

# 面向自动可视化的可解释数据变换推荐

吴子梁<sup>1</sup>, 陈为<sup>1</sup>, 马昱欣<sup>2</sup>, 徐彤<sup>1</sup>, 严凡<sup>1</sup>, 吕樛<sup>1</sup>, 钱中昊<sup>1</sup>, 夏佳志<sup>3</sup>

<sup>1</sup>浙江大学计算机辅助设计与图形学国家重点实验室, 中国杭州市, 310058

<sup>2</sup>南方科技大学计算机科学与工程系, 中国深圳市, 518055

<sup>3</sup>中南大学计算机学院, 中国长沙市, 410083

**摘要:** 自动可视化技术能够为不熟悉可视化设计的用户生成有意义的可视化, 以支持他们的数据分析和模式发现需求。当前, 主流的自动可视化方法采用聚合与过滤从原始数据抽取模式信息。然而, 这些有限的的数据变换并不能捕获聚类、关联等复杂的模式。尽管特征工程领域的最新进展为更加广泛的自动数据变换提供了可能, 其结果却缺少可解释性, 导致变换后的模式无法与原始数据特征建立联系。为应对上述挑战, 我们面向自动可视化中广泛的数据变换类型, 提出一种创新的可解释推荐方法。我们通过回顾既往文献总结可行的数据变换空间, 通过开展预实验总结变换可解释性的度量。我们的推荐算法能够计算最优的数据变换, 这种变换能够在维持可解释性的同时揭示数据的模式信息。真实场景下的使用案例与用户实验验证了我们方法的有效性。

**关键词:** 数据变换; 数据变换推荐; 自动可视化; 可解释性

<https://doi.org/10.1631/FITEE.2200409>