elastic Kernelized Averaging of Time Series (eKATS) Vademecum and Expected Results

pierre-francois DOT marteau AT irisa DOT fr

December 2015

Contents

1 Introduction

eKATS, that stands for "elastic Kernalized Averaging of Time Series", is a code aiming at providing time elastic centroid estimates (or average) for a set of (multidimensional) time series. This kind of averaging technique revolves around the Dynamic Time Warping measure (DTW). eKats is an ANSI C++ implementation of the algorithms described in [2] and [1]. It includes also a C++ port of the DBA algorithm described in [3].

This software and description is free delivered "AS IS" with no guaranties for work at all. Its up to you testing it or modify it as you like. I will answer short relevant questions and help as my time allow it. I have tested it, played with it (on linux UBUNTU 14.* 15.* platforms), and found no problem in stability or malfunction so far. Thank you for sending me some feedback about it would you have any experience with it.

2 Dependencies, external codes or Libraries

1. DBA code has been translated (from java to C++) from the source code available at https://github.com/fpetitjean/DBA
   For more details on DBA, see F. Petitjean & P. Gançarski, "Summarizing a Set of Time Series by Averaging: from Steiner Sequence to Compact Multiple Alignment," Theoretical Computer Science; 2012.

2. The C clustering library, Copyright (C) 2002 Michiel Jan Laurens de Hoon. This library was written at the Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, * 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. Contact: mdehoon 'AT' gsc.riken.jp

3. Requires cmake http://www.cmake.org/

3 Installing the code

1. Unzip files in /eKATS1.0

2. Edit `CMakeLists.txt` and update the path where you will install the eKATS1.0 files. Source files are in `KATS1.0/src`

3. in eKATS1.0 directory, type the two following commands:

   ```
   $ cmake .
   $ make
   ```

   Should compile libkats in `./src` directory and should produce executable eKats in `eKATS1.0/bin` directory

4 Running the code

The executable is in `eKATS1.0/bin`: the command `./eKats -h` produces the following help message:

```
./eKats -sigma <sigma> -path <PATH> -DS <dataset>
        -TYPE <DTW_MED|KDTW_MED|iDBA|pKTDW-PWA|iKDBA|ALL>

Options:
-sigma <sigma> : value for sigma (sigma = 1.0/nu in the papers) for KDTW_MED,
   pKDTW-PWA and iKDBA
-path <PATH> : path to locate the data set
-DS <dataset> : name of the data set. <dataset> should exist in <PATH>
    directory in the UCR or MDTS format
-TYPE <DTW_MED|KDTW_MED|iDBA|pKTDW-PWA|iKDBA|ALL> is the type of centroid
   or methoid method
-MDTS : the dataset file is in the MDTS format.
-UCR : the dataset file is in the UCR format.
```

Examples :

```
./eKats -h // help: list of options
./eKats -PATH ../DATASETS/ -sigma 1 -DS CBF -TYPE ALL // all methods are
        processed, KDTW based methods with sigma = 1
./eKats -PATH ../DATASETS/ -DS Trace  -TYPE pKDTW-PWA -sigma .5
     // pKDTW-PWA method only, with sigma = .5
```

5 Expected Results

Some results (medoids and centroid estimates) are given in directory ./eKATS1.0/octave.

In this same directory the octave/matlab file `viewAverage.m` allows for visualizing some results (`viewAverage(<dataset>, <nclass>)`

5.1 CBF dataset

The command `./eKats -PATH ../DATASETS/ -sigma 1 -DS CBF -TYPE ALL` will compute the medoids and centroid estimates that could be viewed by running the command `viewAverage('CBF', 3)` in matlab or octave. The figures reproduced in Table 1 should then be generated.

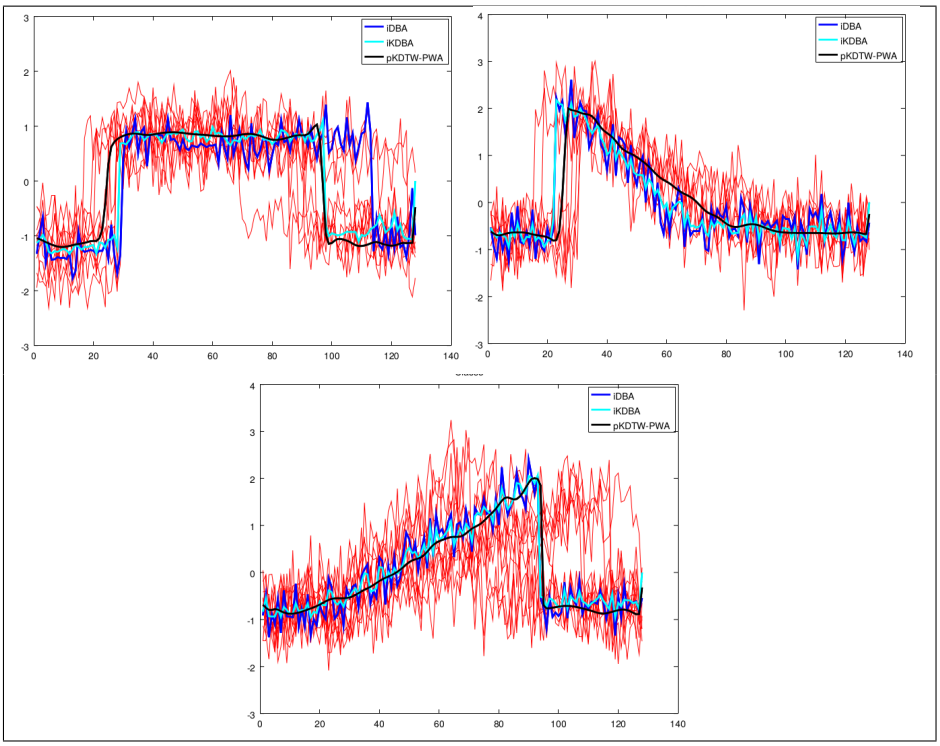

Table 1: Centroid estimation for the three categories of the CBF dataset. For each category, the centroid estimates are indicated as bold lines superimposed on some sample time series (in light red). The centroid estimates provided by the DBA algorithm are shown in blue bold lines, the estimates provided by the iKDBA algorithm in cyan bold lines and the estimates provided by the pKDTW-PWA algorithm in black bold lines.

5.2 synthetic_control dataset

The command `./eKats -PATH ../DATASETS/ -sigma .5 -DS synthetic_control -TYPE ALL` will compute the medoids and centroid estimates for dataset CBF that could be viewed by running the command `viewAverage('CBF', 3)` in matlab or octave. The figures reproduced in Table 2 should then be generated.

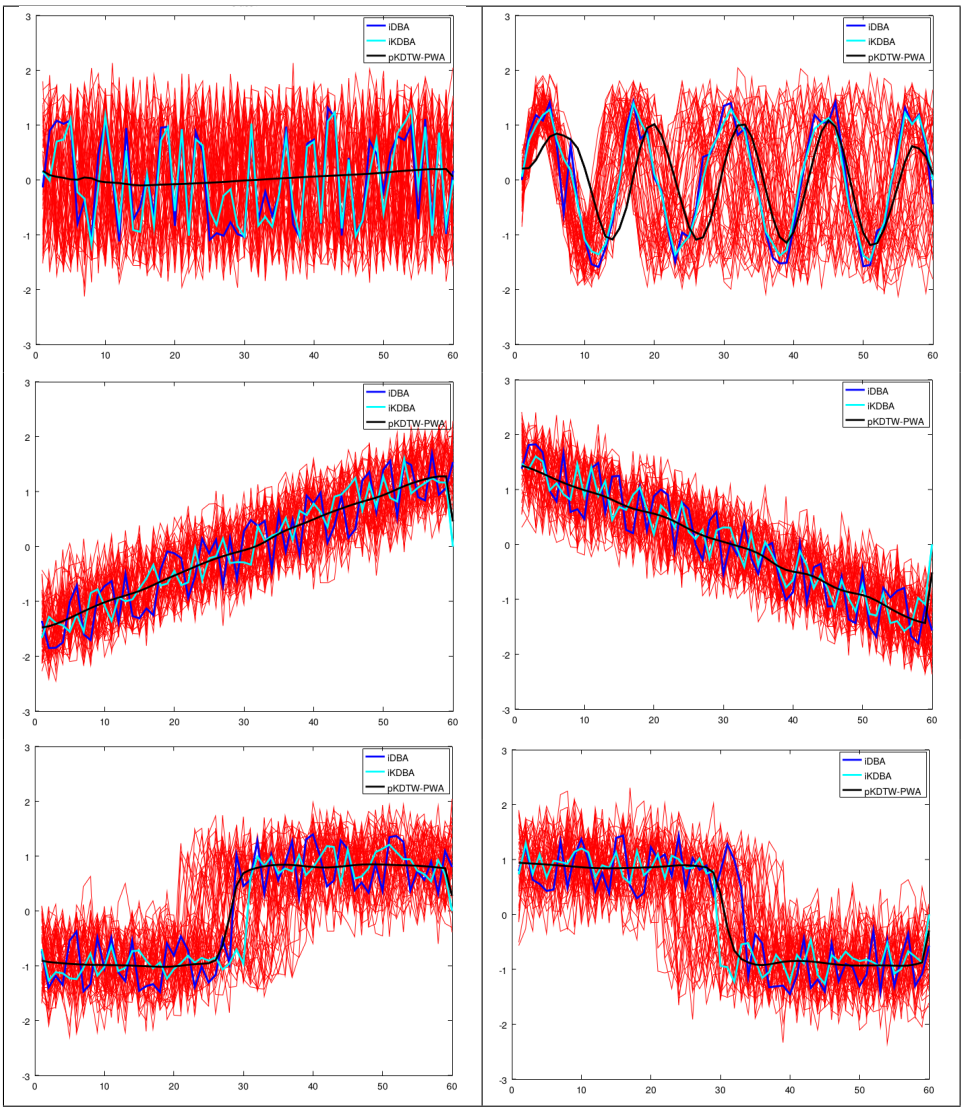

Table 2: Centroid estimation for the six categories of the synthetic_control dataset. For each category, the centroid estimates are indicated as bold lines superimposed on some sample time series (in light red). The centroid estimates provided by the DBA algorithm are shown in blue bold lines, the estimates provided by the iKDBA algorithm in cyan bold lines and the estimates provided by the pKDTW-PWA algorithm in black bold lines.

6 Dataset File Formats

6.1 UCR (ASCII)

1. each line corresponds to a time series (ASCII `'\n'` is the end of line)
2. the token separator is the SPACE character (ASCII code 32)

3. first token (in general an INT) corresponds to the class label
4. following tokens (generally floats) correspond to the 1D samples

6.2 MDTS (ASCII) : local (eKATS) format for multidimensional time series

1. each line corresponds to a time series (ASCII `'\n'` is the end of line)
2. the token separator is the SPACE character (ASCII code 32)
3. first token (in general an INT) corresponds to the subject id
4. second token (in general an INT) corresponds to the class label
5. For a D dimensional time series of length N, the following tokens are presented in sequence as follows:

```
0:<time_stamp_1> 1:<sample_1_1> 2:<sample_1_2> ... D:<sample_1_D>
D+1:<time_stamp_2> D+2:<sample_2_1> D+3:<sample_2_2> ... 2D+1:<sample_2_D> ...
(N-1)(D+1):<time_stamp_N> (N-1)(D+1)+1:<sample_N_1> (N-1)(D+1)+2:<sample_N_2> ...
N(D+1)-1:<sample_N_D>|
```

References

[1] Pierre-François Marteau. Times series averaging from a probabilistic interpretation of time-elastic kernel. *CoRR*, abs/1505.06897, 2015.

[2] Pierre-François Marteau and Sylvie Gibet. On Recursive Edit Distance Kernels with Application to Time Series Classification. *IEEE Trans. on Neural Networks and Learning Systems*, pages 1–14, June 2014.

[3] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn.*, 44(3):678–693, March 2011.