VARTS：代表性时间序列数据的可视化与分析工具

Duosi Jin计算机科学与技术学院 南京航空航天大学中国南京 jinduosi@nuaa.edu.cn

Jianqiu Xu*计算机科学与技术学院 南京航空航天大学中国南京 jianqiu@nuaa.edu.cn

Guidong Zhang自动化学院 广东工业大学中国广东 guidong.zhang@gdut.edu.cn

arXiv:2601.01361v1 [cs.GR] 4 Jan 2026

摘要—**大规模时间序列可视化常常面临视觉混乱过多和冗余模式等问题，使用户难以理解主要的时间趋势。为应对这一挑战，我们推出了 VARTS，这是一款用于代表性时间序列选择与可视化的交互式可视分析工具。基于我们此前的 M4-Greedy 工作，VARTS 将基于M4的采样、基于 DTW 的相似度计算和贪心选择整合到统一工作流中，用于识别和可视化代表性序列。该工具提供响应迅速的图形界面，允许用户导入时间序列数据集、执行代表性选择，并通过多个协调视图可视化原始数据和简化数据。通过减少冗余并保留关键数据模式，VARTS 有效提升了大规模时间序列分析的视觉清晰度和可解释性。演示视频见 https://youtu.be/mS9f12Rf0jo。**

索引词—**时间序列、可视化、代表性选择、采样**

I. 引言

传感、监测和数据采集技术的快速发展，使航空、制造业和物联网等领域积累了海量时间序列数据。可视分析在帮助用户探索时间模式、检测异常并从这类数据中获取洞见方面发挥着重要作用。然而，在可视化大规模时间序列数据集时，用户常常面临两大挑战：（i）每条序列中数量庞大的数据点会造成严重的视觉混乱并降低性能；（ii）同时显示大量相关序列，往往会遮蔽底层的时间趋势。

为应对这些挑战，我们此前的工作探索了可视化框架和数据约简技术。在我们早期的系统中，

FlyTsViz [1],我们开发了一个基于数据降维方法的飞行时间序列可视化平台，支持对大规模飞行数据集进行交互式探索。在此基础上，我们随后提出了*M4-Greedy* 算法 [2], ，这是一种面向可视化的代表性选择方法，可联合优化所选时间序列的多样性和覆

* 通讯作者。作者感谢美国国家科学基金会通过项目编号（U23A20296、62472217）提供的支持。

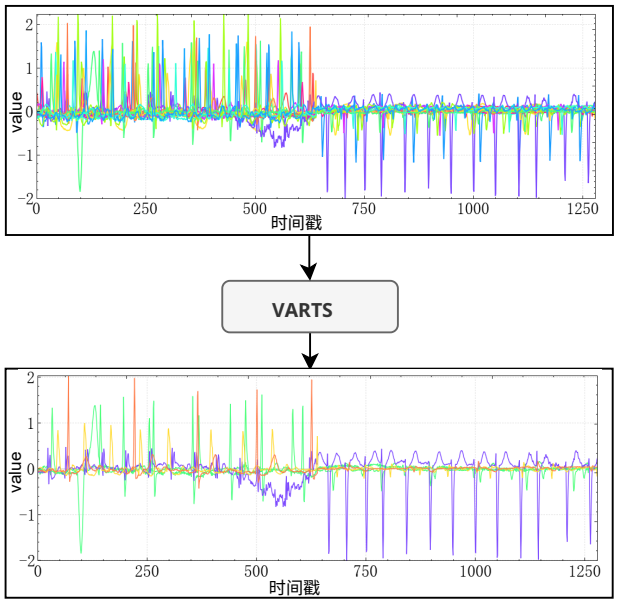


图1. 原始时间序列可视化与代表性时间序列可视化的比较。VARTS 自动选择并可视化一个简洁的子集，在减少视觉杂乱的同时保留主要模式。

盖度。该算法将基于 M4 的时间采样与基于 DTW 的贪心选择相结合，高效识别具有代表性的时间序列子集，在将数据量减少至原来的十分之一左右的同时，保留全局形状特征。尽管该算法在定量评估中表现出色，用户仍缺少一种直观的工具，来应用该算法、观察其选择过程，并在真实的分析场景中进行交互式探索。

本文介绍了**VARTS**（即代表性时间序列数据的可视化与分析），这是一款交互式可视分析工具，将 M4-Greedy 代表性选择算法集成到一个经过增强、面向大规模时间序列数据的可视化平台中。该工具使用 Qt 实现，提供响应灵敏的图形界面，

允许用户上传数据集、执行代表性选择，并通过多个协调视图可视化原始数据和简化数据。VARTS 将理论选择框架转化为交互式、可扩展且面向领域的可视分析工具，从而衔接了算法研究与实际应用。

该工具基于真实世界的飞行时间序列数据集进行演示，用户可以交互式地调整代表性序列数量（k）和多样性-覆盖率权衡（α）等参数，以探索不同抽象层次下的数据行为。VARTS 在保留关键时间模式的同时，有效减少了视觉混乱，从而促进对大规模时间序列数据更快速、更清晰的理解。如图1所示，VARTS 通过自动选择代表性序列，将密集、杂乱的原始时间序列数据转换为简洁且具有可解释性的可视化结果。图的上方面板展示了多条相互重叠的时间序列，由于视觉拥挤而难以解读。应用基于 M4-Greedy 的代表性选择后，下方面板仅展示信息量最大的序列，在保留基本时间模式的同时，显著提升了清晰度。这种精简不仅增强了可读性，还保持了数据集整体结构的保真度。

本演示的主要贡献总结如下：

1) 我们提出了 VARTS，这是一款用于代表性时间序列数据的可视化与分析的工具。2) 该工具实现了 M4-Greedy 算法，可自动识别并可视化代表性子集，从而减少视觉混乱并增强可解释性。3) 本演示展示了用户如何通过参数控制和协调式多视图可视化，交互式地探索飞行时间序列数据。

II. 相关工作

时间序列可视化。Tableau、Power BI 和 ECharts 等传统可视化工具提供交互式图表界面，但在处理大规模时间序列数据时能力有限。近期研究（例如 OM3 [3]，和基于 IoTDB 的可视化 [4]）致力于通过分层聚合或多尺度表示提升可扩展性和交互性。然而，这些方法主要被设计为算法框架，而非面向领域的可视分析平台。

代表性选择与采样。代表性选择已被广泛用于概括大型时间序列数据集。早期研究依赖基于聚类的方法 [5]，，而较新的研究则同时考虑多样性和覆盖度 [6]–[8]。这些方法启发了 M4-Greedy 算法 [2]，，该算法在 DTW（动态时间规整）度量下联合优化多样性和覆盖度。类似的贪心选择策略也已应用于空间众包等相关领域 [9]。

M4-Greedy 算法与工具集成。我们之前的工作 FlyTsViz [1], 提出了一种专注于数据缩减与可视分析的时间序列可视化工具，而 M4-Greedy [2]提出了一个用于代表性选择的算法框架。VARTS 将这两条工作路线整合起来，将 M4-Greedy 算法嵌入实用的可视化环境，在计算优化与交互式探索之间架起桥梁。

III. 框架

除了代表性选择模块外，VARTS 还支持通用可视化和异常可视化。不过，本文主要聚焦于代表性选择机制及其与可视化工作流的集成。

VARTS 的整体框架如图 2 所示。它在统一工作流中整合了数据预处理、代表性选择和交互式可视化。具体而言，该框架由三个主要组件构成：**（i）数据预处理**、**（ii）代表性选择**和**（iii）可视化与交互**。以下各小节将详细介绍每个组件。

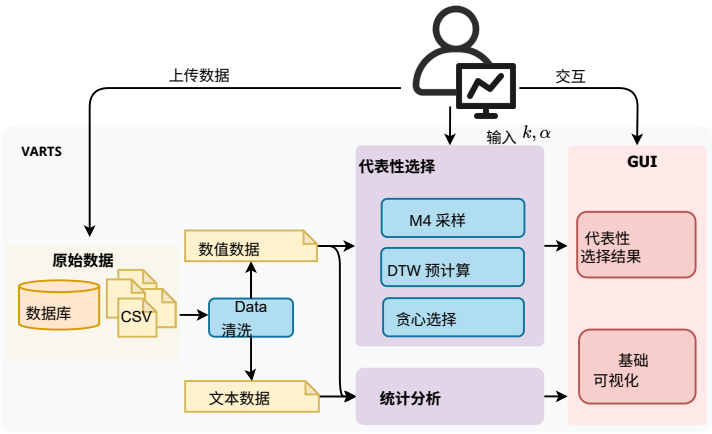


图 2. VARTS 的框架。

A. 数据预处理

VARTS 支持多种数据格式，包括数据库、CSV 文件和基于文本的记录，其中包含数值属性和文本属性。数据上传后，该工具会执行清洗、归一化和类型分类，以区分数值变量和分类变量。此外，系统会自动处理缺失值和不一致的时间戳，以确保数据完整性和时间一致性。

B. 代表性选择

代表性选择是 VARTS 的核心模块，负责减少数据冗余，同时保留大规模时间序列数据集的整体时间模式。该模块整合了三个算法组件，即 *M4* 采样、基于 *DTW* 的相似度计算和 贪心代表性选择，以自动识别能够最佳体现数据集全局特征的紧凑时间序列子集。

1）*M4* 采样：为高效描述每条时间序列的形状，VARTS 采用 M4采样方法 [10],，从每个片段中提取四个关键点：时间戳最小的点、数值最大的点、数值最小的点以及时间戳最大的点。这四个点（M4）能够有效保留局部子序列的振幅和时间形状，从而形成紧凑且信息丰富的表示。此步骤在保持数据整体结构的同时，大幅降低了后续相似度计算的计算负担。

2）基于 *DTW* 的相似度计算：给定采样后的表示，该工具使用 DTW距离计算时间序列之间的两两相似度。与欧氏距离不同，DTW 支持灵活的时间拉伸和压缩，因此对于因传感器采样差异而可能出现相位偏移或局部延迟的时间序列数据，它具有更强的鲁棒性。为提高效率，系统采用预计算和缓存机制来存储中间 DTW 结果，从而显著加快选择过程中的重复相似度查询。

3）贪心代表性选择：VARTS 基于 DTW距离矩阵，采用贪心策略迭代选择具有代表性的时间序列。令 $\mathcal{T} = \{T_1, T_2, \ldots, T_n\}$表示 n 时间序列的集合，令 $D(T_i, T_j)$表示两个序列 T_i 和 T_j之间的 DTW距离。目标是选择大小为 k 的子集 $\mathcal{R} \subset \mathcal{T}$，依据 M4-Greedy公式 [2]，联合优化多样性和覆盖度，以最大限度地提升该数据集的代表性。

目标函数。 优化目标定义如下：

$$\max_{\mathcal{R}\subset\mathcal{T},|\mathcal{R}|=k} \Big[\alpha \cdot \mathrm{Div}(\mathcal{R}) - (1-\alpha)\cdot \mathrm{Cov}(\mathcal{R},\mathcal{T})\Big], \quad (1)$$

其中 α 为非负系数，用于平衡两个目标。

其中多样性项鼓励所选代表彼此充分分离：

$$\mathrm{Div}(\mathcal{R}) = \min_{T_i,T_j\in\mathcal{R},T_i\neq T_j} D(T_i,T_j), \quad (2)$$

该项最大化代表之间最小的两两 DTW 距离。这种极大极小形式可避免选择冗余，并确保所有代表在时间模式上彼此不同。

其中覆盖度 项用于衡量代表对原始数据集的近似程度：

$$\mathrm{Cov}(\mathcal{R},\mathcal{T}) = \frac{1}{n}\sum_{T_i\in\mathcal{T}} \min_{T_j\in\mathcal{R}} D(T_i,T_j), \quad (3)$$

其中，值越小表示在 DTW 度量下，大多数序列都接近至少一个代表。

贪心优化。从空的代表集 $\mathcal{R} = \varnothing$ 开始，该算法会迭代选择能够使公式 1 中定义的综合目标函数获得最大增量改进的时间序列 T^*：

$$T^* = \arg\max_{T\in\mathcal{T}\setminus\mathcal{R}} \Big[\alpha \cdot \Delta\mathrm{Div}(\mathcal{R},T) - (1-\alpha)\cdot \Delta\mathrm{Cov}(\mathcal{R},T)\Big], \quad (4)$$

其中， $\Delta\mathrm{Div}(\mathcal{R},T)$ 和 $\Delta\mathrm{Cov}(\mathcal{R},T)$ 分别表示将候选项 T 加入当前代表集 $\mathcal{R}$ 时多样性和覆盖度的增量变化。该贪心过程会持续进行，直到 $|\mathcal{R}| = k$，从而生成一个代表性子集，在保持集合内部高多样性和全局低覆盖度的同时，保留数据集的基本时间结构，并最大限度减少可视化中的冗余。

C. 可视化与交互

可视化与交互组件使用户能够通过基于 Qt框架构建的图形界面，探索原始时间序列数据和代表性时间序列数据。用户可以直观地比较结果，并评估代表性选择过程的效果。

VARTS 支持多种飞行时间序列数据可视化模式，包括折线图、箱线图和轨迹图。这些协调视图使用户能够考察飞行参数的时间动态、统计分布和空间轨迹，从而从全局和局部视角了解飞行行为。

完成代表性选择后，该工具仅在可视化视图中显示选定的序列。用户可以调整代表项数量（k）和多样性-覆盖率权衡（α）等参数，可视化结果也会相应更新。前端界面与后端计算的集成支持对时间序列数据进行交互式探索。

总之，VARTS 框架将数据预处理、代表性选择和可视化交互无缝结合，构成了一个完整的工作流程，可用于智能且可扩展的时间序列视觉分析。

IV. 演示

基于上一节介绍的工具框架，本演示展示了 VARTS 的交互式工作流程，该流程将多视图可视化与代表性选择相结合，用于大规模时间序列分析。图 3 展示了 VARTS 的屏幕截图。上传数据集后，用户可以探索图 3(a) 中各个时间序列维度的趋势及其相互关系。图 3(b) 提供了整个数据集的概览，帮助用户识别哪些维度会显著影响整体数据分布。对于感兴趣的维度，用户可以返回图 3(a) 进行选择和深入查看。

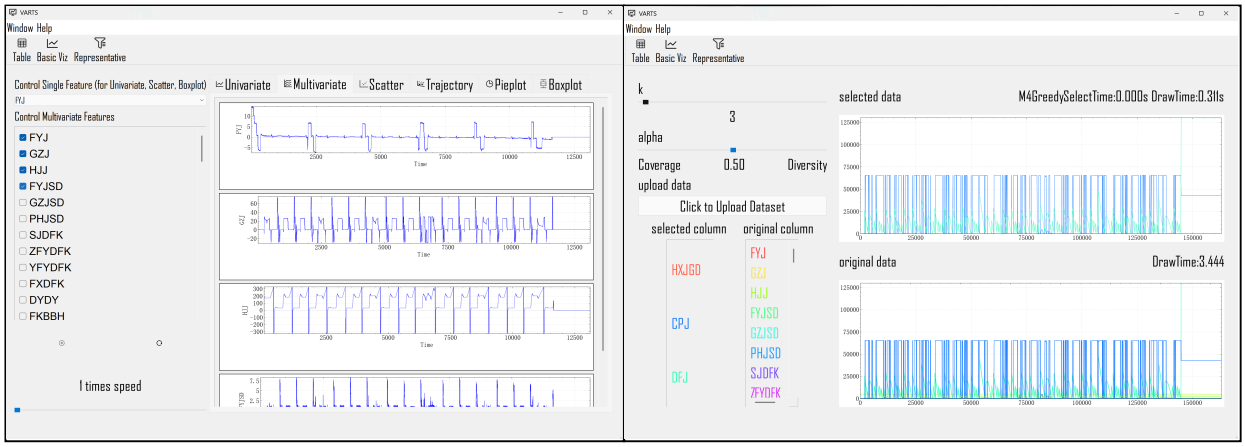
表 I VARTS 中使用的飞行时间序列数据集

数据集	#元组[a]	#变量	Size
FlightData-2021	78181	103	64 MB
FlightData-2023	110115	86	56 MB
FlightData-2024	545850	92	298 MB

[a]每个元组对应于飞行时间序列数据集中的一个时间步。

A. 数据导入与预处理

用户可以从数据库或本地 CSV 文件加载时间序列数据。VARTS 会自动解析、归一化，



(a) 基础可视化。 (b) 代表性选择结果。

图 3. VARTS 的屏幕截图。

并对齐数据，以确保时间一致性。表 I 总结了演示中使用的数据集，这些数据集在元组数量、变量数量和总体规模方面各不相同，展示了该工具处理各种飞行数据源时的可扩展性。此外，只要数据集包含多变量数值数据，就可以独立使用代表性选择功能。

B. 可视化与代表性选择

导入数据后，用户可以通过协调联动的可视化视图探索原始时间序列，包括折线图、箱线图和轨迹图，如图 3(a) 所示。这些视图从不同角度呈现飞行参数的时间变化、统计分布和空间行为。代表性选择模块应用 M4-Greedy 算法，识别出一个精简的代表性序列子集。数据加载后会自动生成初始可视化，如图 3(b) 所示，其中原始时间序列和代表性时间序列并排显示。此对比视图凸显了所选子集如何在大幅减少视觉混乱和冗余的同时，保留数据集的主要时间模式。

C. 交互式探索

VARTS 支持对参数调整和结果比较进行实时交互。用户可以通过图 3(b) 中的滑块调整代表性序列的数量（k）和多样性-覆盖率权衡（α）：当 $\alpha \to 0$ 时，选择会侧重于覆盖度；而 $\alpha \to 1$ 则会强调多样性。初次选择时计算一次 DTW 相似度矩阵后，后续调整 k 或 α 只需根据预计算结果更新可视化。显示代表性序列所需的时间仅约为渲染所有原始时间序列所需时间的 10%；这是因为代表性序列的维度通常少于 10，而原始数据集包含约 100 个变量，从而确保交互流畅、响应迅速。

本演示展示了 VARTS 如何连接算法分析与人类洞察，提供一个有效且可扩展的平台，用于航空分析中的代表性时间序列可视化。

V. 结论

本演示介绍了 VARTS，这是一款面向大规模时间序列数据的交互式可视分析工具。该工具实现了 M4-Greedy 代表性选择算法，将 M4 采样、基于 DTW 的相似度计算和贪心选择整合为实用的可视化工作流。通过在真实航班时间序列数据集上的演示，VARTS 使用户能够探索并比较原始序列与代表性序列、调整选择参数，并在减少视觉混乱的同时分析模式。

参考文献

[1] D. Jin、J. Xu 和 X. Zhao，“FlyTsViz：一种基于降维方法的飞行时间序列数据可视分析系统”，载于第八届国际数据存储与数据工程会议（*DSDE*）论文集，2025年，第 30–36 页。[2] D. Jin、S. Xu、M. Ju、Z. Zhang 和 F. Lin，“M4-Greedy：面向可视化的时间序列数据代表性选择”，载于第一届国际 *APWeb-WAIM* 空间富集大数据管理与应用研讨会（*SRBDMA2025*），待出版。[3] Y. Wang等，“OM3：一种用于时间序列交互式渐进可视化的有序多级最小最大表示”，*ACM* 数据管理会议论文集，第 1 卷，第 2 期，第 145:1–145:24 页，2023年。[4] L. Rui、X. Huang、S. Song、Y. Kang、C. Wang 和 J. Wang，“Apache IoTDB 中的时间序列表示”，*ACM* 数据管理会议论文集，第 2 卷，第 1 期，第 35:1–35:26 页，2024年。[5] J. Paparrizos 和 L. Gravano，“k-Shape：高效且准确的时间序列聚类”，载于*ACM SIGMOD* 国际数据管理会议论文集，2015年，第 1855–1870 页。[6] F. Pan、W. Wang、A. K. H. Tung 和 J. Yang，“从海量数据中寻找代表集”，载于第五届 *IEEE* 国际数据挖掘会议（*ICDM*）论文集，2005年 9 月，第 8–15 页。[7] C. Yang、L. Chen、H. Wang 和 S. Shang，“面向多样性和覆盖度的高效活动轨迹选择”，载于*AAAI* 人工智能会议论文集（*AAAI*），2021年，第 35 卷，第 1 期，第 689–696 页。[8] G. Lee、S. Huang、Z. Bao 和 Y. Zhao，“用于数据探索的代表性时间序列发现”，*VLDB*基金会论文集，第 18 卷，第 3 期，第 915–928 页，2024年。[9] Y. Tong、J. She、B. Ding、L. Wang 和 L. Chen，“空间众包中的在线移动微任务分配”，载于第三十二届 *IEEE* 国际数据工程会议（*ICDE*）论文集，2016年，第 49–60 页。[10] U. Jugel、Z. Jerzak、G. Hackenbroich 和 V. Markl，“M4：面向可视化的时间序列数据聚合”，*VLDB*基金会论文集，第 7 卷，第 10 期，第 797–808 页，2014年。