

周报

刘精昌

2017 年 2 月 10 日

本周工作

- 阅读《A Complexity-Invariant Distance Measure for Time Series》，文章比较简单，阅读的比较粗略，主要思想就是通过通过在距离前面添加系数，解决了序列分类中，将简单形状序列分类到复杂序列的情况。
- 阅读《FastDTW : Toward Accurate Dynamic Time Warping in Linear Time and Space》，文章将 DTW 方法的时间复杂度、空间复杂度从 $O(n^2)$ 降到 $O(n)$ ，文章的成果很具有价值。
- 通过人造数据集和TSDMA数据集的实验，探究 DTW 方法、Euclid 方法分类的精度差异，以及 warp window 大小对 DTW 方法分类精度的影响。人造数据是在一定的间隔上添加随机噪声而产生的，有四类，每类有四个训练集，四个测试集。TSDMA 数据集的类数目，测试集、训练集数目都有给出。分类均采用 1NN 方法，因为计算量特别大，所以实验是在实验室服务器上利用 MATLAB 的并行计算完成的。主要结果如下：

1. 对于人造数据，从 0 到 2 递加间隔，每种间隔下，实验均运行 100 次（噪声是随机的），最后得到 DTW 距离和 Euclid 距离下的平均精度。实验结果曲线如图??。

由实验结果曲线可以看出，DTW 距离下的分类精度明显优于 Euclid 聚类。

2. 对于不同的 TSDMA 数据集，分别应用 DTW 和 Euclid 距离，得到精度表

name	Computers	Trace	FaceFour	WordsSynonyms	Gun_Point	Plane	StrawBerry
lasses	2	4	4	25	2	7	2
training set size	250	100	24	267	50	105	370
test set size	250	100	88	638	150	105	613
sequence length	720	275	350	270	150	144	235
error rate (Euclid)	0.424	0.24	0.21591	0.38245	0.086667	0.038095	0.06199
error rate (DTW)	0.332	0.01	0.15909	0.32445	0.12	0	0.66884

从上面的精度表看以看出，大部分情况下 dtw 距离的精度都要优于 Euclid 距离，个别的数据 dtw 距离的精度反而不如 Euclid 距离。

3. 主要探究 warp window 的变化对 DTW 距离精度的影响。实验数据是间隔为 0.135 的人造数据和部分 TSDMA 数据，方法是将 warp window 从 0 到 1 每次变化 1%，记录精度的变化，结果如图??、??、??、??、??所示。

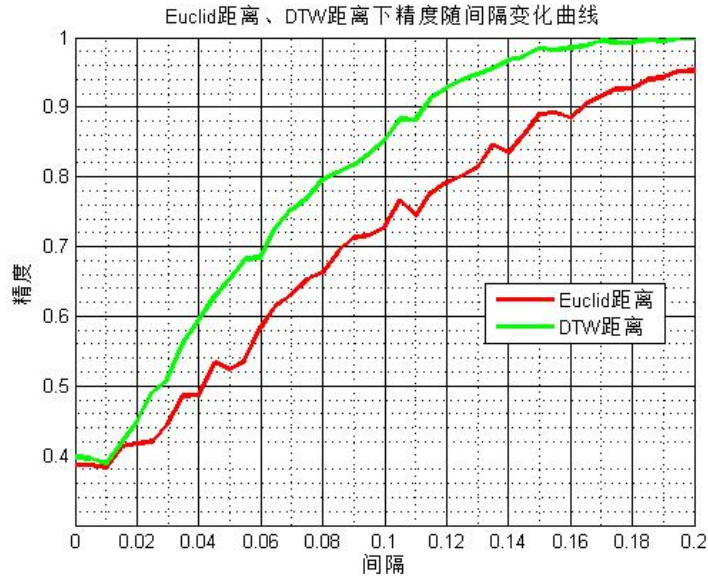


图 1: Euclid、DTW 距离下精度随时间变化

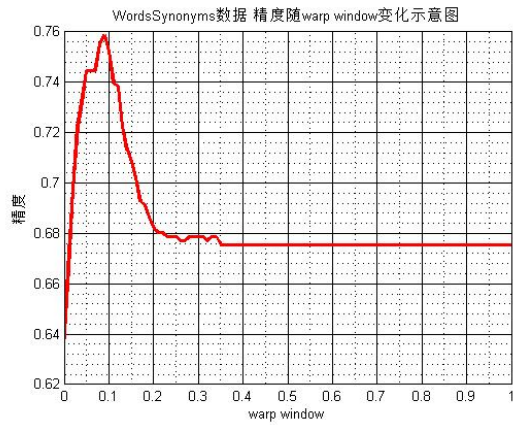


图 2: WordsSunonyms 数据集

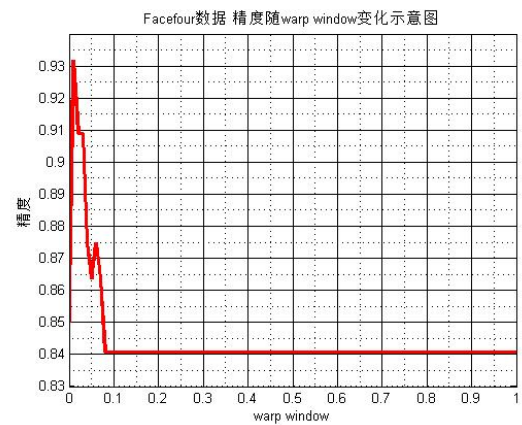


图 3: FaceFoue 数据集

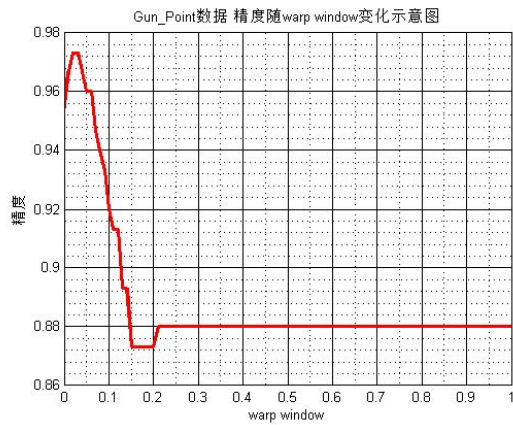


图 4: Gun_Point 数据集

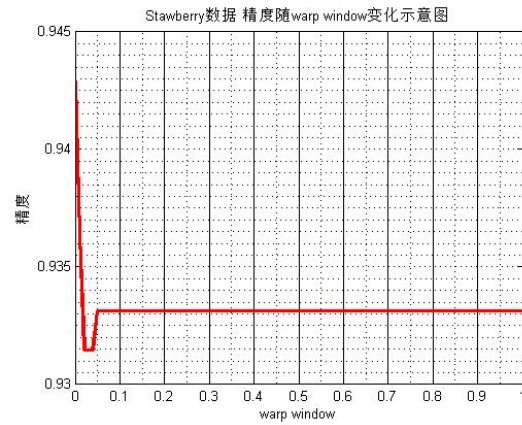


图 5: StrawBerry 数据集

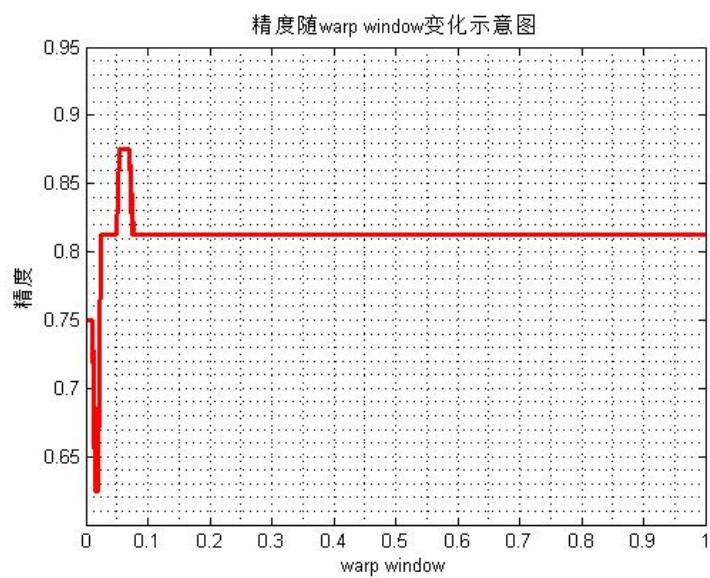


图 6: 人造数据集

从这些图中可以看出，一般而言，一个较小的 warp window（小于 10%）对 DTW 分类精度的提升是有利的。

下周计划

- 主要写毕设论文，灵活地进行实验，参考论文。