# 基于多方表格数据关联策略的数据补全可视分析方法

朱海洋[1,2]，韩东明[1]，潘嘉铖[1]，魏雅婷[3]，封颖超杰[1]，翁罗轩[1]，
毛科添[1]，邢远凯[2]，闾建树[2]，万邱成[2]，陈为[1]
[1]浙江大学计算机辅助设计与图形系统全国重点实验室，中国杭州市，310058
[2]物产中大数字科技有限公司，中国杭州市，310020
[3]物产中大金属集团有限公司，中国杭州市，310005

**摘要：**数据补全是数据治理的一项重要预处理任务，目的是填补不完整的数据。然而，传统的数据补全方法只能通过单张数据表格在一定程度上缓解数据的不完整问题，并未能在补全值的准确性和效率之间达到最佳平衡。本文提出了一种新颖的数据补全可视化分析方法；设计了一套多方表格数据关联策略，采用智能算法识别相似列并在多个表格之间建立列之间的关联关系，然后利用其它表格中的相似数据条目对缺失数据进行初始补全；开发了一个可视分析系统来优化数据补全的候选值。本文中的交互式系统将多方数据补全方法与专家知识相结合，有助于更好地理解数据的关系结构，显著提高了数据补全的准确性和效率，提升了数据治理质量和数据资产内在价值。实验验证和用户调查表明，本文方法支持用户使用领域知识验证判断相关列及相似行。

**关键词：**数据治理；数据不完整；数据补全；数据可视化；交互式可视分析
https://doi.org/10.1631/FITEE.2300480