

一种新的 DTW 最佳弯曲窗口学习方法

陈 乾^{1,2} 胡谷雨²

(解放军理工大学通信工程学院 南京 210007)¹ (解放军理工大学指挥自动化学院 南京 210007)²

摘 要 时间序列相似性查询中,DTW(Dynamic Time Warping)距离是支持时间弯曲的经典度量,约束弯曲窗口的 DTW 是 DTW 最常见的实用形式。分析了传统 DTW 最佳弯曲窗口学习方法存在的问题,并在此基础上引入时间距离的概念,提出了新的 DTW 最佳弯曲窗口学习方法。由于时间距离是 DTW 计算的附属产物,因此该方法可以在几乎不增加运算量的情况下提高 DTW 的分类精度。实验证明,采用了新的学习方法后,具有最佳弯曲窗口的 DTW 分类精度得到明显改善,分类精度优于 ERP(Edit Distance with Real Penalty)和 LCSS(Longest Common SubSequence),接近 TWED(Time Warp Edit Distance)的水平。

关键词 时间序列,相似性度量,动态时间弯曲,弯曲路径,时间距离

中图分类号 TP391.4 文献标识码 A

New Learning Method for Optimal Warping Window of DTW

CHEN Qian^{1,2} HU Gu-yu²

(Institute of Communication Engineering,PLA University of Science and Technology,Nanjing 210007,China)¹

(Institute of Command Automation,PLA University of Science and Technology,Nanjing 210007,China)²

Abstract The dynamic time warping is a classic similarity measure which can handle time warping issue in similarity computation of time series,and the DTW with constrained warping window is the most common and practical form of DTW. After systematically analyzing the traditional learning method for optimal warping window of DTW,we introduced time distance to measure the time deviation between two time series,and proposed a new leaning method for optimal warping window based on time distance. Since the time distance is an appurtenant of the DTW computation,the new method can improve DTW classification accuracy with little additional computation. Experimental data show that the optimal DTW with best warping window gets better classification accuracy when the new learning method is employed. What is more,the classification accuracy is better than the ERP(Edit Distance with Real Penalty) and the LCSS(Longest Common SubSequence),and is close to the TWED(Time Warp Edit Distance).

Keywords Time series, Similarity measure, Dynamic time warping, Warping path, Time distance

时间序列(Time Series)是指按照时间先后顺序排列的各个观测记录的有序集合。时间序列数据产生于日常生活中的各个领域,如网络安全监测数据、股票市场每日的波动数据、医学及生物学上的实验观察数据、工业监控传感器数据等等。相应的时间序列相似性查询具有广泛和实际的应用需求,如希望从网络安全监测历史数据中找到与当前数据相似的,以确定其是否受到攻击,以及学习攻击模式;医学上心电图异常模式发现;工业监控上的故障报警等。

相似性度量是相似性查询的基础。时间序列相似性度量中,最常用的是欧氏距离(Euclidean Distance),其优点是简单、快速,满足三角不等式,支持各种索引技术。欧氏距离的一些改进^[1,4]可以支持时间序列的振幅平移和伸缩,但仍不能处理局部的时间弯曲。支持时间弯曲的经典弹性度量是动态时间弯曲(Dynamic Time Warping)距离^[2],如图 1 所示,DTW 距离根据最小代价的时间弯曲路径进行匹配,可以很

好地解决时间序列发生时间弯曲后的相似性度量问题。DTW 的主要问题是时间复杂度高,且不满足三角不等式。

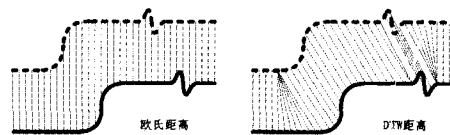


图 1 欧氏距离和 DTW 距离

目前支持时间弯曲的时间序列相似性度量方面的研究主要集中在以下两大方面:

- (1)对 DTW 的改进。主要针对 DTW 的时间复杂度高以及不支持索引这两大问题进行改进。针对 DTW 的高时间复杂度,寻找快速的算法,提高 DTW 计算效率,主要方法有(a)提前终止技术^[7]:累计弯曲代价超过一定的门限,则停止搜索,认为两个序列不相似;(b)采用全局约束设置搜索范围^[8]:

到稿日期:2011-09-18 返修日期:2012-01-19 本文受国家自然科学基金(61001106),国家重点基础研究发展计划(“973”项目)(2009CB320400)资助。

陈 乾(1980-),男,博士生,讲师,主要研究方向为时间序列数据挖掘,E-mail:qqianchen@126.com;胡谷雨(1963-),男,博士,教授,博士生导师,主要研究方向为网络管理技术。

将搜索路径控制在弯曲窗口内部。针对 DTW 不支持索引, 寻找运算量小且满足三角不等式的 DTW 边界函数, 利用下边界函数构造索引, 再在剩余部分 DTW 精确查找^[9,12]。

(2) 寻找其他的弹性度量方式: 主要有 ERP(Edit Distance with Real Penalty)、LCSS(Longest Common SubSequence)、TWED(Time Warp Edit Distance)等。ERP^[6]结合了编辑距离和 L_p 规范构造; LCSS^[3-5]采用两个序列中相同最长子串的长度比例作为相似性度量; TWED^[13]基于编辑距离, 在构造代价函数时考虑了时间偏离的影响, 其性能从目前的实验结果来看是最好的。

绝大多数文献将 DTW 的值作为相似性度量的标准, 时间上的偏离仅采用全局约束设置允许偏离范围。近期也开始有文献关注时间偏离对相似度的影响, 如 TWED 中设置了的时间惩罚系数, 即偏离越远, 代价越高。本文认为, 弯曲路径对相似度是有影响的, 所以在采用 DTW 的相似性度量中, 引入对弯曲路径的考量, 应该能带来更好的结果。

本文第 1 节简要介绍 DTW 的相关背景知识; 第 2 节引入时间距离的概念, 提出一种新的 DTW 最佳弯曲窗口学习方法; 第 3 节用一个分类实验对比采用新的学习方法后的 ODTW(Optimal DTW with best warping window)和其他相似性度量的分类效果, 最后是总结及下一步工作的方向。

1 相关背景

1.1 DTW 和弯曲路径^[9]

设时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle, Y = \langle y_1, y_2, \dots, y_m \rangle$, 长度分别为 n 和 m , 距离矩阵 D 的元素 d_{ij} 表示 x_i 和 y_j 之间的距离 $d(x_i, y_j)$ 。如图 2 所示, 从该距离矩阵中寻找一条弯曲路径 $W = \langle w_1, w_2, \dots, w_K \rangle, \max(n, m) \leq K \leq n + m - 1$, 元素 $w_k = (i, j)$ 表示序列 X 的第 i 个点和序列 Y 中第 j 个点匹配。并满足:

- W 起于矩阵左下角, 止于右上角, 即 $w_1 = (1, 1), w_K = (n, m)$
- W 上任意相邻两个元素在矩阵 D 中也相邻, 且向前发展。即 $w_k = (a_k, b_k), w_{k+1} = (a_{k+1}, b_{k+1})$ 满足 $0 \leq a_{k+1} - a_k \leq 1, 0 \leq b_{k+1} - b_k \leq 1$ 。

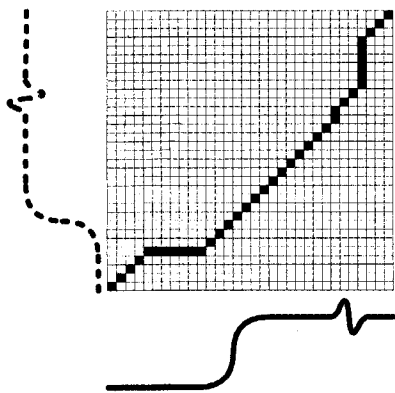


图 2 时间序列的 DTW 弯曲路径

DTW 路径是所有可能的弯曲路径中累计距离最小的路径, DTW 值即为累计距离:

$$DTW(X, Y) = \min_W \left\{ \sum_{k=1}^K d_k, W = \langle w_1, w_2, \dots, w_K \rangle \right\}$$

DTW 距离可通过动态规划方式求得:

$$DTW(X, Y) = f(n, m), f(0, 0) = 0, f(1, 0) = f(0, 1) = \infty$$

$$f(i, j) = d(x_i, y_j) + \min\{f(i-1, j), f(i, j-1), f(i-1, j-1)\} \quad (i=1, \dots, n; j=1, \dots, m)$$

计算 DTW 需要遍历整个 f 矩阵, 复杂度是 $O(n * m)$ 。

1.2 DTW 的全局约束

当弯曲路径偏离对角线太远时, 形成病态的匹配, 此时即使 DTW 值很小, 两个序列的差别也可能很大。如图 3 所示的两个时间序列, 虽然 DTW 距离为 0, 但两者可能并不属于同一类。

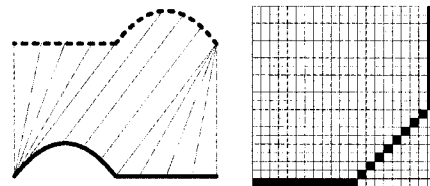
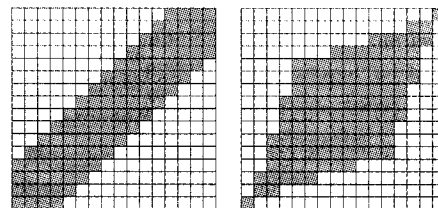


图 3 DTW 中的病态匹配

为了防止这种病态的匹配, DTW 的原始文献就已经提出采用全局约束(Global Constraint)来限制动态匹配时允许的偏离范围^[14], 距离矩阵中弯曲路径可达到的范围称之为弯曲窗口(Warping Window)。文献[15]中用实例说明了更大的弯曲窗口并不一定能带来更准确的分类精度。图 4 给出了 DTW 中最常用的两个全局约束。采用弯曲窗口的 DTW 实现简单, 可减小一部分运算量, 轻微地提高 DTW 运算的效率。更重要的是, 选择合适的弯曲窗口能提高分类精度, 采用最佳弯曲窗口的 DTW 的分类效果比普通 DTW 的好^[16]。



(a) Sakoe-Chiba band (b) Itakura Parallelogram

图 4 最常用的两个全局约束

采用了 S-C 带的 DTW, 又称为具有统一弯曲窗口的 DTW(DTW with Uniform Warping Window), 是最常见的 DTW 实用形式, 和原始 DTW(Full DTW)相比, 它可以利用 LB_Keogh 下界^[9]进行精确索引及快速的搜索算法。很多文献中都将具有最佳弯曲窗口的 DTW 和其他的相似性度量进行有效性对比^[13,15,16], 下面引用文献[13]中的叫法 ODTW(Optimized DTW with best warping window)。

2 最佳弯曲窗口学习方法

2.1 传统的最佳弯曲窗口学习方法

如果 ODTW 的最佳弯曲窗口大小为 r , 传统的学习方法是从 0 开始遍历所有 r 的可能取值, 采用留一验证(Leave-one-out)的方式对训练集进行测试, 即交替选择训练集中一个序列作为测试序列, 剩余的序列作为训练序列, 采用最近邻法则对其进行分类, 从而得到一个分类准确率。选择获得最优分类准确率的 r 值, 如果最优分类准确率对应于多个 r 值, 则选择其中最小的作为训练结果^[15]。

该方法充分利用了训练集中序列间 DTW 的距离信息,

仅就 DTW 距离而言, r 值已是最佳训练结果, 但如果能再充分利用 DTW 弯曲路径提供的信息, r 值的学习还应该有提升的空间。

2.2 序列的时间距离

为了能充分利用 DTW 弯曲路径提供的信息, 我们引入时间距离的概念。时间距离的定义基于 1.1 节中 DTW 弯曲路径。

定义 1 弯曲路径中包含 X 序列第 i 样点的路径点构成的路径子集, 称为 X 的第 i 路径段, 记为

$$W_X^i = \langle w_k = (a_k, b_k) \mid a_k = i, w_k \in W \rangle$$

根据定义, 弯曲路径对于 X 共有 n 个路径段。

定义 2 X 的第 i 路径段中, 对应的 Y 序列编号和 i 的最大距离称为 X 的第 i 路径段最大时间偏离, 记为

$$MaxTimeDev_X(i) = \max\{|b_k - i| \mid W_X^i\}$$

定义 3 序列 X 和 Y 构成的 DTW 弯曲路径中, X 的所有路径段最大时间偏离之和为其时间距离, 记为

$$TimeDist(W) = \sum_{i=1}^n MaxTimeDev_X(i)$$

时间距离的物理含义是 DTW 弯曲路径和斜对角路径 $W_{best} = \langle w_k = (a_k, b_k) \mid a_k = b_k \rangle$ 所夹的面积, 可以用来表示两个序列在时间轴上的偏离程度。如图 5 所示, 对于两个等长的时间序列, 最佳弯曲路径应该是 $a_k = b_k$ 构成的对角线(图中灰色所示), 此时两个序列在时间轴上一一对应, 偏差为 0, 而实际 DTW 弯曲路径(图中黑色所示)和对角线的偏离程度可以用它们所夹的面积来表示。

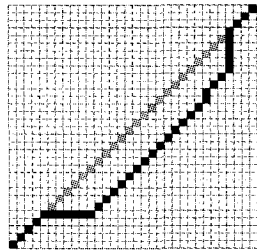


图 5 DTW 弯曲路径和最佳弯曲路径

该时间距离也同样适用于不等长的两个序列的时间距离度量。由定义可知, 不等长的序列的时间距离度量必大于 0, 且随着长度差距扩大, 时间距离也变大, 这和我们日常的经验相符。

DTW 的弯曲路径其实是 DTW 运算的附属产物, 在计算 DTW 时, 如果记录每个 f 值的指向, 构成指向矩阵 g , 从 $g(n, m)$ 开始倒推到 $g(1, 1)$, 即得到 DTW 弯曲路径。再根据时间距离的定义, 计算得到的时间距离。

DTW 距离表示幅值上的偏离程度, 时间距离表示时间上的偏离程度。为了验证时间距离对分类精度的影响, 定义时值组合距离。将这两个距离进行归一化处理, 一种简单的方法就是除以各自的最大值, 然后再加入比例系数 a, b 来调整两者的权重。

定义 4 处理后得到的距离组 $(a \cdot dv_norm, b \cdot dt_norm)$ 可看成是二维空间中的一个点, 该点到原点的距离为序列 X 和 Y 的时值距离 (Time and Value Distance), 记为 $TVD(X, Y, a, b)$:

$$TVD(X, Y, a, b) = (a^2 \cdot dv_norm^2 + b^2 \cdot dt_norm^2)^{1/2}$$

2.3 双距离最佳弯曲窗口学习方法

下面以一个实际的数据集为例, 介绍基于 DTW 距离和时间距离的最佳弯曲窗口学习方法。

首先对 r 值从小到大, 计算训练集中两两间的 DTW 距离和时间距离, 分别利用 DTW 距离和时间距离进行留一测试。两个距离的测试结果分别代表了幅值距离和时间距离对分类精度的影响。如表 1 所列, 此处用分类错误个数来替代分类错误率, $Ev(r)$ 和 $Et(r)$ 分别表示采用窗口为 r 时的 DTW 距离和时间距离对训练集进行留一测试得到的分类错误个数。选择弯曲窗口时, 应综合考虑两者对分类精度的影响。

表 1 不同弯曲窗口下分类错误个数 (ECG200 数据集)

r	Ev	Et	Esum
0	14	50	64
1	16	36	52
2	16	25	41
3	16	19	35
4	16	15	31
5	16	18	34
6	17	16	33
...

鉴于最终对测试集分类时采用的是 DTW 距离, 弯曲窗口 r 本应该从获得最小分类错误率的 r 值中选取, 但考虑到时间距离的影响, 以及训练集可能和测试集存在微小偏差, r 的候选区选择为接近最小分类错误率对应的 r 值, 即满足

$$Ev(r) \leq \min(Ev) + \Delta$$

Δ 的选择根据具体的数据集来确定, 允许的误差率可小比例 r_t 浮动, 如 $r_t = 1\%$, 则 $\Delta = \text{floor}(\text{train_num} \times r_t)$, 其中 train_num 表示训练集的大小, $\text{floor}(\ast)$ 是取整函数。另外由于 DTW 对不同数据集分类精度不一样, 因此还需考虑实际的分类效果, 允许的误差率可按最差分类错误率的一个比例 r_e 浮动, 如 $r_e = 10\%$ 。综合考虑以上因素, Δ 取值为

$$\Delta = \text{floor}(\max(\text{train_num} \times r_t, \max(Ev) \times r_e))$$

如图 6 所示, 红色标记的为候选点。 $r=0$ 时, DTW 距离达到最小的分类错误率, 此时可能时间距离的分类错误率很高, 说明由于受弯曲窗口的限制, 弯曲路径不够准确, 时间距离影响到分类精度, 应当适当放宽弯曲窗口 r , 来获得更准确的弯曲路径。可能在 r 增大的同时, DTW 距离的分类错误率会提高, 所以应该综合考虑, 选取使两者平均分类错误率最小时的 r 值, 如此例中 r 的取值应为 4。

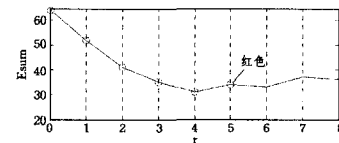


图 6 Esum 曲线

最佳窗口 r 的学习方法详细步骤如下:

- 1) 计算不同 r 值对应的留一验证分类错误数。
- 2) 计算 Δ , 设定 r 的候选区, 如果候选区呈多段, 选择包含最小分类错误率的那一段。
- 3) 考察候选区内 Ev, Et 之和 $Esum$, 选择 $Esum$ 首个极小值对应的 r 值。
- 4) 如果候选区内首个极小值对应多个 r 值, 选择中间的 r 值。

5) 如果不存在极小值(E_{sum} 单调减), 选择当前最小值对应的首个 r 值。

其中, 步骤 4) 和 5) 列出了两种特殊情况下的选择方法。

当前的研究已经指明^[16]: 提出性能接近 DTW, 但运算量更小的度量没有太大的意义, 因为采用的 LB_Keogh 下边界^[9]以及全局约束的 DTW 快速算法, 使其计算复杂度趋近线性, 而且该文献的实验也说明了, 当前绝大多数度量的准确性方面的性能和 DTW 差不多。近期文献^[13]提出的 TWED 从实验数据上看, 性能要好于 DTW, 主要是因为其弯曲路径搜索时引入了时间惩罚。

由于弯曲路径是计算 DTW 时的附属产物, 因此采用基于时间距离的最佳弯曲窗口学习方法的 ODTW 可以在几乎没有增加运算量的情况下, 提高 DTW 的分类准确度, 同时由于其本质还是 ODTW, 因此可以沿用 DTW 的研究成果, 如利用 LB_Keogh 下边界函数进行精确的索引和快速匹配算法等。

3 实验

为验证本文所述最佳弯曲窗口学习方法的有效性, 设计了一个分类实验。下面把用传统方法学习最佳弯曲窗口的 ODTW 称为 TODTW (Traditional ODTW), 采用本文所述新方法的称为 NODTW (New ODTW)。

实验采用文献^[13, 16]使用的 UCR 数据集^[17], 分别采用 ED、DTW、ERP、LCSS、TWED、TODTW 和 NODTW 作为距

离度量; 采用文献^[13, 16]中的方法, 即最近邻法则进行分类, 比较分类错误概率。其中 DTW、ODTW 采用 L2 规范, 本文中的 NODTW 计算采用 L1 规范, 因为在文献^[11]的附加部分, Daniel Lemire 用实验说明了 DTW 用于分类时间序列时, 距离 L_p 规范中 p 的取值为 1 是一个比较好的选择。利用训练集学习最佳弯曲窗口 r 时, $r_t=1\%$, $r_e=10\%$ 。

时值组合距离采用的弯曲窗口和 NODTW 一致, 比例系数 (a, b) 分别取值 $0\sim 1$, 以 0.1 递进, $\{(a=1; b=0, 0.1, 0.2, \dots, 0.9), (a=1, b=1), (a=0.9, 0.8, \dots, 0; b=1)\}$ 共计 21 种组合的时值组合距离, 训练时选择最小分类错误率对应的比例系数。

实验的详细结果如表 2 所列。图 7(a) 所示为 L1 规范下 ODTW 和 L2 规范下 ODTW 的对比结果。图 7 中的对角线表示两者分类错误率相同, 如果某数据集的分类错误率结果落在该线上方, 表示 X 轴对应的度量方式错误率小于 Y 轴对应的度量方式, 所以落在该线下方的点越少, 表示 X 轴对应的度量方式越优越。从图 7(a) 可以看出, L1 规范下的 ODTW 稍优于 L2 规范下的 ODTW, 这和文献^[11]的结论吻合。图 7(b) 为 NODTW 和时值距离的对比结果, 可以看到, 采用了比例系数后的时值距离的分类精度和未采用比例系数的 NODTW 基本一致。从数值上看, 采用了时值距离后, 分类精度有变差的, 如 Gun Point 数据集, 也有变好的, 如 Swedish Leaf 数据集, 而大部分保持不变, 总的来看, 时间距离的引入未能改善分类精度。

表 2 各种相似性度量在 1NN 分类器下的错误率

数据集	类别数 训练集数 测试集数 长度	ED	LCSS	ERP	TWED	DTW	TODTW L2 (r)	TODTW L1 (r)	NODTW L1 (r)	TVD (a, b)
Synthetic Control	6 300 300 60	0.12	0.047	0.036	0.023	0.007	0.017 (6)	0.023 (3)	0.013 (24)	0.013 (1, 0)
Gun-Point	2 50 150 150	0.087	0.013	0.04	0.013	0.093	0.087 (0)	0.087 (0)	0.04 (7)	0.06 (1, 0.2)
CBF	3 30 900 128	0.148	0.009	0.003	0.008	0.003	0.004 (11)	0.008 (10)	0 (21)	0 (1, 0)
Face (all)	14 560 1690 131	0.286	0.201	0.202	0.189	0.192	0.192 (3)	0.215 (6)	0.211 (5)	0.211 (1, 0.2)
OSU Leaf	6 200 242 427	0.483	0.202	0.397	0.248	0.409	0.384 (7)	0.417 (13)	0.405 (16)	0.405 (1, 0)
Swedish Leaf	15 500 625 128	0.213	0.117	0.12	0.102	0.210	0.157 (2)	0.138 (2)	0.147 (4)	0.118 (1, 0.1)
50Words	50 450 455 270	0.369	0.213	0.281	0.187	0.310	0.242 (6)	0.218 (23)	0.209 (18)	0.220 (1, 0.1)
Trace	4 100 100 275	0.24	0.02	0.17	0.05	0.0	0.01 (3)	0.01 (12)	0.01 (18)	0.01 (1, 0)
Two Patterns	4 1000 4000 128	0.09	0.0	0.0	0.001	0.0	0.0015 (4)	0.0013 (5)	0 (7)	0 (1, 0)
Wafer	2 1000 6174 152	0.005	0.0	0.009	0.004	0.020	0.005 (1)	0.0045 (5)	0.0034 (1)	0.0034 (1, 0)
Face (four)	4 24 88 350	0.216	0.068	0.102	0.034	0.170	0.114 (2)	0.102 (3)	0.091 (8)	0.057 (1, 0.3)
Lightning-2	2 60 61 637	0.246	0.18	0.148	0.213	0.131	0.131 (6)	0.115 (8)	0.148 (6)	0.131 (1, 0.4)
Lightning-7	7 70 73 319	0.425	0.452	0.301	0.247	0.274	0.288 (5)	0.274 (3)	0.219 (9)	0.274 (1, 0.3)
ECG200	2 100 100 96	0.12	0.10	0.13	0.10	0.23	0.12 (0)	0.12 (0)	0.1 (4)	0.12 (1, 0.8)
Adiac	37 390 391 176	0.389	0.425	0.378	0.376	0.396	0.391 (3)	0.389 (0)	0.414 (2)	0.407 (1, 0.1)
Yoga	2 300 3000 426	0.170	0.137	0.147	0.130	0.164	0.155 (2)	0.149 (11)	0.149 (11)	0.149 (1, 0)
Fish	7 175 175 463	0.217	0.091	0.120	0.051	0.167	0.160 (4)	0.149 (8)	0.16 (9)	0.16 (1, 0)
Beef	5 30 30 470	0.467	0.533	0.5	0.533	0.5	0.467 (0)	0.467 (0)	0.467 (0)	0.467 (1, 0)
Coffee	2 28 28 286	0.25	0.214	0.25	0.214	0.179	0.179 (3)	0.25 (3)	0.214 (7)	0.214 (1, 0)
Olive Oil	4 30 30 570	0.133	0.8	0.167	0.167	0.133	0.167 (1)	0.133 (0)	0.133 (0)	0.133 (1, 0)
平均分类错误率		0.234	0.191	0.175	0.145	0.179	0.164	0.164	0.157	0.158

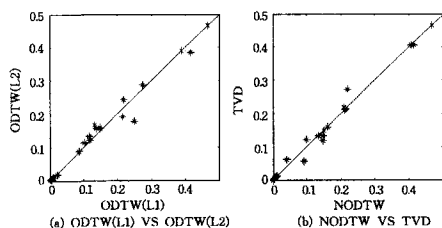


图 7 对比结果

图 8 所示分别是各种度量方式和 NODTW 的对比结果。

从图 8 可以看出, NODTW 明显优于 TODTW、ED、DTW、ERP, 与 LCSS 结果相当, 稍差于 TWED。实际上, 如果从数值上看, NODTW 和 LCSS 相比, 20 个数据集有 9 个优于 LCSS, 8 个劣于 LCSS, 3 个相等, 而且 NODTW 的平均分类错误率要远远小于 LCSS, 所以 NODTW 要优于 LCSS; NODTW 和 TWED 相比, 9 个优于 TWED, 9 个劣于 TWED, 2 个相等, 仅仅是平均分类错误率稍大于 TWED, 两者的实际分类效果十分接近。

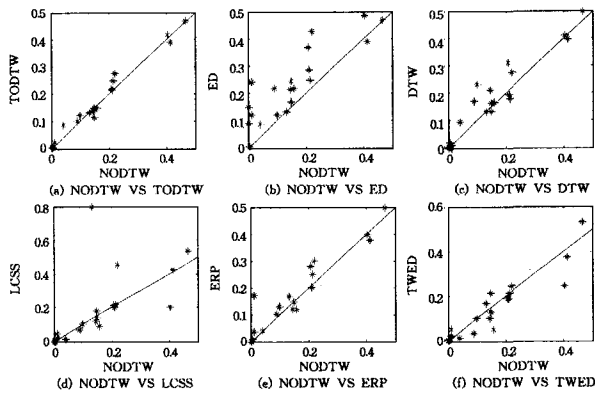


图8 距离度量对比

TWED 中包含了两个参数:硬度系数(时间惩罚系数) γ 和非匹配代价常数 λ ;而 ODTW 只有一个参数:最佳窗口 r 。计算复杂度方面,TWED 计算和 DTW 一样要搜索整个矩阵范围,运算复杂度同为 $O(n * m)$ 。而 ODTW 对搜索范围进行了限制, r 越小,计算复杂度越小。训练阶段运算量 NODTW 要小于 TWED;在分类阶段,NODTW 还可以进一步利用 LB_Keogh 下界实现快速搜索,运算速度远高于 TWED。表 3 列出了实验中部分数据集的 TWED 和 NODTW 训练时间和分类时间,其中 TWED 训练时参考文献 [13], λ 取值 $\{0, 0.25, 0.5, 0.75, 1\}$, γ 取值 $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$,共 30 种组合;NODTW 训练时 r 取值 $0 \sim 30$,分类时采用了 LB_Keogh 下界快速搜索算法^[9]。从实验结果看,NODTW 运算复杂度明显低于 TWED。根据以上分析,从实用性角度来看,改进后的 ODTW 显然更好。

表3 NODTW 和 TWED 运行时间对比

数据集	训练用时(s)		分类用时(s)	
	TWED	NODTW	TWED	NODTW
Gun-Point	3129.8	813.2	636.6	18.6
CBF	784.5	236.5	1669.5	565.1
Face(four)	4368.1	1141.8	1277.2	221.1
ECG200	5049.8	1776.1	340.6	18.4

结束语 本文对时间序列的相似性度量中经典的 DTW 进行了深入的研究,主要贡献有:

- (1)在 DTW 弯曲路径基础上,引入时间距离来描述时间序列间的时间偏离程度;
- (2)用实验证明了采用 DTW 时,时间距离对于改善分类精度效果很有限,但可用来改进 ODTW 最佳弯曲窗口的学习;
- (3)提出了利用时间距离改进 ODTW 最佳弯曲窗口的学习方法,改进后的 DTW 分类精度接近于 TWED。

下一步的研究方向是适于其它不同的全局约束(如 Itakura Parallelogram 和 R-K Band)的最佳弯曲窗口学习方法。

参考文献

[1] Rafiei D, Mendelson A. Efficient retrieval of similar time sequences using DFT [C]//Proceedings of 5th International Conference on Foundations of Data Organizations and Algorithms. Kobe, Japan, 1998;249-257

[2] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series [C]//Proceedings of the AAAI Workshop

on Knowledge Discovery in Databases. Menlo Park, CA; AAAI, 1994;359-370

[3] Das G, Gunopulos D, Mannila H. Finding Similar Time Series [C]//Proceedings of the Conference on Principles of Knowledge Discovery and Data Mining. Springer-Verlag, London, UK, 1997;88-100

[4] Loh W K, Kim S W, Whang K Y. Index interpolation: an approach for subsequence matching supporting normalization transform in time-series databases [C]//Proceedings of the 9th International Conference on Information and Knowledge Management. New York; ACM Press, 2000;314-325

[5] Vlachos M, Hadjieleftheriou M, Gunopulos D, et al. Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures[C]//Proceedings of 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, 2003;216-225

[6] Chen L, Ng R. On the marriage of L_p -norm and edit distance[C]//Proceedings of the 30th International Conference on Very Large Data Bases. Toronto, Canada; VLDB Endowment, 2004;792-801

[7] 陈胜利,李俊奎,刘小东.基于提前终止的加速时间序列弯曲算法[J]. 计算机应用, 2010, 30(4):1068-1071

[8] 尚福华,孙达辰,吕海霞.提高 DTW 运算效率的改进算法[J]. 计算机工程与设计, 2010, 31(15):3518-3520

[9] Keogh E. Exact indexing of dynamic time warping [C]// Proceedings of 28th International Conference on Very Large Databases Conference. Hong Kong; VLDB Endowment, 2002;406-417

[10] Kim S, Park S, Chu W. An index-based approach for similarity search supporting time warping in large sequence databases[C]// Proceedings of the 17th International Conference on Data Engineering. Heidelberg, IEEE Computer Society, 2001;607-614

[11] Lemire D. Faster Retrieval with a Two-Pass Dynamic-Time-Warping Lower Bound[J]. Pattern Recognition, 2009, 42 (9): 2169-2180

[12] Niennattrakul V, Ruengronghirunya P, Ratanamahatana C A. Exact Indexing for Massive Time Series Databases under Time Warping Distance[J]. Data Mining and Knowledge Discovery, 2009, 21(3):509-541

[13] Marteau P F. Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence archive, 2009, 31 (2): 306-318

[14] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE Transactions on Acoustics, Speech and Signal Process, 1978, 26(1):43-49

[15] Ratanamahatana C A, Keogh E. Making Time-series Classification More Accurate Using Learned Constraints[C]//Proceedings of SIAM International Conference on Data Mining. Lake Buena Vista, Florida, 2004;11-12

[16] Ding H, Trajcevski G, Scheuermann P, et al. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures[J]. Proceedings of the 34th International Conference on Very Large Data Bases. Auckland, New Zealand, VLDB Endowment, 2008, 1(2):1542-1552

[17] Keogh E J, Xi E, Wei X, et al. The UCR Time Series Classification/Clustering[EB/OL]. http://www.cs.ucr.edu/~eamonn/time_series_data/, 2006