

数据库与信息处理

多数据流上的联机方差分析研究

王小龙, 马瑞民

大庆石油学院 计算机与信息技术学院, 黑龙江 大庆 163318

收稿日期 修回日期 网络版发布日期 2007-5-9 接受日期

摘要 多数据流上的联机方差分析是一个有意义的研究问题。针对以元组为单位流入的具有相同属性集的多支单数据流组成的多数据流, 提出了分别对每支单数据流进行蓄水池抽样, 构造一一对应于各单数据流的若干个多快照窗口, 即两者之间是双射关系, 可以将多快照窗口串行置于主存中, 将元组包含的属性与多快照窗口中的各个快照窗口一一对应, 且使得同一快照窗口中的各基本窗口与取自其对应的单数据流的属性值样本一一对应, 然后对这些相互独立的样本进行方差分析。按顺序串行处理各个多快照窗口中的数据, 就可以用串行化的方法来实现并行的多数据流上的联机方差分析。理论分析与实验表明, 该方法是合理的和有效的。

关键词 [多数据流](#) [快照窗口](#) [双射](#) [模型](#) [联机方差分析](#)

分类号

Research on on-line analysis of variances over data streams

WANG Xiao-long, MA Rui-min

School of Computer and Information Technology, Daqing Petroleum Institute, Daqing, Heilongjiang 163318, China

Abstract

It is important to analyze variances over data streams. In this paper, each of single data stream that is included in data streams has same attributes set as other, the unit of single data stream is tuple, and same attributes set is included in each of these tuples. Reservoir algorithm is used to sample from these single data stream respectively, then some multiple snapshot windows are constructed, the relationship between these single data stream and multiple snapshot windows is bijective mapping, the relationship between attributes that are included in tuple and snapshot windows that are included in relative multiple snapshot windows is bijective mapping, and the relationship between basic windows that belong to same snapshot window and the attribute values that come from different single data stream is bijective mapping as well, that is, these attribute values come from same attribute that is comprised by different single data stream. Variances are analyzed based on these independent sample values. Data of these multiple snapshot windows can be processed orderly. In a word, a serialization method is used to analyze parallel data streams. The analytical and experimental results show that this analytic method is logical and effective.

Key words [data streams](#) [snapshot window](#) [bijective mapping](#) [model](#) [on-line analysis of variance](#)

DOI:

通讯作者 王小龙 [E-mail: wangxiaolongdqi@yahoo.com.cn](mailto:wangxiaolongdqi@yahoo.com.cn)

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(893KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“多数据流”的相关文章](#)

▶ [本文作者相关文章](#)

· [王小龙](#)

· [马瑞民](#)