

doi:10.1631/FITEE.1400376

题目：基于趋势特征的时间序列符号化方法

目的：提出一种通用方法用于长时间序列的知识发现过程。

创新点：提出一种基于并行分割的时间序列符号化方法----趋势特征符号化近似法 (trend feature symbolic approximation, TFSA)，对长时间序列进行快速分割，并且保留原始序列大多数趋势特征，将分割后的子序列用特征符号表示。本文的贡献在于改进了长时间序列的分割效率，而且 TFSA 专注于保留原始时间序列的大多数趋势特征，使得挖掘后的规则更加容易理解和解释。

方法：首先，通过一个两步 (two-step) 分割机制将时间序列分割成一系列不等长的子序列。然后，采用趋势特征符号化近似 (TFSA) 将子序列符号化并获得符号项集。最后通过一个基于 apriori 的关联规则算法来实现时序数据的知识发现。

结论：针对长时间序列，基于累积和控制图方法研究一种海量数据环境下序列的并行分割机制。可以通过分布式结点来实现，随结点数增加，其效率将进一步提高。TFSA 符号化方法不同于传统的方法，它致力于保留原始时间序列的大部分趋势特征及模式，通过规定的趋势符号来表示时间序列，并且其表达方式也考虑后续的时间序列挖掘研究。实验证明，本文方法在时间序列的分割效率以及分类准确性上相比于已有的方法均有所提高。

关键词：长时间序列；分割；趋势特征；符号化；知识发现