

## 多元时间序列相似性搜索研究综述

李正欣<sup>†</sup>, 张凤鸣, 张晓丰, 陈继成, 李超

(空军工程大学 装备管理与安全工程学院, 西安 710051)

**摘要:** 多元时间序列相似模式挖掘是数据挖掘领域的研究热点,它主要包括特征表示、相似模式度量和相似性搜索3个方面.目前,大部分研究成果主要集中在特征表示和相似模式度量,相似性搜索则成为制约问题突破的关键环节.为此,主要针对多元时间序列的相似性搜索进行综述,归纳了主要的相似模式度量方法,对比了不同相似模式度量下的序列搜索方法,并分析了不同方法的优缺点,以期为进一步研究多元时间序列相似性搜索提供帮助.

**关键词:** 多元时间序列; 数据挖掘; 相似模式度量; 相似性搜索

中图分类号: TP311

文献标志码: A

### Survey of similarity search for multivariate time series

LI Zheng-xin<sup>†</sup>, ZHANG Feng-ming, ZHANG Xiao-feng, CHEN Ji-cheng, LI Chao

(Equipment Management and Safety Engineering College, Air Force Engineering University, Xi'an 710051, China)

**Abstract:** Similar patterns mining for multivariate time series is becoming a hot topic in the area of data mining, which consists of three consecutive parts: Feature representation, similarity measure and similarity search. Most researches mainly focus on feature representation and similarity measure, which make similarity search still a tough problem in similar patterns mining. Therefore, the existing similarity measures are summarized, different methods of similarity search for multivariate time series are compared, their merits and demerits are analyzed, and the further research direction is provided.

**Keywords:** multivariate time series; data mining; similarity measure; similarity search

## 0 引言

时间序列是一种普遍存在的数据类型,广泛存在于金融、医疗、电子信息和气象等工程领域<sup>[1]</sup>.例如:股市逐日的交易情况、商品销售数据、气象数据、“黑匣子”记录的飞行数据等.由于数据存储和处理能力的不断提高,随着时间的推移,现实应用中时间序列的存储规模呈现爆炸式增长<sup>[2]</sup>.

按时间顺序获取的一系列观测值  $x_t(j)$  称为时间序列,其中  $t(t = 1, 2, \dots, n)$  表示第  $t$  个时刻,  $j(j = 1, 2, \dots, m)$  表示第  $j$  个变量,  $x_t(j)$  表示第  $j$  个变量在第  $t$  个时刻上的记录值<sup>[3]</sup>.当  $m = 1$  时,  $x_t(j)$  为一元时间序列(UTS);当  $m > 1$  时,  $x_t(j)$  为多元时间序列(MTS).

时间序列数据挖掘旨在探索隐含在海量时间序列之中的深层知识,它已成为目前数据挖掘领域中10个最具挑战性的问题之一<sup>[4]</sup>,在图像识别、语音处

理、声纳技术、遥感技术、机械工程等工程技术领域以及金融分析、人口统计、地震检测等社会经济领域都具有广阔的应用前景<sup>[5]</sup>.

时间序列模式是指一段时间内,序列形态的变化趋势,它体现系统在一个阶段中的演化规律.时间序列中的一些模式具有明确含义,例如,某个模式的飞行数据对应着特定的一个飞行动作,或复杂状态下的潜在故障.

相似模式挖掘以时间序列的典型模式为基础,从海量时间序列中寻找与其形态相似的时间序列.其过程可以描述为:给定时间序列数据集TB、时间序列  $x$ 、具有典型模式的查询序列  $Q$  和相似模式度量模型  $\text{Sim}(\cdot, \cdot)$ ,从TB中找出与  $Q$  相似的所有时间序列集合  $R$ ,即

$$R = \{x | x \in \text{TB} \wedge \text{Sim}(Q, x) = \text{True}\}. \quad (1)$$

多元时间序列相似模式挖掘是众多应用领域

收稿日期: 2016-09-14; 修回日期: 2016-12-09.

基金项目: 国家自然科学基金项目(61502521, 71601183).

作者简介: 李正欣(1982—),男,博士,从事信息系统工程与智能决策、数据挖掘、机器学习等研究;张凤鸣(1963—),男,教授,从事信息系统工程与智能决策等研究.

<sup>†</sup>通讯作者. E-mail: lizhengxin\_2005@163.com

共同面对的一个关键问题,例如:寻找相似走势的股票<sup>[6]</sup>、查找具有相似病情的患者脑电图<sup>[7]</sup>、公共健康监测<sup>[8]</sup>、图像检索<sup>[9]</sup>、视频检索<sup>[10]</sup>、气象监测<sup>[11]</sup>、互联网事态监控<sup>[12]</sup>等,因此,针对该问题的研究具有广阔的应用前景。

## 1 相关研究

时间序列相似模式挖掘主要涉及3个问题:

1)特征表示.用简洁的形式表达与问题相关的时间序列特征。

2)相似模式度量.以特征表示的结果作为输入,定量刻画两条时间序列的相似程度,且能够较好地支持时间序列的多种形变<sup>[13]</sup>,如图1所示。

3)相似性搜索.以某种相似模式度量方法为基础,在时间序列数据集中,寻找与给定模式相似的时间序列。

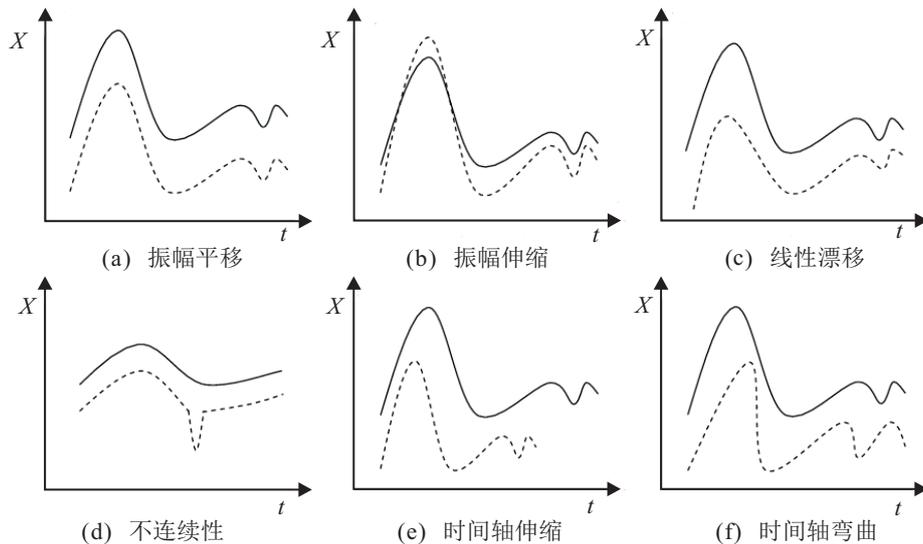


图1 时间序列的多种形变

根据搜索对象不同,可分为全序列搜索和子序列搜索:已知时间序列数据集 $TB = \{S_1, S_2, \dots, S_t\}$ 、典型模式查询序列 $Q$ ,寻找与 $Q$ 相似的所有序列 $S_i (1 \leq i \leq t)$ 称为全序列搜索,寻找与 $Q$ 相似的所有序列 $S_i[j:k] (1 \leq i \leq t, 1 \leq j \leq k \leq |S_i|)$ 称为子序列搜索.与全序列搜索相比,子序列搜索的复杂性主要体现在搜索对象的急剧增长。

国内外研究人员对上述问题做了相关的研究工作,但绝大部分研究都是针对一元时间序列,针对多元时间序列的研究还不够成熟,还有较多尚未解决的问题<sup>[14]</sup>。

现实世界中的状态刻画通常涉及多个变量,与一元时间序列相比,多元时间序列更具普遍性.例如,股票交易可以用开盘价、收盘价、最高价、最低价、交易量等变量描述.广义上讲,任何包含多变量的数据存储都可以被视为多元时间序列<sup>[15]</sup>。

与一元时间序列相比,多元时间序列的维度更高,且变量间存在关联关系,如图2所示.因此,不能将多元时间序列看成多个一元时间序列的简单叠加,一元时间序列相似模式挖掘方法很难直接用于多元时间序列<sup>[16]</sup>。

3个问题是承接递进的一个整体:特征表示和相

似模式度量是基础,决定着相似性匹配的准确性;相似性搜索是前两个环节的延伸和扩展,是相似模式挖掘的最终目标,它使得在海量数据中寻找给定序列模式成为可能。

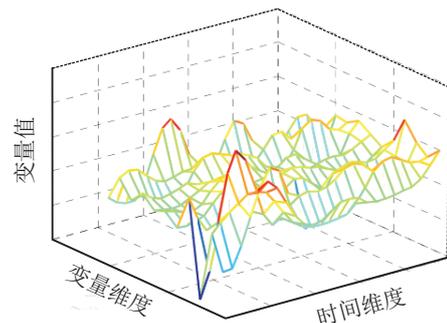


图2 多元时间序列的三维图形

目前,大部分研究成果主要集中在多元时间序列的特征表示和相似模式度量,相似性搜索成为制约问题突破的关键环节。

特征表示决定着相似模式度量方法的选取,相似模式度量又直接影响着序列搜索方式.如果只针对其中的某一环节进行探讨,不将3个环节作为一个整体,不考虑与其他环节的承接关系,则问题很难取得实质性进展.因此,在探讨相似性搜索之前,必须先分析其前一个环节——相似模式度量。

## 2 多元时间序列相似模式度量

目前,多元时间序列的相似模式度量方法主要包括:Euclidean距离<sup>[17]</sup>、动态时间弯曲距离(DTW)<sup>[18]</sup>、奇异值分解方法(SVD)<sup>[19]</sup>、基于点分布特征(PD)<sup>[20]</sup>的方法等。

1) Euclidean距离. Euclidean距离简单直观,是最常用的一种距离度量,其计算复杂度不高,与序列长度成线性关系.然而,它对时间序列的各种形变比较敏感,且要求时间序列的长度必须相等<sup>[21]</sup>.

2) DTW距离. DTW距离允许序列点自我复制后再进行对齐匹配,它能够度量不等长时间序列的相似性,并且支持时间轴的弯曲和伸缩等形变,因此在时间序列相似模式度量中被广泛采用<sup>[22]</sup>.

3) SVD方法. SVD方法在部分文献中被称为扩展Frobenius范数,它将时间序列中的变量视为随机变量,将各个时刻的记录值视为随机变量的样本点,以相关系数矩阵作为特征表示的基础,用扩展Frobenius范数来度量两个多元时间序列的相似性. SVD方法能够体现变量间的相互关系,支持不同长度多元时间序列的相似模式度量.但它是一种统计意义上的度量方法,不能体现记录值的时序关系,因此存在误判风险,且计算复杂度较高<sup>[16]</sup>.如果任意交换一个多元时间序列中各时刻记录值的次序,使用SVD方法时,新序列与原序列一定是相似的,这显然不符合人们对相似性的直观判断.

4) PD方法. PD方法在三维空间中描述多元时间序列,并抽取局部重要点作为特征;然后根据局部重要点的统计分布,构建一个特征模式向量,并采用Euclidean距离度量特征模式向量之间的差异. PD方法能够有效地刻画多元时间序列的形状特征,且支持不等长序列的相似模式度量,对小规模的多元时间序列能够得到相对较好的匹配效果.然而,该方法提取局部重要点时,将不同变量的记录值放在一起比较,没有考虑不同变量间的量纲差异,具有一定的局限性.此外,它也是一种统计意义上的度量方法,不能体现记录值的时序关系,也存在误判风险<sup>[16]</sup>.

5) 其他相似模式度量方法. 此外,还有最长公共子串距离(LCS)<sup>[23]</sup>和编辑距离(ED)<sup>[24]</sup>,它们将时间序列的记录值离散化为单个字符,将时间序列转换为字符串. LCS方法用最长公共子串的长度来度量两个时间序列的相似性. ED方法通过替换、删除和插入等编辑操作,将两个字符串变成完全一致,以最小操作代价度量两个时间序列的相似性. LCS方法和ED方法都以时间序列符号化为前提,如何选择合适

的离散算法,如何保持变量之间的相互关系,如何合理确定字符表的大小都是符号化方法面临的难题.

比较现有的多元时间序列相似模式度量方法,DTW距离能够度量长度不同的多元时间序列,支持时间轴的弯曲和伸缩等形变,且不存在误判风险,不需要进行符号化处理.但它存在两个缺陷,给后续的相似性搜索带来了困难:计算复杂度较高,当数据规模较大时,采用顺序扫描的方法进行逐一匹配不切实际;不满足距离三角不等式,直接进行索引查询会遗漏正确结果.

## 3 搜索策略与完备性

顺序扫描是一种最直观的搜索策略,即用查询序列逐一与数据集中的所有序列进行模式度量,找出相似序列.然而,通常情况下,数据集中的序列数量众多,子序列数量更为庞大,且相似模式度量的计算复杂度较高,因此顺序扫描往往是不可行的.为了提高搜索效率,需要一个更好的搜索策略.

常用方法遵循两步搜索的基本流程,如图3所示,执行顺序用编号明确.

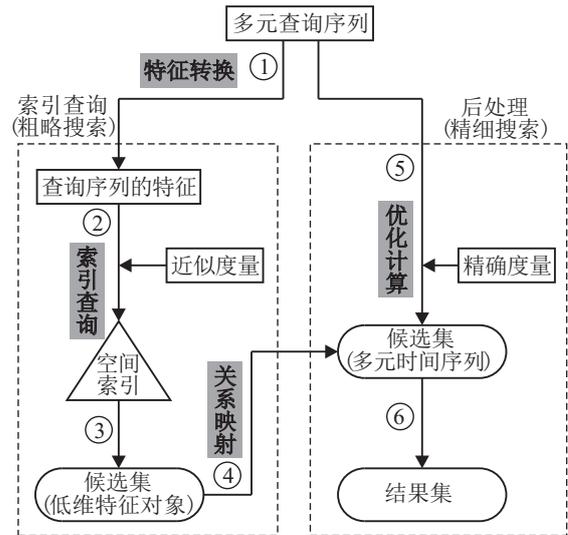


图3 两步搜索的基本流程

1) 索引查询. 首先,将时间序列映射到低维特征空间,将其转换为低维特征空间中的对象;然后,利用空间索引结构对低维空间对象进行组织;最后,使用查询序列的低维特征在索引结构上执行查询,利用索引的剪枝功能提高搜索效率.粗略搜索去除大部分不相似序列,得到候选集.与原始数据集相比,候选集中的序列数量大为减少.

2) 后处理. 计算多元查询序列与候选集中每个多元时间序列的相似性,去除不相似序列,得到结果集.

其中所涉及的研究内容主要包括:如何将多元

时间序列映射到低维特征空间,用索引进行组织;如何构造近似度量方法,利用索引进行粗略搜索;如何降低后处理阶段的计算代价;如何将全序列搜索拓展至子序列搜索.

完备性是衡量搜索策略的一个重要标准,它包括完全性和准确性. 设  $S$  是数据集  $TB$  中满足相似性匹配要求的序列集合,  $R$  是实际搜索到的序列集合. 若  $S - R \neq \emptyset$ , 则搜索不是完全的,  $S - R$  表示遗漏的正确结果, 称发生漏报; 反之, 若  $R - S \neq \emptyset$ , 则搜索是不准确的,  $R - S$  表示引入的错误结果, 称发生误报.

通常准确性较容易满足, 只要结果集中的序列都满足相似性要求即可, 而完全性却不容易保证. 文献 [25] 证明了下界距离引理, 它能够保证时间序列在变换到特征空间后的相似性搜索不发生漏报.

**下界距离引理**<sup>[25]</sup> 设时间序列  $Q$  和  $C$ , 通过特征提取函数  $F$  映射到特征空间, 为了保证特征空间的搜索不产生漏报, 必须满足

$$D_{\text{feature}}(F(Q), F(C)) \leq D_{\text{true}}(Q, C), \quad (2)$$

其中  $D_{\text{feature}}$  和  $D_{\text{true}}$  分别表示特征空间和原始空间的相似模式度量方法.

## 4 多元时间序列相似性搜索

现有的相似性搜索方法主要针对 Euclidean 距离或 DTW 距离, 采用两步搜索策略, 满足下界距离定理, 实现搜索的完备性. 其中, 针对一元时间序列的搜索方法, 虽然不能直接应用于多元时间序列, 但对研究多元时间序列的相似性搜索具有一定借鉴意义.

### 4.1 支持 Euclidean 距离的相似性搜索

索引结构很大程度上依赖于相似模式度量, 目前大部分相似模式度量都基于 Euclidean 距离, 因此索引方式大多采用空间索引结构.

Agrawal 等<sup>[26]</sup> 提出了处理等长度全序列搜索问题的 F-Index 索引方法, 它采用离散傅立叶变换 (DFT) 将一元时间序列映射到频域空间, 提取主要系数作为特征, 从而将一元时间序列映射为特征空间中的向量点, 再使用 R\*-Tree 对特征向量点进行组织. 相似性搜索时, 先在 R\*-Tree 上进行索引查询, 然后进行后处理, 去除不满足相似性要求的序列, 得到结果集. 在 F-Index 索引方法的基础上, Chan 等<sup>[27]</sup> 用离散小波变换 (DWT) 将一元时间序列映射到频域空间, 其余步骤均与 F-Index 方法类似.

Faloutsos 等<sup>[25]</sup> 将 F-Index 方法扩展到子序列搜索, 提出了处理等长子序列搜索问题的 ST-Index 方法. 首先采用滑动窗口 (Sliding Window) 技术将一元时间序列划分为等长子序列; 然后使用文献 [26] 中的方法将等长子序列转换为特征空间中的向量点, 在特

征空间中, 这些向量点形成一条轨迹; 最后对轨迹进行分割, 每个分割部分都用最小边界矩形 (MBR) 表示, 采用 R\*-Tree 对所有 MBR 进行组织. 查询序列经过邻接窗口 (Disjoint Window) 分割、特征映射与提取后, 转换为特征空间上的点集. 执行查询时, 先在 R\*-Tree 上搜索包含查询序列特征点的 MBR, 这些 MBR 内的向量点对应的子序列经过扩展后, 转换为候选序列; 然后在原始数据空间上使用 Euclidean 距离对候选序列进行顺序扫描, 去除不满足相似性要求的子序列.

Moon 等<sup>[28]</sup> 进一步提出了处理等长子序列搜索的 DualMatch 方法, 可视为对 ST-Index 方法的改进. 它的基本思路与 ST-Index 方法类似, 但采用邻接窗口分割一元时间序列, 利用滑动窗口分割查询序列 (与 ST-Index 方法相反), 使得一元时间序列被分割的子序列数量大大减少, 可以直接在 R\*-Tree 上存储特征向量点. 执行查询时, 能够显著地降低候选集大小, 从而提高搜索效率.

离散傅立叶变换和离散小波变换都具有保持 Euclidean 距离不变的特性, 即满足下界距离引理, 因此上述方法都不会产生漏报. 但由于采用的相似模式度量均为 Euclidean 距离, 且都使用 DFT 或 DWT 将时间序列映射到频域空间, 上述方法只支持 Euclidean 距离下的等长一元时间序列的全序列或子序列搜索.

### 4.2 支持 DTW 距离的相似性搜索

DTW 距离支持不等长时间序列的相似模式度量, 其计算复杂度较高, 直接针对 DTW 距离构造索引结构比较困难. 为此, 需要寻找一种计算更简单且易于进行索引查询的距离度量来粗略地估计 DTW 距离, 称为 DTW 下界距离. 利用它在索引结构上执行查询, 过滤掉大部分不满足相似性要求的序列, 从而提高搜索效率. DTW 下界距离要满足 3 个条件<sup>[29]</sup>:

- 1) 正确性. 所有满足条件的序列应该在经下界距离过滤得到的候选集中, 即不允许出现漏报.
- 2) 有效性. 下界距离的计算复杂度尽量低.
- 3) 紧致性. 下界距离的度量结果尽量逼近 DTW 距离, 从而减少后处理的计算量.

#### 4.2.1 支持 DTW 距离度量的全序列搜索方法

现有的支持 DTW 距离度量的相似性搜索方法, 基本都遵循如下思路: 首先, 将时间序列映射为低维特征对象, 利用空间索引技术对低维特征对象进行组织; 然后, 寻找一种计算代价较低, 支持距离三角不等式的 DTW 下界距离, 用其在索引上执行搜索, 过滤掉大部分不满足相似性要求的序列; 最后, 对剩下的候选集中的序列采用 DTW 距离进行筛选, 获得搜索结

果.这类方法对DTW下界距离有3个关键要求:应具有较低的计算复杂度,以提高搜索效率;应具有较高的紧致性,以提高过滤性能;应满足下界距离引理,以保证非漏报性.

Yi等<sup>[30]</sup>、Kim等<sup>[31]</sup>、Keogh等<sup>[32]</sup>分别提出了支持DTW距离的一元相似序列搜索方法,他们分别给出了各自的DTW下界距离,提出了支持相应下界距离的索引构建方法,并且证明了搜索方法的非漏报性.

Yi等在构造一元时间序列的DTW下界距离时,以一条序列作为基准序列,以另一条序列中大于基准序列最大值的点集以及小于基准序列最小值的点集作为特征,以此为基础,构造下界距离LB\_Yi,如图4所示.

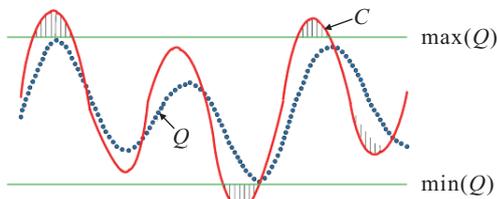


图4 LB\_Yi计算示意图

Kim等将一元时间序列的起始点、结束点、最大值点和最小值点作为特征,构造DTW下界距离LB\_Kim,如图5所示.

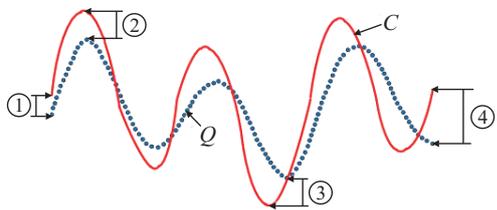


图5 LB\_Kim计算示意图

Keogh等提出了一种DTW下界距离LB\_Keogh:

$$LB\_Keogh(Q, C) = \sqrt{\sum_{i=1}^n \begin{cases} (c_i - u_i)^2, & c_i > u_i; \\ (c_i - l_i)^2, & c_i < l_i; \\ 0, & \text{otherwise.} \end{cases}} \quad (3)$$

其中:  $Q = \langle q_1, q_2, \dots, q_n \rangle$  和  $C = \langle c_1, c_2, \dots, c_n \rangle$  的长度均为  $n$ ;  $u_i$ 、 $l_i$  分别定义为  $u_i = \max\{q_{i-r}, \dots, q_{i+r}\}$ 、 $l_i = \min\{q_{i-r}, \dots, q_{i+r}\}$ ;  $r$  是弯曲路径的全局约束条件,如图6所示.

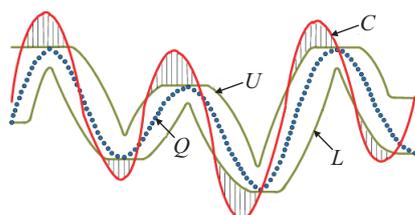


图6 LB-Keogh计算示意图

Keogh等<sup>[17]</sup>首先使用分段累积近似方法(PAA),将数据集中的一元时间序列转换为空间向量点,

用R-Tree对向量点进行组织;然后提取查询序列的上、下边界序列作为查询特征,利用DTW下界距离LB\_Keogh在空间索引上执行查询,索引查询的结果构成候选集;最后使用DTW距离计算查询序列与候选集中每个一元时间序列的DTW距离,去除不符合相似性条件的序列,得到结果集. Keogh等通过大量实验验证了下界距离LB\_Keogh的紧致性优于LB\_Yi和LB\_Kim.

Zhu等<sup>[33]</sup>、Zhou等<sup>[34]</sup>分别对LB\_Keogh进行改进,进一步提高了下界距离在索引查询中的紧致性.

以上几种DTW下界距离的计算复杂度均低于DTW距离,并且对应的搜索方法不会遗漏正确结果,且在形式上同Euclidean距离较为接近,是相似性搜索的有效方法.但是,它们针对一元时间序列,仅适用于全序列搜索,且要求被度量的序列长度相等,因此存在一定的局限性.

Michail等<sup>[35]</sup>和李正欣等<sup>[36]</sup>将多元时间序列中每一时刻的记录值视为向量,分别将LB\_Keogh方法推广到多元时间序列,但只能处理DTW距离下的全序列搜索,不能解决子序列搜索问题.

Wang等<sup>[37]</sup>将多元时间序列拆分为多个一元时间序列,分别对每个一元时间序列进行DTW距离度量下的相似性搜索,再将搜索结果进行合成,作为多元时间序列的搜索结果.该方法仅适用于多元全序列搜索,它将多元时间序列视为多个一元时间序列的组合,割裂了变量之间的关系,对于变量相关性较强的多元时间序列并不适用.

#### 4.2.2 支持DTW距离度量的子序列搜索方法

与全序列搜索相比,子序列搜索的复杂性主要体现在搜索对象的急剧增长,这直接影响着索引构建以及相似性搜索方法.

后缀树(Suffix Tree)<sup>[38]</sup>是一种支持DTW距离度量的索引结构,以该索引为基础能够进行相似子序列搜索,结果具有非漏报性.但后缀树索引占用的磁盘空间比较大,甚至大于原始数据所占空间1到2个数量级<sup>[32]</sup>,较长的时间序列在构建后缀树索引时,基于内存的算法往往不能正常工作,因此很难应用于大型时间序列数据集.

此外,现有的基于后缀树的子序列搜索方法主要针对一元时间序列,如果要应用于多元时间序列,还需要进一步研究如何将各个时刻的记录值向量转换为离散字符,如何设计高效的多元子序列搜索方法,如何保证相应的搜索方法具有非漏报性等问题.

Sakurai等<sup>[39]</sup>、Zhou等<sup>[40]</sup>、Machiko等<sup>[41]</sup>分别提

出了时间序列数据流中支持DTW距离的相似子序列搜索方法,但这些方法都只针对一元时间序列,不能直接应用于多元时间序列。

Teddy等<sup>[29]</sup>借鉴Keogh提出的下界距离的基本思路,结合Euclidean距离下子序列搜索中的滑窗分割技术<sup>[25]</sup>,提出了一种支持DTW距离的子序列搜索方法,搜索效率优于顺序扫描方法,结果具有非漏报性,且索引占用空间明显小于后缀树方法。Doruk等<sup>[42]</sup>和Huang等<sup>[43]</sup>分别利用高性能硬件计算平台,优化DTW距离下的子序列搜索计算,进一步提高了计算效率。

然而,现有的基于滑窗分割的子序列搜索方法主要针对一元时间序列,如果应用于多元时间序列,还需要进一步研究如何用索引结构对多元时间序列进行组织,如何有效避免“维度灾难”,如何设计高效的多元子序列搜索方法等问题。

## 5 结 论

从国内外研究成果看,多元时间序列相似模式挖掘是当前研究的一个热点,目前大部分研究成果主要集中在多元时间序列的特征表示与相似模式度量,而序列搜索(尤其是子序列搜索)已成为制约问题突破的关键环节。序列搜索受相似模式度量的直接影响,现有的多元时间序列相似模式度量方法中,DTW距离能够度量长度不同的多元时间序列,支持时间轴的弯曲和伸缩等形变,且不存在误判风险,不需要进行符号化处理,具有较好的度量精度和鲁棒性。

现有的支持DTW距离度量的相似性搜索方法中,一元序列搜索多于多元序列搜索,全序列搜索多于子序列搜索,大部分研究成果主要集中在一元全序列搜索。因此,研究支持DTW距离度量的多元子序列搜索方法,是实现多元时间序列相似模式挖掘的一个研究方向。

### 参考文献(References)

- [1] 李海林. 基于变量相关性的多元时间序列特征表示[J]. 控制与决策, 2015, 30(3): 441-447.  
(Li H L. Feature representation of multivariate time series based on correlation among variables[J]. Control and Decision, 2015, 30(3): 441-447.)
- [2] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, Teh Ying Wah. Time-series clustering — A decade review[J]. Information Systems, 2015, 53: 16-38.
- [3] Zoltán Bankó, János Abonyi. Correlation based dynamic time warping of multivariate time series[J]. Expert Systems with Applications, 2012, 39(17): 12814-12823.
- [4] Yang Q, Wang X. 10 challenging problems in data mining research[J]. Int J of Information Technology and Decision Making, 2006, 5(4): 597-604.
- [5] Joan Serrà, Josep Ll Arcos. An empirical evaluation of similarity measures for time series classification[J]. Knowledge-Based Systems, 2014, 67: 305-314.
- [6] Robert Garthoff, Vasyi Golosnoy, Wolfgang Schmid. Monitoring the mean of multivariate financial time series[J]. Applied Stochastic Models in Business and Industry, 2014, 30(3): 328-340.
- [7] Bogdan Pogorelc, Matjaž Gams. Detecting gait-related health problems of the elderly using multidimensional dynamic time warping approach with semantic attributes[J]. Multimedia Tools and Applications, 2013, 66(1): 95-114.
- [8] Maëlle Salmon, Dirk Schumacher, Michael Höhle. Monitoring count time series in R: Aberration detection in public health surveillance[J]. J of Statistical Software, 2016, 70(10): 1-35.
- [9] Vicente Palazón-González, Andrés Marzal. On the dynamic time warping of cyclic sequences for shape retrieval[J]. Image and Vision Computing, 2012, 30(12): 978-990.
- [10] Huang Zi, Liu Jiajun, Cui Bin, et al. A gram-based string paradigm for efficient video subsequence search[J]. IEEE Trans on Multimedia, 2013, 15(3): 608-620.
- [11] 韩敏, 许美玲, 任伟杰. 多元混沌时间序列的相关状态机预测模型研究[J]. 自动化学报, 2014, 40(5): 822-829.  
(Han M, Xu M L, Ren W J. Research on multivariate chaotic time series prediction using mRSM model[J]. Acta Automatica Sinica, 2014, 40(5): 822-829.)
- [12] Ngan Chun-Kit, Brodsky Alexander. Optimal event monitoring through internet mashup over multivariate time series[J]. Int J of Decision Support System Technology, 2013, 5(2): 46-69.
- [13] Young-Seon Jeong, Myong K Jeong, Olufemi A Omitaomu. Weighted dynamic time warping for time series classification[J]. Pattern Recognition, 2011, 44(9): 2231-2240.
- [14] 吴虎胜, 张凤鸣, 钟斌. 基于二维奇异值分解的多元时间序列相似匹配方法[J]. 电子与信息学报, 2014, 36(4): 847-854.  
(Wu H S, Zhang F M, Zhong B. Similar pattern matching method for multivariate time series based on two-dimensional singular value decomposition[J]. J of Electronics & Information Technology, 2014, 36(4): 847-854.)
- [15] Helmuth Pree, Benjamin Herwig, Thiemo Gruber, et al. On general purpose time series similarity measures and their use as kernel functions in support vector machines[J]. Information Sciences, 2014, 281: 478-495.
- [16] 李正欣, 张凤鸣, 李克武. 多元时间序列模式匹配方法研究[J]. 控制与决策, 2011, 26(4): 565-570.  
(Li Z X, Zhang F M, Li K W. Research on pattern matching method for multivariate time series[J]. Control and Decision, 2011, 26(4): 565-570.)

- [17] Keogh E, Kaushik Chakrabarti, Michael J Pazzani, et al. Dimensionality reduction for fast similarity search in large time series databases[J]. Knowledge and Information Systems, 2001, 3(3): 263-286.
- [18] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series[C]. Proc of the Workshop on Knowledge Discovery in Databases. Seattle, 1994: 229-248.
- [19] Kiyong Yang, Cyrus Shahabi. An efficient  $k$  nearest neighbor search for multivariate time series[J]. Information and Computation, 2007, 205(1): 65-98.
- [20] 管河山, 姜青山, 王声瑞. 基于点分布特征的多元时间序列模式匹配方法[J]. 软件学报, 2009, 20(1): 67-79. (Guan H S, Jiang Q S, Wang S R. Pattern matching method based on point distribution for multivariate time series[J]. J of Software, 2009, 20(1): 67-79.)
- [21] Li H L, Guo C H, Qiu W R. Similarity measure based on piecewise linear approximation and derivative dynamic time warping for time series mining[J]. Expert Systems with Applications, 2011, 38(12): 14732-14743.
- [22] Vladimir Kurbalija, Miloš Radovanovic, Zoltan Geler, et al. The influence of global constraints on similarity measures for time-series databases[J]. Knowledge-Based Systems, 2014, 56: 49-67.
- [23] Vlachos M, Gunopulos D, Kollios G. Discovering similar multidimensional trajectories[C]. Proc of the 18th Int Conf Data Engineering. San Jose, 2002: 673-684.
- [24] Chen L, Ozsu M T, Oria V. Robust and fast similarity search for moving object trajectories[C]. Proc of the 2005 ACM SIGMOD Int Conf on Management of Data. Baltimore, 2005: 491-502.
- [25] Faloutsos C, Ranganathan M, Manolopoulos Y. Fast subsequence matching in time-series databases[C]. Proc of the ACM SIGMOD Conf on Management of Data. New York, 1994: 419-429.
- [26] Agrawal Rakesh, Srikant Ramakrishnan. Mining sequential patterns[C]. Proc of the 11th Int Conf on Data Engineering. Taipei, 1995: 3-14.
- [27] Chan Franky, Fu Wai-chee. Efficient time series matching by wavelets[C]. Proc of the 15th IEEE Int Conf on Data Engineering. Sydney, 1999: 126-133.
- [28] Yang-Sae Moon, Kyu-Young Whang, Woong-Kee Loh. Duality-based subsequence matching in time-series data-bases[C]. Proc of the Int Conf on Data Engineering. Heidelberg, 2001: 263-272.
- [29] Teddy Siu Fung Wong, Man Hon Wong. Efficient subsequence matching for sequences databases under time warping[C]. Proc of the 7th Int Database Engineering and Applications Symposium. Hong Kong, 2003: 139-148.
- [30] Yi B, Jagadish H V, Faloutsos C. Efficient retrieval of similar time sequences under time warping[C]. Proc of the 14th Int Conf on Data Engineering. Washington DC, 1998: 201-208.
- [31] Sang-Wook Kim, Park Sanghyun, Wesley W Chu. An index-based approach for similarity search supporting time warping in large sequence databases[C]. Proc of the 17th Int Conf on Data Engineering. Heidelberg, 2001: 607-614.
- [32] Keogh E, Ratanamahatana C. Exact indexing of dynamic time warping[J]. Knowledge and Information Systems, 2005, 7(3): 358-386.
- [33] Yunyue Zhu, Dennis Shasha. Warping indexes with envelope transforms for query by humming[C]. Proc of the 2003 ACM SIGMOD Int Conf on Management of Data. San Diego, 2003: 181-192.
- [34] Mi Zhou, Man Hon Wong. Boundary-based lower-bound functions for dynamic time warping and their indexing[J]. Information Sciences, 2011, 181(19): 4175-4196.
- [35] Michail Vlachos, Marios Hadjieleftheriou, Dimitrios Gunopulos, et al. Indexing multidimensional time series[J]. The VLDB J, 2006, 15(1): 1-20.
- [36] 李正欣, 张凤鸣, 李克武, 等. 一种支持DTW距离的多元时间序列索引结构[J]. 软件学报, 2014, 25(3): 560-575. (Li Z X, Zhang F M, Li K W, et al. Index structure for multivariate time series under DTW distance metric[J]. J of Software, 2014, 25(3): 560-575.)
- [37] Jimin Wang, Yuelong Zhu, Shijin Li, et al. Multivariate time series similarity searching[J]. The Scientific World J, 2014: 1-8.
- [38] Sanghyun Park, Wesley W Chu, Jeehee Yoon, et al. Similarity search of time-warped subsequences via a suffix tree[J]. Information Systems, 2003, 28(7): 867-883.
- [39] Sakurai Y, Faloutsos C, Yamamuro M. Stream monitoring under the time warping distance[C]. Proc of the 23th Int Conf on Data Engineering. Istanbul, 2007: 1046-1055.
- [40] Mi Zhou, Man Hon Wong. Efficient online subsequence searching in data streams under dynamic time warping distance[C]. Proc of the 24th Int Conf on Data Engineering. Cancun, 2008: 686-695.
- [41] Machiko Toyoda, Yasushi Sakurai, Yoshiharu Ishikawa. Pattern discovery in data streams under the time warping distance[J]. The VLDB J, 2013, 22(3): 295-318.
- [42] Doruk Sart, Abdullah Mueen, Walid Najjar, et al. Accelerating dynamic time warping subsequence search with GPUs and FPGAs[C]. IEEE Int Conf on Data Mining. Sydney, 2010: 1001-1006.
- [43] Sitao Huang, Guohao Dai, Yuliang Sun, et al. DTW-based subsequence similarity search on AMD heterogeneous computing platform[C]. The 15th IEEE Int Conf on High Performance Computing and Communications. Zhangjiajie, 2013: 1054-1063.

(责任编辑: 曹洪武)