Using derivatives in longest common subsequence distance to time series classification

Tomasz Górecki

Faculty of Mathematics and Computer Science
Adam Mickiewicz University

GPSDAA, 2013

Introduction

Time-series classification has been studied extensively by machine learning and data mining communities, resulting in a variety of different approaches ranging from neural and Bayesian networks to genetic algorithms and support vector machines. We can find many distance measures for similarity of time series data (a very good overview can be found in Ding et al. (2008)).

Introduction

The simplicity and efficiency of Euclidean distance makes it the most popular distance measure. It requires that both input sequences be of the same length, and it is sensitive to distortions. Such a problem can be handled by elastic distance measures such as Dynamic Time Warping (DTW) and Longest Common SubSequence (LCSS). DTW searches for the best alignment between two time series, attempting to minimize the distance between them. LCSS finds the length of the longest matching subsequence. Of the three measures, LCSS is the least sensitive to noise because it includes a threshold to define a "match" (Vlachos et al. (2002)).

Introduction

It seems that in the classification domain there could be objects for which function value comparison is not sufficient. There could be cases where assignment to one of the classes depends on the general shape of objects (signals, functions) rather than on strict function value comparison. Especially for time series it seems that some variability in the "time" domain could have a great influence on the classification process.

Introduction

In mathematics, an object associated with a function that responds to its variability in "time" is the derivative of the function. The function's derivative determines areas where the function is constant, increases or decreases, and the intensity of the changes. The derivative determines the general shape of the function rather than the value of the function at an actual point. While the first derivative gave some information about the shape of the function (increasing or decreasing), the second derivative adds additional information as to where the function is convex or concave.

It seems that especially in the case of time series such an approach to classification can be very effective. We cannot expect that it is sufficient to compare time series only as their derivatives. It seems that the best approach is to create a method which considers both function values of time series and values of the derivative (or derivatives) of the function (shape comparison). The intensity of the influence of these approaches should be parameterized. Then we can expect that for different time series the method will select the appropriate intensity of these kinds of comparisons and give the best classification results.

The longest common subsequence similarity measure is a variation of edit distance used. The basic idea is to match two sequences by allowing them to stretch, without rearranging the sequence of the elements but allowing some elements to be alertunmatched or left out (e.g. outliers), where as in Euclidean and DTW, all elements from both sequences must be used, even the outliers. The overall idea is to count the number of couple of points from two sequences that matches. One point can never be associated twice to a point of the other sequence, so that the maximum number of associations is the minimum length of the two sequences. LCSS measure has two parameters, δ and ε. The constant δ, which is usually set to the percent of the sequence length, controls the window size in order to match a given point from on sequence to a point in another sequence. The constant $0 < \varepsilon < 1$ is the matching threshold.

Longest Common Subsequence (LCSS)

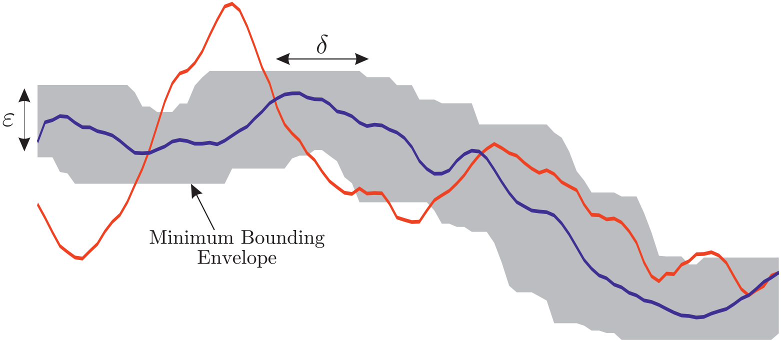

Figure : Matching within δ time and ε in space. Everything that is outside the bounding envelope can never be matched.

Let LCSS is a longest common subsequence distance measure for two time series x and y. Distance measure which considers both function values of time series and values of the first derivative is defined by:

$$\mathrm{DD}_{\mathrm{LCSS}}(x,y) := a\,\mathrm{LCSS}(x,y) + b\,\mathrm{LCSS}(\nabla x, \nabla y),$$

where ∇x and ∇y are first discrete derivatives of x, y, and $a, b \in [0,1]$ are parameters. The discrete derivative of a time series x with length n is defined by:

$$\nabla x(i) = x(i+1) - x(i),\ \ i = 1, 2, \ldots, n-1.$$

LCSS based on derivatives

Distance measure which considers both function values of time series and values of the first and second derivatives is defined by:

$$2\mathrm{DD}_{\mathrm{LCSS}}(x, y) := a\,\mathrm{LCSS}(x, y) + b\,\mathrm{LCSS}(\nabla x, \nabla y) + c\,\mathrm{LCSS}(\nabla^2 x, \nabla^2 y),$$

where $\nabla^2 x$ and $\nabla^2 y$ are second discrete derivatives of x, y, and $a, b, c \in [0, 1]$ are parameters.

We performed experiments on 47 data sets. The data sets originate from the UCR Time Series Classification/Clustering Homepage (Keogh et al. (2011)). For each data set we calculated the classification error rate on a test subset. We found all parameters using the training subset. An appropriate distribution of the training and test sets was proposed by the authors of the repository (each data set is divided into a training and testing subset). We use the leave-one-out cross-validation method to find the best parameter α (the best pair of parameters α, β) in our classifier. If the minimal error rate is the same for more than one value of parameter α we choose the smallest one (the smallest pair – minimizing α first, then β). Finite subsets of parameters α and β are chosen, from 0 to 1 with fixed step 0.01.

The values we used for δ and ε are clearly dependent on the application and the data set. We set δ to 100%. The determination of ε is application dependent. We used a value equal to the smallest standard deviation between the two trajectories that were examined at any time (Vlachos et al. (2002)).

Table : Average relative testing error rates on all data sets.

	$\frac{DD_{LCSS} - LCSS}{LCSS}$	$\frac{2DD_{LCSS} - LCSS}{LCSS}$	$\frac{2DD_{LCSS} - DD_{LCSS}}{DD_{LCSS}}$
MEAN	-33.21	-37.48	-8.25

The average relative error reduction for all data sets is equal to 33.21% for DD_{LCSS} and 37.48% for $2DD_{LCSS}$ compared to LCSS. We see that the new methods are clearly superior to LCSS distance on most of the examined data sets. Additionally we can see that $2DD_{LCSS}$ method outperforms DD_{LCSS} (average relative error reduction for all data sets is equal to 8.25%).

Results

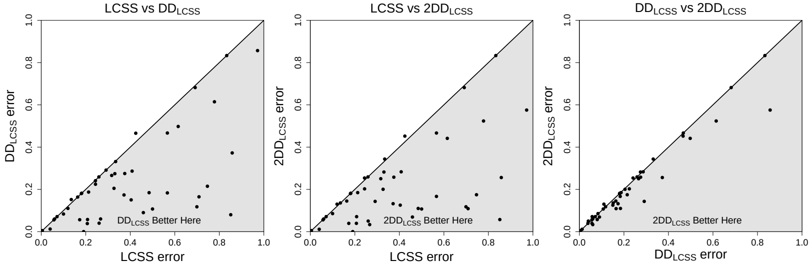

Figure : Comparison of test errors.

To find differences between the methods we used the Iman and Davenport (1980) test, which is a nonparametric equivalent of ANOVA. Due to the fact that the *p*-value is equal to 0, we can proceed with the post-hoc test in order to detect significant pairwise differences among all the classifiers. As a post-hoc test we used Bergmann and Hommel (1988) dynamic procedure.

Procedure	Ranks mean		
$2DD_{LCSS}$	1.51	*	
DD_{LCSS}	1.83	*	
LCSS	2.66		*

Finally, we have two homogeneous disjoint groups of classifiers. The best classifiers are in the first group.

References

- Górecki, T.: Two parametrical derivative dynamic time warping. In: Data analysis methods and its applications (eds. J. Pociecha, R. Decker) C.H. Beck. pp. 31–40 (2012).
- Górecki, T., Łuczak, M.: Using derivatives in time series classification. Data Mining and Knowledge Discovery. 26(2), pp. 310–331 (2013).
- Górecki, T., Łuczak, M.: First and second derivative in time series classification using DTW. Communications in Statistics - Simulation and Computation – accepted.
- Keogh, E., Zhu, Q., Hu, B., Hao. Y., Xi, X., Wei, L. & Ratanamahatana, C. A. (2011). The UCR Time Series Classification/Clustering Homepage: http://www.cs.ucr.edu/~eamonn/time_series_data/
- Vlachos, M., Gunopoulos, D., Kollios, G.: Discovering Similar Multidimensional Trajectories. In: Proceedings of the 18th International Conference on Data Engineering (2002).