

文章编号: 1001-0920(2011)04-0565-06

## 多元时间序列模式匹配方法研究

李正欣, 张凤鸣, 李克武

(空军工程大学 工程学院, 西安 710038)

**摘要:** 针对多元时间序列模式匹配的方法难以高效、准确地刻画序列相似程度的问题, 在考虑变量的量纲和特征差异的基础上, 对多元时间序列进行多维分段拟合; 然后, 选取各个变量维度上拟合线段的倾斜角和时间跨度作为模式的描述方式, 提出一种基于动态时间弯曲(DTW)的多元时间序列趋势距离匹配方法. 实验结果表明, 所提出的模式匹配方法对由连续型变量组成、时间跨度较大且体现一个连续、完整动作过程的多元时间序列, 具有较好的匹配效果.

**关键词:** 多元时间序列; 多维分段拟合; 动态时间弯曲; 模式匹配

**中图分类号:** TP311

**文献标识码:** A

### Research on pattern matching method for multivariate time series

LI Zheng-xin, ZHANG Feng-ming, LI Ke-wu

(Engineering Institute, Air Force Engineering University, Xi'an 710038, China. Correspondent: LI Zheng-xin, E-mail: lizhengxin\_2005@163.com)

**Abstract:** Common methods for matching multivariate time series can't measure their similarity rapidly and accurately. Multivariate time series are fitted with multidimensional piecewise method on the basis of considering feature difference of different variables. Then the angle of inclination and time span of a fitting line segment in a certain variable dimension are chosen as feature pattern. A pattern matching method based on dynamic time warping(DTW) is proposed for multivariate time series. Finally, the experimental results show that the proposed method can measure the similarity of multivariate time series rapidly and accurately, which are composed of continuous variables and can present a whole action process in a comparatively long time.

**Key words:** multivariate time series; multidimensional piecewise fitting; dynamic time warping; pattern matching

## 1 引言

现实世界中存在大量的多元时间序列类型的数据, 如航天飞船等重要仪器的运行状态数据、互联网中关键服务器的通讯流量数据、应用于多种行业的人体运动捕捉数据等. 广义上讲, 任何包含多变量数据存储的数据集均可视为多元时间序列.

时间序列数据挖掘是针对时间序列的模式发现过程, 旨在研究隐含在时间序列中更深层次的知识<sup>[1]</sup>. 时间序列数据挖掘包括时间序列的相似性查询、趋势分析、周期分析、时态关联规则挖掘、异常检测等内容. 时间序列相似性查询是从时间序列数据库中发现用户感兴趣的模式, 它不仅是时间序列数据挖掘的主要研究内容之一, 而且是实现其他挖掘任务的基础<sup>[2]</sup>. 模式匹配是度量时间序列相似程度的方法, 同时也是时间序列相似性查询的一项核心技术, 在时间

序列分析处理中占据基础性地位.

当前的模式匹配方法研究主要针对一元时间序列, 并产生了一套比较成熟的理论<sup>[3]</sup>; 而针对多元时间序列的模式匹配研究相对较少, 还很不成熟<sup>[3-4]</sup>. 事物状态的刻画往往需要多个变量共同确定, 各个变量间通常存在一定的关联关系, 不能将多元时间序列看成多个一元时间序列的简单叠加. 因此, 研究多元时间序列的模式匹配方法具有现实的理论意义和广阔的应用前景.

本文对多元时间序列进行多维分段拟合, 选取各变量维度上拟合线段的倾斜角和时间跨度作为特征模式, 提出一种基于动态时间弯曲(DTW)的多元时间序列趋势距离匹配方法, 并通过实验表明了该方法的有效性.

收稿日期: 2010-01-03; 修回日期: 2010-05-05.

作者简介: 李正欣(1982-), 男, 博士生, 从事信息系统工程与智能决策、数据挖掘的研究; 张凤鸣(1963-), 男, 教授, 博士生导师, 从事信息系统工程与智能决策等研究.

## 2 相关研究

多元时间序列不同变量之间的相关性使其相似性度量的研究成果远不及一元时间序列<sup>[4]</sup>. 目前, 多元时间序列的模式匹配方法主要有 Minkowski 距离、动态时间弯曲 (DTW) 距离、奇异值分解 (SVD) 和基于点分布特征 (PD) 等方法.

Minkowski 距离的优点是计算简单, 容易理解, 在正交变换下保持不变. 但是 Minkowski 距离度量方法要求两条时间序列的长度必须相等, 对时间轴伸缩和弯曲问题无能为力.

Berndt 和 Clifford<sup>[5]</sup>把在语音识别中广泛使用的 DTW 距离引入时间序列的相似性研究中. 之后, 有人将 DTW 距离扩展到多元时间序列的相似性度量<sup>[6]</sup>. DTW 距离定义了时间序列之间的最佳对齐匹配关系, 支持时间轴的伸缩和弯曲, 但其计算复杂度高, 不适于海量时间序列的数据挖掘.

SVD 方法将时间序列中的变量视为随机变量, 将各时间点上的观察值视为样本点, 采用相关系数矩阵作为特征提取的基础, 依据线性空间中的坐标变换原理构造模式匹配模型. 它是一种统计意义上的度量方法, 没有体现出观察值的时序关系, 因此存在误判风险, 且计算复杂度较高.

PD 方法<sup>[4]</sup>抽取局部重要点作为多元时间序列的特征, 根据重要点的分布特征构建模式匹配模型. 然而该方法在提取局部重要点时, 没有考虑多元时间序列不同变量间的量纲和特征差异, 将不同变量的观察值放在一起进行比较, 因此具有一定的局限性.

## 3 多元时间序列的特征模式提取

### 3.1 多元时间序列的表示方法

一系列观察值  $x_t(j)$  称为多元时间序列. 其中:  $t(t = 1, 2, \dots, n)$  表示第  $t$  个时间点,  $j(j = 1, 2, \dots, m)$  表示第  $j$  个变量,  $x_t(j)$  表示第  $j$  个变量在第  $t$  个时间点上的观察值. 一个多元时间序列可以用一个  $m \times n$  的矩阵表示,  $m$  为变量个数,  $n$  为时间点数量.

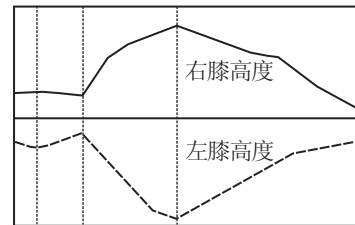
### 3.2 多元时间序列的多维分段拟合

已有的算法大多使用全局特征刻画多元时间序列<sup>[7]</sup>. 由于大部分多元时间序列具有短期波动频繁、大量噪声干扰、非稳态及混沌等非线性特征, 对其采用局部特征进行刻画更能体现其形状细节.

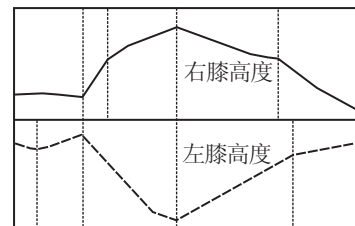
提取局部特征, 最直接的方法是在各个变量维度中独立计算, 然后将所获得的特征顺次排列以构成特征向量, 即将多元问题转化为一元问题来解决. 然而, 这种将各变量维度割裂开来的方法忽视了各变量间相关信息的存在. 例如, 在图 1(a) 所记录的人体行走动作中, 左膝盖与右膝盖的位置就是两个高度相关

的维度.

一个变量维度中出现的特征在其他相关的变量维度中也会有所体现. 所以, 理想的分段结果应是在全部相关的变量维度上生成相同或近似相同的序列段. 由于多元时间序列包含噪声和误差的特点, 如果在各个变量维度上独立进行操作, 则产生的分段结果将不同, 如图 1(b) 所示. 这种分段结果不仅没有意义, 而且仍然存在问题. 基于以上分析, 本文同时对全部相关维度进行分段操作, 即多维分段拟合.



(a) 多维分段结果



(b) 单维分段结果

图 1 对高度相关的多元时间序列进行分段的两种方式

为了更有利于获得理想的分段结果<sup>[8]</sup>, 本文采用自底向上的分段策略. 首先将一个长度为  $n$  的多元时间序列分为  $n/2$  个段, 代表最细致的分段; 然后, 递归地合并产生拟合误差最低的一对相邻段; 直至任何两个相邻段合并的拟合误差都大于给定的最大误差 ( $\maxError$ ) 时, 分段完成.

用  $d_1, d_2, \dots, d_j$  表示一个拟合段中  $j$  个数据点与相应的拟合曲线的垂直距离, 则该拟合段在第  $i$  个变量维度上的标准误差可定义为

$$e_i = \left( \sum_{k=1}^j d_k^2 \right)^{\frac{1}{2}}, \quad i = 1, 2, \dots, m. \quad (1)$$

将全部变量维度上的标准误差进行求和, 可得到一个拟合段的总误差, 即

$$e_{seg} = \sum_{k=1}^m \sigma_k e_k, \quad \sum_{k=1}^m \sigma_k = 1. \quad (2)$$

因为不同变量的量纲和特征存在差异, 在模式匹配中的重要性也不完全相同, 所以式 (2) 在计算一个拟合段的总误差时, 加入了变量维度的权重系数. 其中:  $\sigma_k$  为第  $k$  个变量的误差权重值,  $e_{seg}$  为使用多维分段拟合方法在一个拟合段中的信息丢失量. 为了实现多维分段拟合, 在合并一对相邻段时, 误差的构造反映了全部变量维度上的信息丢失量, 而不是仅限于个别

维度.

文献[7]使用不同阶次的多项式分别对4种多元时间序列数据集进行分段拟合实验.结果表明,使用一阶多项式进行拟合所得的结果精度较高;而更高阶的多项式拟合并不能产生更好的效果,且需要花费额外的计算时间.因此,本文采用一阶多项式对多元时间序列进行多维分段拟合操作.

### 3.3 多元时间序列的特征提取与转换

多维分段拟合后,选取拟合多项式的一次项系数 $k$ 和该段序列的时间跨度 $l$ 作为一个分段在某一变量维度上的特征,即只需要一个特征向量 $(k, l)$ 便可表示一个分段在某个变量维度上的特征.即当一个含有 $m$ 个变量维度的多元时间序列被分割为 $S$ 段时,该序列可用一个 $m \times S$ 的矩阵表示,即

$$\begin{bmatrix} (k_{11}, l_1) & (k_{12}, l_2) & \cdots & (k_{1S}, l_S) \\ (k_{21}, l_1) & (k_{22}, l_2) & \cdots & (k_{2S}, l_S) \\ \vdots & \vdots & \ddots & \vdots \\ (k_{m1}, l_1) & (k_{m2}, l_2) & \cdots & (k_{mS}, l_S) \end{bmatrix}. \quad (3)$$

其中: $k_{ij}(i = 1, 2, \dots, m; j = 1, 2, \dots, S)$ 为拟合线段的斜率,其定义区间为 $(-\infty, +\infty)$ ;  $l_t(t = 1, 2, \dots, S)$ 为拟合线段的时间跨度,其定义区间为 $(0, +\infty)$ .由于 $k_{ij}$ 和 $l_t$ 的定义区间不是有限的,使用模式匹配模型度量两条“基本相似”的多元时间序列时,可能会因某两个拟合段在某一变量维度上的差异值过大而造成“不相似”的误判.为避免这种误判,在模式匹配前,先通过下式:

$$\alpha_{ij} = \arctan k_{ij}, \alpha_{ij} \in (-\pi/2, \pi/2), \quad (4)$$

$$i = 1, 2, \dots, m, j = 1, 2, \dots, S,$$

将斜率转换为倾斜角;然后通过下式:

$$d_i = l_i / \sum_{j=1}^S l_j, d_i \in (0, 1), i = 1, 2, \dots, S, \quad (5)$$

将时间跨度转换为时间跨度与时间序列长度的比值;最后得到转换后的特征矩阵为

$$\begin{bmatrix} (\alpha_{11}, d_1) & (\alpha_{12}, d_2) & \cdots & (\alpha_{1S}, d_S) \\ (\alpha_{21}, d_1) & (\alpha_{22}, d_2) & \cdots & (\alpha_{2S}, d_S) \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha_{m1}, d_1) & (\alpha_{m2}, d_2) & \cdots & (\alpha_{mS}, d_S) \end{bmatrix}. \quad (6)$$

## 4 多元时间序列的模式匹配模型

经过特征提取和转换后,多元时间序列特征矩阵的行数是相同的,即它们的变量维度是一一对应的关系;由于分段数量可能不同,特征矩阵的列数一般不同.于是要求构造的模式匹配模型应能对行数相同、列数不同的特征矩阵进行比较.

DTW实际上就是确定两个序列之间的对齐匹配

关系.当序列发生时间轴弯曲时,可以在弯曲部分进行自我复制,使两条时间序列之间的相似波形进行对齐匹配.此外,它还支持不同长度的序列之间的匹配.为此,本文提出一种基于DTW的趋势距离(TD)匹配方法.设多元时间序列 $T_a$ 和 $T_b$ 经过特征提取与转换后的特征矩阵分别为 $A, B$ ,即

$$A = \begin{bmatrix} (\alpha_{11}, d_1) & (\alpha_{12}, d_2) & \cdots & (\alpha_{1S}, d_S) \\ (\alpha_{21}, d_1) & (\alpha_{22}, d_2) & \cdots & (\alpha_{2S}, d_S) \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha_{m1}, d_1) & (\alpha_{m2}, d_2) & \cdots & (\alpha_{mS}, d_S) \end{bmatrix} = [a_1, a_2, \dots, a_s], \quad (7)$$

$$B = \begin{bmatrix} (\alpha'_{11}, d'_1) & (\alpha'_{12}, d'_2) & \cdots & (\alpha'_{1S'}, d'_{S'}) \\ (\alpha'_{21}, d'_1) & (\alpha'_{22}, d'_2) & \cdots & (\alpha'_{2S'}, d'_{S'}) \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha'_{m1}, d'_1) & (\alpha'_{m2}, d'_2) & \cdots & (\alpha'_{mS'}, d'_{S'}) \end{bmatrix} = [b_1, b_2, \dots, b_{s'}]. \quad (8)$$

其中: $a_i(i = 1, 2, \dots, S)$ ,  $b_j(j = 1, 2, \dots, S')$ 分别为特征矩阵 $A, B$ 的列向量,代表多元时间序列的每一分段在各个变量维度上的局部特征.

$T_a, T_b$ 的趋势距离匹配方法定义为

$$\text{TD}_{\text{dtw}}(T_a, T_b) =$$

$$\text{TD}_{\text{base}}(a_1, b_1) + \min \begin{cases} \text{TD}_{\text{dtw}}(A, B[2: -]), \\ \text{TD}_{\text{dtw}}(A[2: -], B), \\ \text{TD}_{\text{dtw}}(A[2: -], B[2: -]). \end{cases} \quad (9)$$

其中: $X[i: -]$ 表示特征矩阵 $X$ 的第 $i$ 个列向量到最后一个列向量组成的子序列; $\text{TD}_{\text{base}}(a_i, b_j)$ 表示 $a_i$ 与 $b_j$ 的趋势距离,其计算方法为

$$\text{TD}_{\text{base}}(a_i, b_j) = \sum_{t=1}^m \omega_t [\varepsilon_t |\alpha_{ti} - \alpha'_{tj}| + \lambda_t |d_i - d'_j|], \quad (10)$$

它体现了 $T_a$ 的第 $i$ 段拟合与 $T_b$ 的第 $j$ 段拟合在各个变量维度上的累计差异.

多元时间序列在进行模式匹配时,各个变量的量纲不同,重要程度也有差别,因此,需要分配不同的权重值加以调整.在式(10)中, $\omega_t$ 表示第 $t$ 个变量在计算趋势距离时的权重值,同时有下式成立:

$$\sum_{t=1}^m \omega_t = 1; \quad (11)$$

$|\alpha_{ti} - \alpha'_{tj}|$ 和 $|d_i - d'_j|$ 分别表示 $T_a$ 的第 $i$ 段拟合与 $T_b$ 的第 $j$ 段拟合在第 $t$ 个变量维度上的倾斜角与时间跨度的差异; $\varepsilon_t$ 和 $\lambda_t$ 分别表示在第 $t$ 个变量维度上,倾斜角差异与时间跨度差异的权重值,同时有下式成

立:

$$\varepsilon_t + \lambda_t = 1, t = 1, 2, \dots, m. \quad (12)$$

## 5 实验结果与分析

### 5.1 实验方法与实验数据

为了便于比较模式匹配方法的性能, 选用已知分类结果的多元时间序列, 采用  $k$ -近邻的方法进行实验.

假定实验数据集中含有  $n$  个多元时间序列, 从中任意抽取一个作为输入样本  $X$ . 使用相应的模式匹配方法找出与  $X$  最相似的“ $k$  个样本” (实验中  $k$  分别取 1, 5 和 10). 统计找出的这“ $k$  个样本”中与  $X$  类别相同的样本个数  $n_0$ , 并按下式计算准确率  $e$ :

$$e = n_0/k. \quad (13)$$

按照文中的实验方法,  $e$  的取值可能为  $\{0, 0.1, 0.2, \dots, 1\}$ .

将数据集中其他多元时间序列一一作为输入样本, 重复以上实验, 便可得到  $n$  个模式匹配的准确率. 如果将准确率作为一个离散型随机变量  $\varepsilon$ , 则准确率的数学期望可按下式确定:

$$e^* = \sum_{t=1}^{11} p(\varepsilon = e_t) \times e_t, \quad (14)$$

并用其作为各种模式匹配方法的比较依据.

实验中选取两组多元时间序列数据集: 手语信号数据集<sup>[9]</sup>和 Robot Execution Failure 数据集<sup>[10]</sup>.

手语信号是由高性能的定位跟踪器和测量手

套捕捉获取, 采样周期为 10 ms, 包含 22 个连续型变量. 其中: 左、右手的动作特征各用 11 个变量刻画: 6 个变量 (对应 6 个自由度) 表示手的位置, 5 个变量表示各手指的弯曲程度. 手语信号数据集包含 95 种语意 (95 个类), 每种语意都有 27 组序列. 不失一般性, 选取前 8 种语意对应的序列作为实验数据集 (记为 SL), 共有 216 个实验样本. 8 种语意分别为: alive, all, answer, boy, building, buy, change-mind, cold. 216 组多元时间序列的时间跨度在 47~95 之间, 每组序列都体现一个连续、完整的手语动作过程.

Robot Execution Failure 是对 Robot 进行故障监控的数据集, 采样周期为 21 ms, 包含 6 个离散型变量: 3 个变量表示受力, 3 个变量表示扭矩. 该数据集包含 5 个子数据集, 选取第 1 个子数据集 LP1 进行实验. 已知 LP1 数据集分为 4 类, 即 normal, collision, fr-collision 和 obstruction. 共有 88 个样本, 每个样本均为  $6 \times 15$  的矩阵, 属于时间跨度较小、体现某些状态点的多元时间序列.

### 5.2 TD 方法的模型参数选择

TD 方法的模型参数选择就是确定多维分段拟合中的  $\sigma$ , maxError 和式 (10) 中的  $\varepsilon$ ,  $\lambda$  以及  $\omega$  的取值. 以 SL 数据集为例, 不失一般性, 不妨先假设各个变量的重要性相同, 即各个变量维度上的  $\sigma$ ,  $\omega$  相等;  $\varepsilon$  和  $\lambda$  在变量维度上不存在差异. maxError 取 0.03 时, 考虑选择不同的  $\varepsilon$  和  $\lambda$  时, TD 方法准确率的数学期望, 结果如表 1 所示.

表 1 TD 方法中选择  $\varepsilon$ ,  $\lambda$  的实验结果

$k$	$\varepsilon = 0, \lambda = 1$			$\varepsilon = 0.2, \lambda = 0.8$			$\varepsilon = 0.5, \lambda = 0.5$			$\varepsilon = 0.8, \lambda = 0.2$			$\varepsilon = 1, \lambda = 0$		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
准确率	0.22	0.22	0.23	0.58	0.51	0.45	0.90	0.85	0.78	0.97	0.96	0.91	0.99	0.94	0.89

实验结果表明, 针对 SL 数据集的模式匹配, 当  $\varepsilon = 0.8, \lambda = 0.2$  时, 准确率的数学期望最高.  $\varepsilon = 1, \lambda = 0$  时的准确率不如  $\varepsilon = 0.8, \lambda = 0.2$  时的准确率, 这表明了在 TD 方法中考虑时间跨度因素的必要性.

maxError 决定着多元时间序列的分段数量: 取值过小, 分段数量会过多, 算法效率降低; 取值过大, 导致分割过于粗糙, 无法准确地刻画特征, 造成算法的准确率降低. maxError,  $\sigma$ ,  $\omega$  均采用对样本数据集进行训练的方法加以确定, 具体步骤类似于上文中  $\varepsilon$  和  $\lambda$  的确定方法.

### 5.3 几种模式匹配方法的比较

针对 SL 数据集, 分别采用 TD, SVD 和 PD<sup>[4]</sup> 这 3 种方法进行模式匹配, 并计算相应的准确率 (保留 2 位有效数字), 结果见表 2. 其中: TD 方法模型参数的

选择与上节实验相同; PD 方法中, 分割形式为  $X[i-1:i+1, j-1:j+1]$ . 在成功率为小概率事件的情况下, TD 方法所对应的次数比 SVD 方法和 PD 方法都少; 而在成功率为大概率事件的情况下, TD 方法所对应的次数比 SVD 方法和 PD 方法都要多. 特别地, 当准确率为 100% 时, TD 方法所对应的次数远远多于 SVD 方法和 PD 方法.

进一步, 分别计算 3 种方法准确率的数学期望, 见表 3. 可以看出, TD 方法的准确率期望值明显高于 SVD 方法和 PD 方法.

为了形象地说明不同方法的匹配结果, 采用 SL 数据集中第 3 个样本 (记为 SL\_3) 作为输入样本, 使用 3 种方法分别进行模式匹配, 得到最相似模式所对应的样本图形, 如图 2 所示. 相比之下, TD 方法得到

表2 3种模式匹配方法的实验结果(N表示个数,R表示比率)

k	TD						SVD						PD					
	1		5		10		1		5		10		1		5		10	
e	N	R	N	R	N	R	N	R	N	R	N	R	N	R	N	R	N	R
0	6	0.03	0	0	0	0	54	0.25	16	0.07	11	0.05	91	0.42	33	0.15	14	0.06
10%	—	—	—	—	0	0	—	—	—	—	14	0.06	—	—	—	—	13	0.06
20%	—	—	0	0	0	0	—	—	32	0.15	27	0.13	—	—	25	0.12	17	0.08
30%	—	—	—	—	2	0.01	—	—	—	—	16	0.07	—	—	—	—	28	0.13
40%	—	—	4	0.02	3	0.01	—	—	36	0.17	23	0.11	—	—	36	0.17	18	0.08
50%	—	—	—	—	9	0.04	—	—	—	—	23	0.11	—	—	—	—	26	0.12
60%	—	—	9	0.04	8	0.04	—	—	29	0.13	17	0.08	—	—	34	0.16	17	0.08
70%	—	—	—	—	7	0.03	—	—	—	—	11	0.05	—	—	—	—	13	0.06
80%	—	—	15	0.07	22	0.10	—	—	20	0.09	7	0.03	—	—	23	0.11	6	0.03
90%	—	—	—	—	30	0.14	—	—	—	—	7	0.03	—	—	—	—	12	0.06
100%	210	0.97	188	0.87	135	0.63	162	0.75	83	0.38	60	0.28	125	0.58	65	0.30	52	0.24

表3 SL数据集的准确率期望值

k	TD	SVD	PD
10	0.91	0.57	0.56
5	0.96	0.64	0.57
1	0.97	0.75	0.58

表4 LP1数据集的准确率期望值

k	TD	SVD	PD
10	0.51	0.46	0.76
5	0.59	0.47	0.81
1	0.64	0.48	0.83

的样本与输入样本在形状上具备很大的相似性,而SVD方法得到的样本与输入样本在形状上差异较大.这是因为TD方法不是依据统计原理,而是按照序列的对齐匹配关系计算相似程度,能有效刻画出多元时间序列的局部形状特征.

原因在于,TD方法在特征提取时,用拟合线段的斜率和时间跨度作为特征,体现的是多元时间序列的连续变化趋势,而非各个状态点的具体数值,因而无法很好地刻画出时间跨度较小、只体现某些状态点的多元时间序列的特征,所以对其匹配效果不佳.

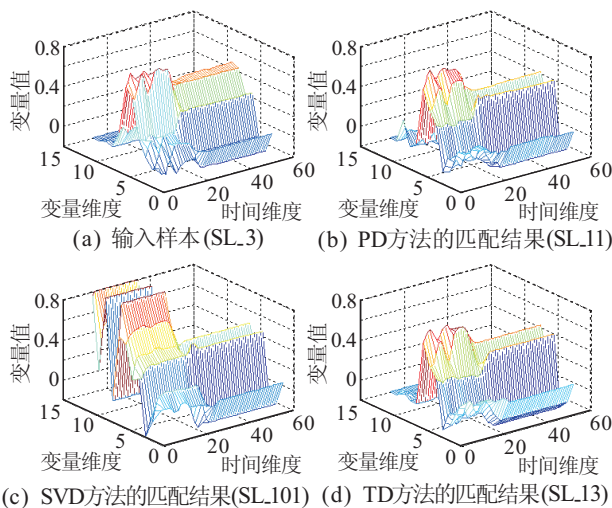


图2 SL数据集的相似模式匹配结果

下面针对LP1数据集,分别采用3种方法进行模式匹配,并计算相应准确率的数学期望,所得结果如表4所示.其中TD方法模型参数的选择与第1个实验相同;PD方法中,分割形式为 $X[i-1:i+1, j-1:j+1]$ .

由表4可见,PD方法的准确率期望值均在80%左右,明显优于TD方法和SVD方法;TD方法的准确率期望值较低,很难达到模式匹配的应用要求.主要

## 6 结论

本文使用多维分段拟合的方法,提取多元时间序列的局部形状特征,提出了一种基于DTW的趋势距离匹配方法,并通过实验表明了方法的有效性.与其他几种模式匹配方法相比,TD方法的优点是:1)同时考虑了拟合线段的斜率和时间跨度,采用DTW算法实现序列的对齐匹配,因此能够支持多元时间序列的振幅平移和伸缩、时间轴伸缩和弯曲;2)与DTW方法相比,TD方法实现了原始序列的大幅度降维,提高了计算效率;3)与SVD和PD方法相比,TD方法能够体现观察值的时序关系,消除了误判风险.实验结果表明,TD方法对由连续型变量组成、时间跨度较大且体现一个连续、完整动作过程的多元时间序列具有较好的匹配效果,但不适于由离散型变量组成、时间跨度较小、只体现某些状态点的多元时间序列.

## 参考文献(References)

[1] Roddick J F, Spiliopoulou M. A survey of temporal knowledge discovery paradigms and methods[J]. Trans on Data Engineering, 2002, 14(4): 750-767.  
 [2] 曲吉林. 时间序列挖掘中索引与查询技术的研究[D]. 天津: 天津大学管理学院, 2006.

- (Qu J L. Indexing and querying of time series in data mining[D]. Tianjin: School of Management, Tianjin University, 2006.)
- [3] 杨兴江, 周勇. 多元时间序列相似性研究[J]. 西南民族大学学报, 2007, 33(4): 864-869.  
(Yang X J, Zhou Y. Research on similarity of the multivariate time series[J]. J of Southwest University for Nationalities, 2007, 33(4): 864-869.)
- [4] 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.)
- [5] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series[C]. Working Notes of the Knowledge Discovery in Databases Workshop. Seattle, 1994: 359-370.
- [6] Michail Vlachos, Marios Hadjieleftheriou, Dimitrios Gunopulos, et al. Indexing multi-dimensional time-series with support for multiple distance measures[C]. Proc of the 9th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. Washington, 2003: 216-255.
- [7] 高歌. 多维时间序列分类技术[D]. 杭州: 浙江大学计算机学院, 2008.  
(Gao G. Extending the classification paradigm to multivariate time series[D]. Hangzhou: College of Computer Science, Zhejiang University, 2008.)
- [8] Eamonn J Keogh, Selina Chu, David Hart, et al. An online algorithm for segmenting time series[C]. Proc of IEEE Int Conf on Data Mining. Los Alamitos, 2001: 289-296.
- [9] Mohammed Waleed Kadous. High-quality recordings of Australian sign language signs[EB/OL]. (2002). [http://kdd.ics.uci.edu/databases/High-quality Australian Sign Language/High-quality Australian Sign Language.html](http://kdd.ics.uci.edu/databases/High-quality%20Australian%20Sign%20Language/High-quality%20Australian%20Sign%20Language.html).
- [10] Luis Seabra Lopes, Luis M Camarinha-Matos. Robot execution failures[EB/OL]. (1999). <http://kdd.ics.uci.edu/databases/robotfailure/robotfailure.html>.

(上接第564页)

- [3] Ma X-X, Zhao Q-Z, Qu J L. Robust portfolio optimization with a generalized expected utility model under ambiguity[J]. Annals of Finance, 2008, 4(4): 431-444.
- [4] Fabozzi F J, Huang D-H, Zhou G-F. Robust portfolios: Contributions from operations research and finance[J]. Annals of Operations Research, 2010, 176(1): 1-5.
- [5] 高莹, 黄小原. 具有VaR约束的跟踪误差投资组合鲁棒优化模型[J]. 中国管理科学, 2007, 15(1): 1-5.  
(Gao Y, Huang X Y. Robust optimal tracking error portfolio models based on VaR[J]. Chinese J of Management Science, 2007, 15(1): 1-5.)
- [6] 高金窑, 李仲飞. 模型不确定性条件下的Robust投资组合有效前沿与CAPM[C]. 第6届风险管理国际研讨会暨第7届金融系统工程研讨会. 南京, 2009: 1383-1394.  
(Gao J Y, Li Z F. On the robust portfolio frontier and CAPM under model uncertainty[C]. Visited Talk at The 6th Int Conf on Risk Management & The 7th Int Conf on Financial System Engineering. Nanjing, 2009: 1383-1394.)
- [7] Tütüncü R H, Koenig M. Robust asset allocation[J]. Annals of Operations Research, 2004, 132(2): 157-187.
- [8] Lu Z. Robust portfolio selection based on a joint ellipsoidal uncertainty set[J/OL]. <http://www.informaworld.com/smpp/content/content=a915794216/db=all/jumtype=rss>.
- [9] Lu Z. A computational study on robust portfolio selection based on a joint ellipsoidal uncertainty set[J/OL]. <http://www.springerlink.com/content/w5j56614r8033221/>.
- [10] El Ghaoui L, Oks M, Oustry F. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach[J]. Operations Research, 2003, 51(5): 543-556.
- [11] Zhu S, Fukushima M. Worst-case conditional value-at-risk with application to robust portfolio management[J]. Operations Research, 2009, 57(5): 1155-1168.
- [12] Bertsekas D. Nonlinear programming[M]. 2nd ed. New York: Athena Scientific, 1984: 511-519.
- [13] Boyd S, El Ghaoui L, Feron E, et al. Linear matrix inequalities in system and control theory[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1994: 10-32.
- [14] Horn R A, Johnson C R. Matrix analysis[M]. New York: Cambridge University Press, 1985: 112-131.
- [15] Toh K-C, Todd Michael J, Tutuncu Reha H. SDPT3 version 4.0-a MATLAB software for semidefinite quadratic linear programming[J]. Mathematical Programming Ser B, 2003, 95(2): 189-217.