [24] S. Cheng, Z. Jiang, Q. Qi, S. Li, X. Meng. The polar parallel coordinates method for time-series data visualization. *Proceedings of the Fourth International Conference on Computational and Information Sciences* (2012), pp. 179-182
- [25] Perin, C., Vuillemot, R., Fekete, J. D., & Vernier, F. (2014). SoccerStories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2654-2663.
- [26] P. Ordez, M. Desjardins, C. Feltes, C.U. Lehmann, J. Fackler. Visualizing multivariate time series data to detect specific medical conditions. *Proceedings of the Annual Symposium*

/ AMIA Symposium. AMIA Symposium, American Medical Informatics Association (2008)