Time Series Classification using Dynamic Time Warping and Time Series Averaging

Kosate Limpongsa 5731012721
Sirinthra Chantharaj 5730635521

1. Introduction

Classifications is one of the most common problem in time series mining. At this moment, Nearest Neighbour algorithm (NN) is the most accurate method for time series classification. It simply compares a test time series to all training data and measure similarity. Then, the test data is classified as the same class as the most similar data.

One of the most popular method for measuring similarity between two time series is Dynamic Time Wrapping (DTW). However, DTW contains many parameters such as weight and number of cells used in calculation. Thus, we aim to find the best parameters of DTW for performing 1-NN.

Nevertheless, practically, we do not perform 1-NN on large dataset since it consumes a lot of time. There is an idea to find a time series which can represent all data in train set which is Nearest Centroid Classifier (NCC). The key idea is that instead of comparing test data to all training data, we reduce the training data to some centroid value of each class then compare training data to that value. Unfortunately, there is no clear definition of centroid value in time series mining. In this work, we intent to find the good definition of centroid value that can produce the same level of accuracy as NN algorithm.

2. Objectives

- To find the best DTW weight of the three neighbour cells on our dataset
- To find the optimal DTW pattern that maximize the 1-NN accuracy on our dataset
- To find the best time series averaging method for different kind of datasets

3. Literature Review

For NCC, there are two main averaging techniques used for defining the centroid which are amplitude average and shape average. Figure 1 compares the different between amplitude averaging (left) and shape averaging (right).

Amplitude averaging can be calculated by arithmetic mean while shape averaging requires more complex method. There are many existing work studied on shape averaging methodology.

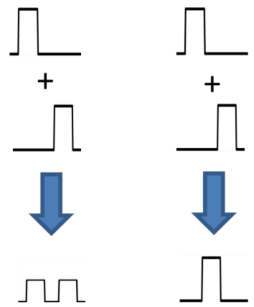

Figure 1. DBA iteratively adjusting the average of two sequences

One of the well-known method is DTW Barycenter Averaging (DBA). DBA is the iterative method for finding the centroid. Each iteration can be executed by computing DTW between each individual sequence and the temporary average sequence to find associations between its coordinates and coordinates of the sequences. Then, the mean is updated according to the associations just computed. The more number of iterations, the more accurate the centroid. Figure 2 illustrates an example of an average sequence computed with DBA.

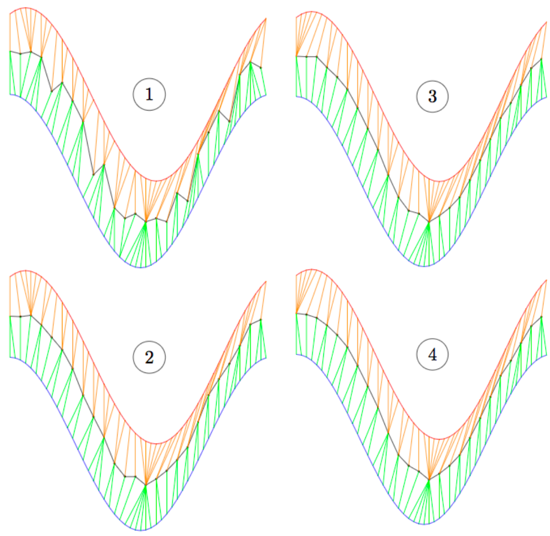

Figure 2. DBA iteratively adjusting the average of two sequences

However, DBA is time consuming; hence, there is a n algorithm which takes advantage of a smoothed formulation of DTW, called soft-DTW, that computes the soft-minimum of all alignment costs. Figure 3 compares the results from Euclidean barycenter, DBA, and Soft-DTW barycenter.

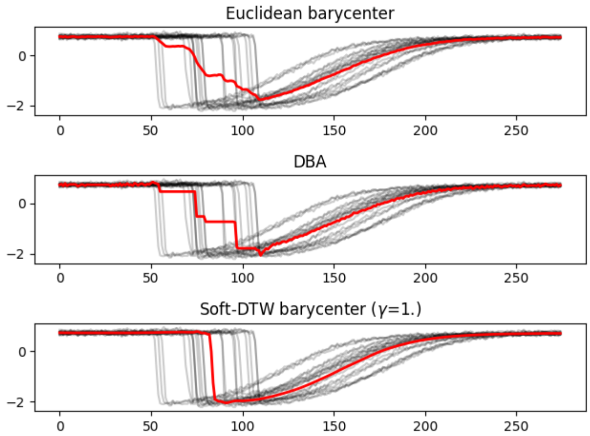

Figure 3. Centroids calculated by Euclidean barycenter, DBA, and Soft-DTW barycenter

Another approach of defining the best centroid is PK-Bayesian method proposed by Poomarin Phloyphisut and Korrawe Karunratanakul, who are the students of Time Series Mining class 2015. The method treats each index of time series data as one dimension and extract mean and standard deviation on each index of each class. Then calculate Bayesian probability for each class.

4. Dataset

In 1-NN experiments, we used 6 datasets which are UCR time series dataset consisting of 50words, CBF, synthetic control, Adiac, Coffee, and OSULeaf.

For NCC, we used 17 datasets to perform the task consisting of 50words, Adiac, Beef, CBF, Coffee, FISH, FaceAll, FaceFour, Gun_Point, Lightning7, OSULeaf, OliveOil, SwedishLeaf, Trace, Two_Patterns, synthetic_control, and yoga dataset.

5. Experiment and Result

Experiment 1-1: Best DTW Weight

In this experiment, we aim to find the best DTW weight of the three neighbor cells. We carefully chose three random integers which are not greater than four. The chosen weights must not be the multiplication of other weights. For example, (2,4,4) is multiplication of (1,2,2). In this case, (2,4,4) was not selected. We had run the experiment with 10 random set of weights. The results are listed in Table 1.

Experiment 1-2: Best DTW Pattern

Instead of adjust the weights in experiment 1-1, we change the structure of neighbouring cells for DTW calculation. We designed five different patterns. All patterns are symmetric weights and each weight value is not over than three. The maximum number of neighboring cells was not exceed eight. The results and implementation details are listed in Table 2.

Experiment 2: Best Centroid

We performed the different methods of averaging time series data including Amplitude Average, DTW Barycenter, Sotf-DTW Barycenter, and our proposed method. The proposed method is modified from PK-Bayesian method. However, instead of comparing the centroid with the test data point by point, we used DTW alignment and find mean and standard deviation of the each point on DTW Bary Center. The goodness of centroid is evaluated by classification accuracy which can be achieved by comparing test data with centroid of each class. Then, the test data is classify as the most similar centroid class. The results, as represented in Table 3, show that the method which produce the best centroid is Soft-DTW.

6. Discussion and Conclusion

According to the experiments, there are different set of weights that performing the best accuracy depends on specific dataset as shown in Table 1. The best weight which is an asymmetric weight (1,1,2) produced 79.51% of accuracy.

In DTW pattern experiment, there is another method that perform better than 3-neighboring cells. The result end up with 80.59% of average accuracy which is slightly better than the 3-neighbouring method. The pattern which achieve the highest accuracy also contains 3-neighbouring cells like the previous one but the position of each neighbor is changed. Dynamic programming equation of the best pattern is showed as follows:

$$D(i,j) = min \begin{cases} D(i-2,\ j-1) + 2 \times d(i,j) \\ D(i-1,\ j-2) + 2 \times d(i,j) \\ D(i-1,\ j-1) + d(i,j) \end{cases}$$

For NCC part to find the best centroid, all the shape averaging methods gave the better performance than the amplitude averaging method. The best shape averaging method is Sotf-DTW Barycenter Averaging which achieved 71.38% of mean accuracy. However, DBA also produced high accuracy which is 70.40%.

After exploring the datasets, we found that the DBA method performed better on complex data which are hard to classify. The 50Words and FaceAll dataset are good examples of complex data. Figure 4 and 5 demonstrates the data of OliveOil and Beef dataset respectively.

The Soft-DTW Barycenter Averaging method gave better accuracy on normal data which can be classify easily by humans. Figure 6 and 7 demonstrates the data of 50Words and FaceAll dataset which are the examples of normal data.

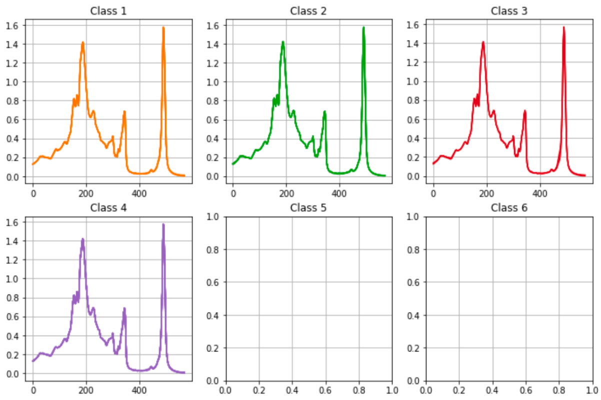

Figure 4. Data of all classes from OliveOil dataset

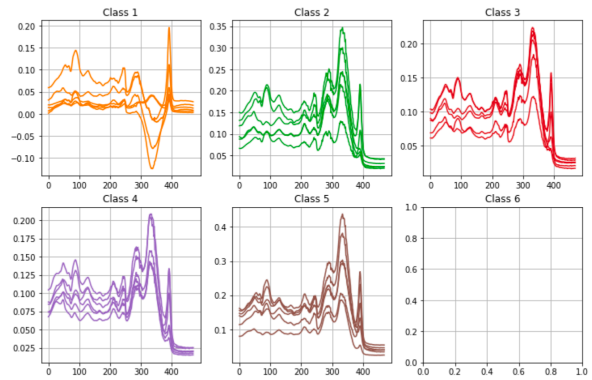

Figure 5. Data of all classes from Beef dataset

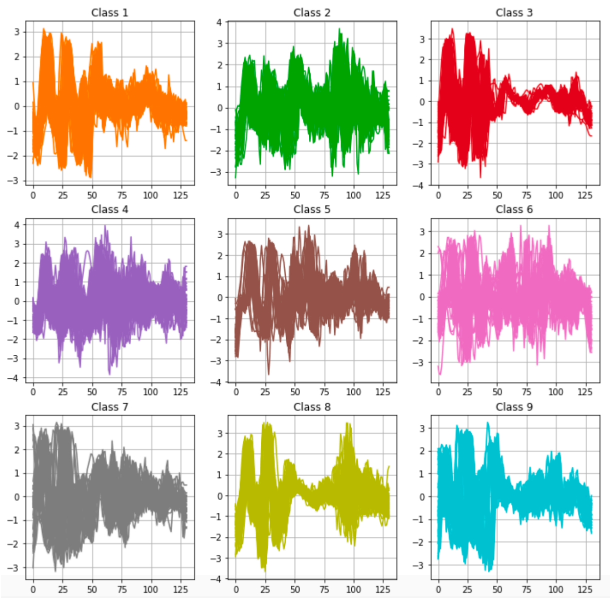

Figure 6. Data of the first nine classes of FaceAll dataset

For the proposed Bayesian-DTW method, it worked only on Two_Patterns dataset which might be resulted from the size of train data. The Two_Patterns dataset contains 1,000 time series in total while the other datasets consists of less than 561 sequences. Table 4 shows the train size of each dataset. When performing Bayesian probability, all data have to be normal distribution which required n of each class to be grater than 30.

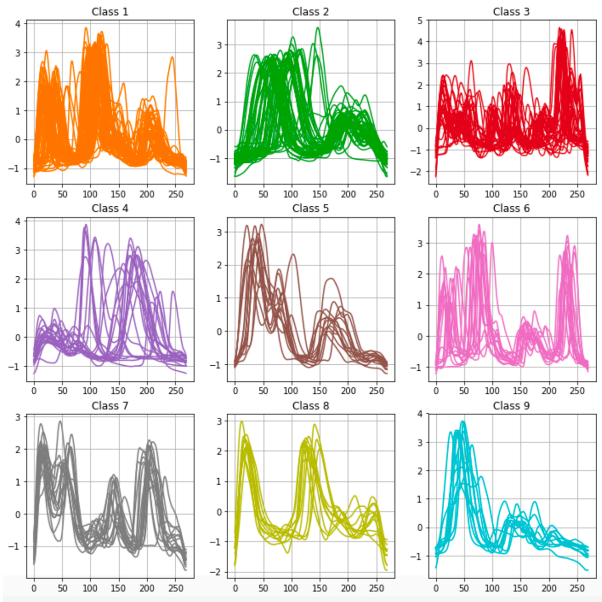

Figure 7. Data of all classes from 50Words dataset

Table 4. Train dataset size

Dataset	Train size
50words	450
Adiac	390
Beef	30
CBF	30
Coffee	**28**
FISH	175
FaceAll	560
FaceFour	24
Gun_Point	50
Lightning7	60
OSULeaf	300
OliveOil	30
SwedishLeaf	500
Trace	100
Two_Patterns	1000
synthetic_control	300
yoga	300
Average	249

Table 1. DTW weight experiment results

w 1	w 2	w 3	50words	CBF	synthetic control	Adiac	Coffee	OSULeaf	Average
1	**1**	**2**	**0.7187**	**1.0000**	**0.9900**	0.5959	**0.8214**	**0.6446**	**0.7951**
1	**1**	**1**	0.7165	**1.0000**	0.9867	0.5882	**0.8214**	0.6322	0.7908
2	**1**	**1**	0.6989	0.9978	0.9633	**0.6061**	0.7857	0.6116	0.7772
1	**3**	**2**	**0.7187**	0.9967	0.9833	0.5217	0.7857	0.6405	0.7744
1	**2**	**1**	0.7099	0.9700	0.9467	0.5243	0.7857	0.6281	0.7608
0	**1**	**1**	0.5912	0.9978	0.9800	0.4859	**0.8214**	0.5413	0.7363
3	**2**	**1**	0.6462	0.9633	0.9133	0.5780	0.7500	0.5620	0.7355
1	**4**	**1**	0.7077	0.8578	0.8333	0.4987	0.7143	0.6322	0.7073
2	**3**	**1**	0.6659	0.8589	0.8433	0.5269	0.7143	0.5661	0.6959
1	**0**	**1**	0.0615	0.3856	0.3633	0.1125	0.3571	0.2851	0.2609

Table 2. DTW pattern experiment results

DP-Equation	50words	CBF	synthetic control	Adiac	Coffee	OSULeaf	Average
min { D(i-1, j-1) + 2*d(i,j) D(i-1, j) + d(i,j) D(i, j-1) + d(i,j) }	0.7187	**1.0000**	**0.9900**	0.5959	**0.8214**	0.6446	0.7951
min { D(i, j-2) + 2*d(i,j) D(i-2, j) + 2*d(i,j) D(i-1, j-1) + d(i,j) }	0.6967	0.9589	0.8567	0.5780	**0.8214**	0.6281	0.7566
min { D(i, j-2) + 3*d(i,j) D(i-2, j) + 3*d(i,j) D(i, j-1) + 2*d(i,j) D(i-1, j) + 2*d(i,j) D(i-1,j-1) + d(i,j) }	0.7319	0.9844	0.9133	0.6061	**0.8214**	0.6364	0.7823
min { D(i-2, j-1) + 2*d(i,j) D(i-1, j-2) + 2*d(i,j) D(i-1, j-1) + d(i,j) }	**0.8022**	0.9978	0.9833	0.5473	0.7857	**0.7190**	**0.8059**
min { D(i-2, j-1) + 2*d(i-1,j) + d(i,j) D(i-1, j-2) + 2*d(i,j-1) + d(i,j) D(i-1, j-1) + d(i,j) }	0.7956	0.9944	0.9700	**0.6113**	0.7500	0.6860	0.8012
min { D(i, j-2) + 2*d(i,j-1) + d(i,j) D(i-2, j) + 2*d(i-1,j) + d(i,j) D(i-1, j-1) + d(i,j) }	0.7055	0.9911	0.9267	0.5703	0.7857	0.6240	0.7672

Table 3. Accuracy of the classification result performed by amplitude average, DBA, Soft-DTW, and Bayesian-DTW on different datasets

Dataset	Amplitude Average	DBA	Soft-DTW	Bayesian-DTW
50words	0.4571	0.6242	**0.7626**	0.3341
Adiac	0.4834	0.4629	**0.6726**	0.4936
Beef	0.4667	**0.5333**	0.4667	0.5000
CBF	0.7144	**0.9656**	0.9511	0.9500
Coffee	**0.3214**	0.2500	0.2857	0.2857
FISH	0.5771	0.6571	**0.8400**	0.4914
FaceAll	0.4479	0.7964	**0.8438**	0.7107
FaceFour	0.8068	0.8523	**0.8750**	0.6705
Gun_Point	0.6267	**0.7000**	0.6400	0.6933
Lightning7	0.4658	**0.6575**	0.6575	0.4932
OSULeaf	0.3636	0.4380	**0.5165**	0.3347
OliveOil	0.5333	**0.8667**	0.6000	0.7333
SwedishLeaf	0.6288	0.6816	**0.7824**	0.5488
Trace	0.4800	**0.9700**	0.7400	0.8900
Two_Patterns	0.3018	0.9750	0.9695	**0.9760**
synthetic_control	0.7367	**0.9800**	0.9567	0.9567
yoga	0.5197	0.5570	**0.5743**	0.5620
Average	0.5254	0.7040	**0.7138**	0.6249

7. References

[1] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen and Gustavo Batista (2015). The UCR Time Series Classification Archive. URL www.cs.ucr.edu/~eamonn/time_series_data/

[2] François Petitjean, Germain Forestier, Geoffrey I. Webb, Ann E. Nicholson, Yanping Chen and Eamonn Keogh. Dynamic Time Warping Averaging of Time Series allows Faster and more Accurate Classification

[3] Petitjean, F., Ketterlin, A., & Gancarski, P. (n.d.). A global averaging method for dynamic time warping, with applications to clustering.

[3] Cuturi, M., & Blondel, M. (n.d.). Soft-DTW: A Differentiable Loss Function for Time-Series.