CHAPTER IV

NUMERICAL RESULTS

We compare the performance of our distance measures with the Euclidean distance and DTW using the datasets from the UCR time series classification / clustering homepage (Keogh et al., 2006). The experimental setting is done as in the UCR dataset homepage. The 1-NN algorithm was used to perform classification tasks. Each of the datasets is already split as the training set and the testing set. Numbers shown in each column of distance measure are the error rates as classified by the 1-NN algorithm using that distance function as the dissimilarity measure. The results are shown in Table 4.

Every time series in each dataset was already pre-scaled to the same length if needed so that the ℓ^1 distance and the Euclidean distance can be computed between every pair of time series in each dataset.

The distance δ_1 was incompetent because its base distance, the ℓ^∞ metric is usually very sensitive to little variations or noise in time series. Other than δ_1, our proposed subadditive distances did relatively well. The results also showed that the value of π and p had influence on the classification accuracy of δ_4.

Nonetheless, it is interesting to note that the *trace* dataset can be classified with zero error rate by DTW and δ_1. The identical property shared by the two distances is that they are a condensations whose morphs are the stretch operations $\mathcal{S}$ (cf. Equation (3.3) and Equation (3.6)). While the condensations δ_{2-4}, whose morphs are the gap insertions $\mathcal{I}$, gave relatively poor performance. A plausible explanation of the phenomenon is that the stretches are more suitable as morph operations than gap insertions for this particular dataset. It seemed, though, that some dataset have no preference of either the stretches the inserts over the other, the *two patterns* dataset can be classified by 1-NN with zero error rate with every condensations regardless of the morphs involved in each condensation.

It is noticeable that the results of the distance δ_2^2 and the distance δ_3 are the same or almost the same in many datasets. This is due to the fact that if the ℓ^2 norm of every time series in a dataset is the same, then the nearest neighbors of a time series as measured by δ_2^2 will also be the nearest neighbors as measured by δ_3 and vice versa. Those datasets whose time series have small variation among their ℓ^2 norms yielded the same error rates for δ_2^2 and δ_3. For those datasets whose time series have large variation among their ℓ^2 norms the results were different. The time series in the coffee dataset have very large variation of the ℓ^2 norms and the distance δ_3 performed significantly better than other distances including δ_2^2.

Table 4.1: Numbers whose value is minimal in its row are typeset bold. The second column shows the standard deviations of the ℓ^2 norms of the time series samples in the training set of each dataset.

dataset	$\sigma(\|\cdot\|_2)$	ℓ^1	ℓ^2	DTW^1	DTW^2	δ_1	δ_2^1	δ_2^2	δ_3	δ_4^1 $\pi=0.5$	δ_4^1 $\pi=1.0$	δ_4^2 $\pi=0.5$	δ_4^2 $\pi=1.0$
50words	2.512e-08	0.3319	0.3692	0.2835	0.3099	0.4264	0.2813	0.3231	0.3231	0.2352	0.2242	**0.2110**	0.2462
adiac	2.449e-08	0.4015	0.3887	0.4118	0.3964	0.4220	0.3785	**0.3708**	**0.3708**	0.3785	0.3939	0.3862	0.3887
beef	1.037e+00	0.5000	0.4667	0.5000	0.5000	0.4333	0.5000	0.5000	**0.2333**	0.5000	0.5000	0.4667	0.4667
cbf	2.637e-08	0.1111	0.1478	**0.0000**	0.0033	0.0289	0.0022	0.0022	0.0022	0.0100	0.0322	0.0078	0.0156
coffee	4.643e+01	0.2500	0.2500	0.1786	0.1786	0.3214	0.2500	0.2500	**0.0357**	0.2500	0.2500	0.2500	0.2500
ecg200	6.052e-03	0.1100	0.1200	0.2000	0.2300	0.2500	0.1500	0.1800	0.1800	0.1600	**0.1000**	0.1400	0.1300
faceall	2.465e-08	0.2787	0.2864	0.2284	0.1923	0.2379	0.2024	0.2030	0.2036	0.1941	0.1947	**0.1882**	0.1935
facefour	3.670e-08	0.1591	0.2159	0.1591	0.1705	0.4205	0.1023	0.1705	0.1705	0.0568	**0.0455**	0.1364	0.1250
fish	1.068e-01	0.2057	0.2171	0.1429	0.1657	0.3029	**0.1200**	0.1314	0.1314	0.1371	0.1600	0.2000	0.2114
gun point	2.311e-08	0.0467	0.0867	0.1200	0.0933	0.2133	**0.0400**	**0.0400**	**0.0400**	0.0467	0.0467	0.0800	0.0867
lighting2	1.350e-07	0.1803	0.2459	0.1967	0.1311	0.2787	0.1475	**0.1148**	**0.1148**	**0.1148**	**0.1148**	0.1311	**0.1148**
lighting7	1.210e-07	0.2877	0.4247	**0.2329**	0.2740	0.5205	0.3014	0.3288	0.3288	0.2740	0.2603	0.2466	0.2603
oliveoil	3.710e-02	0.1667	**0.1333**	**0.1333**	**0.1333**	0.2000	0.1667	**0.1333**	0.1667	0.1667	0.1667	**0.1333**	**0.1333**
osuleaf	5.997e-03	0.4504	0.4835	0.3678	0.4091	0.4876	0.3967	0.4132	0.4132	**0.3512**	0.3636	**0.3512**	0.3719
swedishleaf	2.513e-06	0.2112	0.2112	0.2096	0.2080	0.2464	0.1200	0.1232	0.1232	**0.1120**	0.1264	0.1184	0.1408
synthetic control	2.614e-08	0.1200	0.1200	0.0133	**0.0067**	0.0467	0.0333	0.0300	0.0300	0.0267	0.0367	0.0300	0.0233
trace	2.585e-08	0.2400	0.2400	0.0100	**0.0000**	**0.0000**	0.1700	0.1600	0.1600	0.1800	0.2000	0.1700	0.1700
two patterns	1.392e-07	0.0387	0.0932	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
wafer	1.274e-07	0.0047	0.0045	0.0161	0.0201	0.0339	0.0088	0.0092	0.0092	0.0049	**0.0028**	0.0049	**0.0028**
yoga	2.620e-08	0.1710	0.1697	0.1613	0.1637	0.1757	0.1470	0.1623	0.1623	**0.1307**	0.1363	0.1487	0.1583