Individualized Warping Window Size for Dynamic Time Warping

Krisztian Buza[1] Ladislav Peška[1,2]

[1] Telekom Innovation Laboratories, Department of Data Science and Engineering, Faculty of Informatics
ELTE - Eötvös Loránd University, Budapest, Hungary, buza@inf.elte.hu
[2] Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

1 Summary

We focus on learning individualized warping window sizes (WWS) for time series classification based on dynamic time warping (DTW). DTW is one of the most popular distance measures for time series classification and it was shown that its warping window size is crucial for the final accuracy of the model [1]. WWS is therefore considered as an important parameter of DTW.
In contrast to the previous works, in which static WWS was used, i.e., the WWS size was selected for the entire dataset, we propose a hubness-aware approach to select WWS for each instance *individually*, denoted as IWW in Results. We evaluate our approach on publicly available real-world datasets and show that the classification accuracy using individualized WWS is significantly higher than the accuracy in case of static WWS.

2 Background

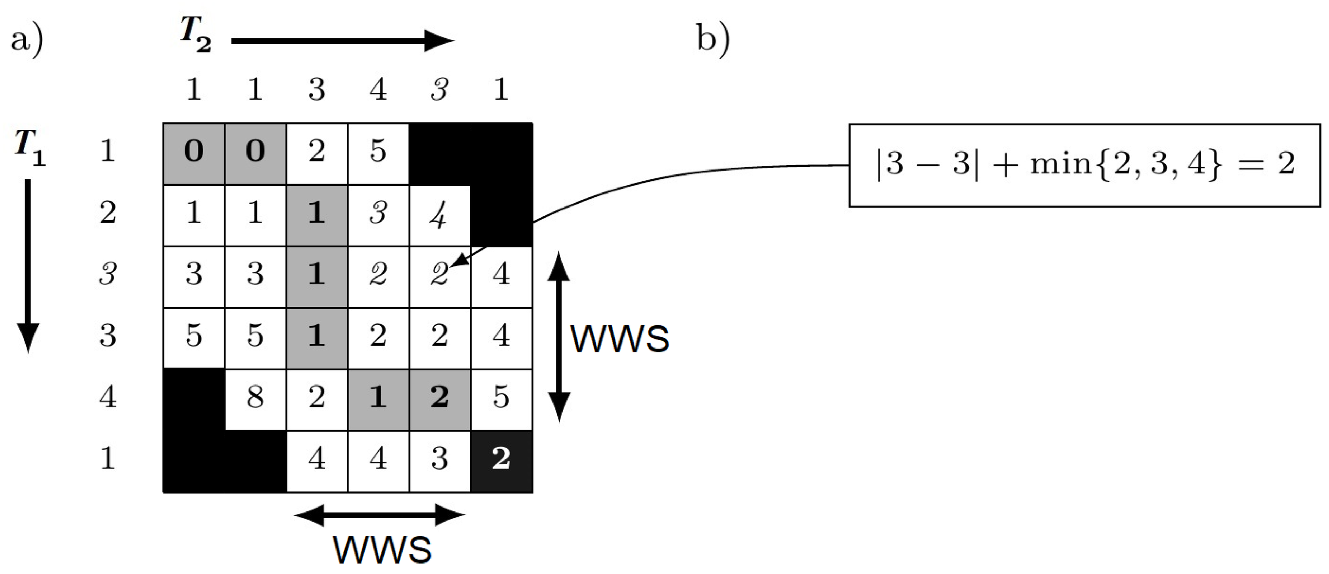
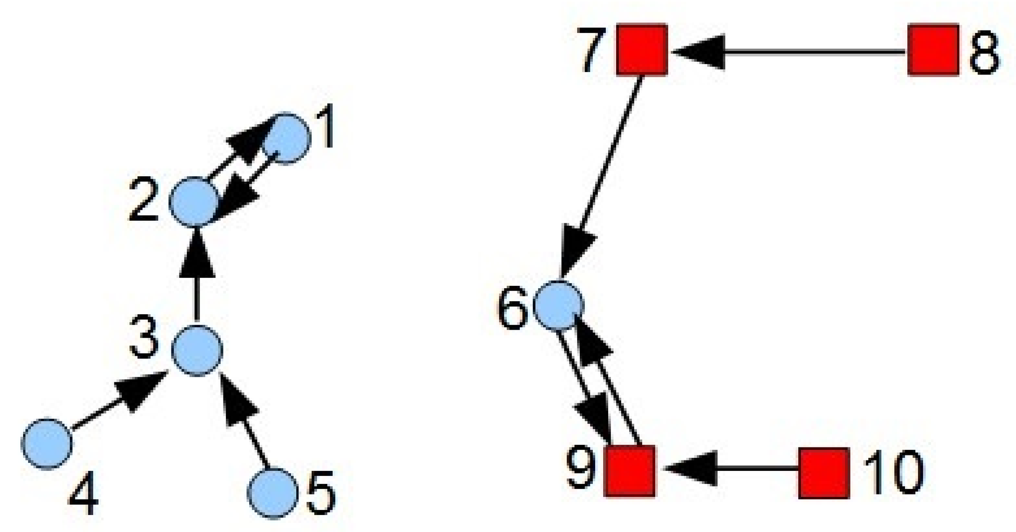


Fig. 1. Dynamic Time Warping



Instance:	1	2	3	4	5	6	7	8	9	10
$N_1(x)$	1	2	2	0	0	2	1	0	2	0
$GN_1(x)$	1	2	2	0	0	0	1	0	1	0
$BN_1(x)$	0	0	0	0	0	2	0	0	1	0

Fig. 2. Good and Bad Neighbors [2]

Acknowledgement

This work was supported by the project no. 20460-3/2018/ FEKUTSTRAT within the Institutional Excellence Program in Higher Education of the Hungarian Ministry of Human Capacities.

3 Our Approach

Algorithm 1 Individualized warping window size selection for a time series T

Input: set of labelled training time series $\mathcal{D}$, time series T, number of nearest neighbors k

$w_{max} = 0.1 \cdot \text{length}(T)$
$w_{best} = 0$
$score_{best} = 0$

for each $w \in 0, \ldots, w_{max}$ **do**
  $score = 0$
  # $\mathcal{N}^{\mathcal{D}}_{k,w}(T)$ = the set of k-nearest neighbors of T
  # among the time series in $\mathcal{D}$, calculated with
  # warping window size w
  for each $T' \in \mathcal{N}^{\mathcal{D}}_{k,w}(T)$ **do**
    $score = score + GN_{k,w}(T') - BN_{k,w}(T')$
    # $GN_{k,w}(T')$ and $BN_{k,w}(T')$ are calculated
    # on $\mathcal{D}$ with warping window size w
  end for
  if $score > score_{best}$ **then**
    $w_{best} = w$
    $score_{best} = score$
  end if
end for
return w_{best}

4 Results

Table 1. Classification accuracy (averaged over the 10×10 folds) $\pm$ its standard deviation for k-nearest neighbor in case of calculating full DTW, DTW with global warping widow sizes of 5% and 10% and individualized warping window sizes (IWW). The best approach is underlined.

DATASET	FULL DTW	WWS=5 %	WWS=10 %	IWW
		$k = 5$		
ARROWHEAD	0.794±0.087	0.817±0.078	0.799±0.087	0.871±0.072
CAR	0.672±0.120	0.708±0.136	0.685±0.116	0.747±0.105
DISTALPHALANXOUTLINEAGEGROUP	0.819±0.049	0.817±0.047	0.818±0.049	0.824±0.042
DISTALPHALANXTW	0.725±0.053	0.727±0.052	0.727±0.052	0.753±0.052
DODGERLOOPDAY	0.398±0.125	0.448±0.127	0.411±0.126	0.537±0.121
		$k = 10$		
ARROWHEAD	0.792±0.083	0.819±0.075	0.800±0.084	0.870±0.069
CAR	0.612±0.139	0.680±0.134	0.624±0.141	0.697±0.134
DISTALPHALANXOUTLINEAGEGROUP	0.808±0.048	0.813±0.046	0.808±0.048	0.821±0.044
DISTALPHALANXTW	0.752±0.051	0.752±0.051	0.752±0.051	0.749±0.051
DODGERLOOPDAY	0.384±0.125	0.405±0.106	0.375±0.122	0.510±0.113

References

[1] Dau et al.: Optimizing dynamic time warping's window width for time series data mining applications. Data Mining and Knowledge Discovery (2018): 1-47.
[2] Tomasev et al.: Hubness-aware Classification, Instance Selection and Feature Construction: Survey and Extensions to Time-Series In: Feature selection for data and pattern recognition, Springer-Verlag (2015): 231-262.