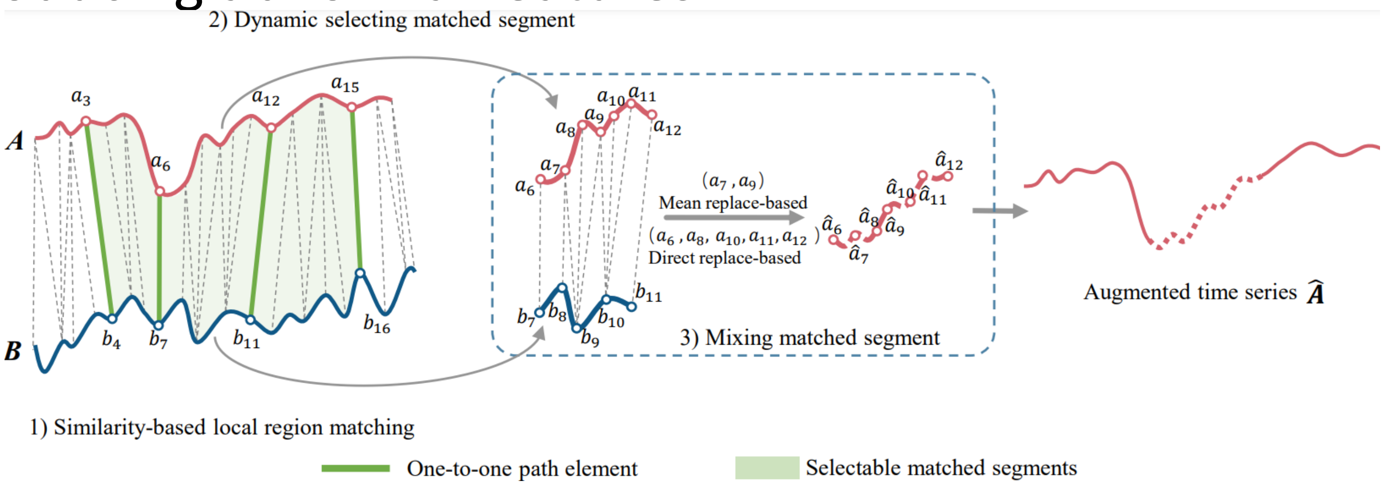
ISM: Intra-class Similarity Mixing for Time Series Augmentation

Pin LIU, Rui WANG, Yongqiang HE, Yuzhu WANG

Frontiers of Computer Science, DOI: 10.1007/s11704-024-40110-9

Problems & Ideas

- The overfitting problem of the time series classification model:
  - The existing method mixes the global region with similar characteristics, making introducing abnormal features easy.
  - Mixing the region without the quantitative indicator can quickly generate abnormal samples, which mislead classification decisions.
- Ideas: An augmentation method that mixes two local segments with similar features from within the same class of time series, thereby improving sample diversity while minimizing the risk of introducing abnormal features



The overall pipeline performs time series augmentation from two original time series to augmented time series. The core steps are as follows: 1) Similarity-based local region matching. 2) Dynamic selecting matched segment. 3) Mixing matched segment.

Main Contributions

- Contributions:
  - We propose a local region mixing method based on intra-class similarity, which is the first to take the local similarity into time series augmentation;
  - The proposed mixing strategy can perform better on the non-equal length segments of time series, overcoming the information loss and noise introduction defects;
  - Comprehensive experiments show that ISM can achieve superior performance and better interpretability than state-of-the-art methods.

Table 1 Classification performance (Mean Acc.%). F and R represent the backbones of the FCN and ResNet, respectively.

Dataset	None		DTWmerge		wDBA		DGW(sD)		ISM	
	F	R	F	R	F	R	F	R	F	R
Wine	58.7	74.0	81.5	77.8	84.0	92.6	73.2	88.1	**88.9**	**94.9**
Car	90.5	92.5	67.5	78.3	90.7	92.7	92.3	93.8	**95.0**	**95.3**
DistalPT.	69.0	66.5	69.8	68.4	73.4	69.3	72.0	69.0	**76.3**	**71.9**
Lightning7	82.7	84.5	78.8	80.8	89.7	86.5	83.6	84.9	**89.9**	**89.0**
Yoga	82.7	87.0	71.4	82.7	79.0	87.4	81.8	88.0	**88.2**	**89.1**
WormsTC.	62.5	74.7	79.2	79.2	80.1	76.4	80.4	80.3	**83.1**	**87.8**
WordS.	46.4	52.2	37.5	52.0	47.4	54.0	49.1	57.4	**51.4**	**60.7**
Computers	72.6	81.5	80.6	79.6	84.0	72.4	71.6	84.0	**86.4**	**86.6**
Mallat	96.7	96.6	75.8	91.6	94.8	92.4	95.7	94.6	**98.9**	**97.6**
FordA	90.4	93.4	90.3	93.1	90.0	93.5	87.1	92.8	**91.6**	**94.7**
Average	75.2	80.3	73.2	78.4	81.3	81.7	78.7	83.3	**85.0**	**86.8**

Table 2 The time consumption (Minute) on the typical datasets.

Method		WordS.	FordA	Mallat	Average
None	Fcn	2.1	9.2	3.8	5.0
wDBA	Fcn	1819.0	12542.5	2781.2	5714.2
DGW(sD)	Fcn	105.2	969.6	204.0	426.3
ISM	Fcn	61.1	467.9	131.8	220.2
None	ResNet	37.2	120.5	56.5	71.4
wDBA	ResNet	1832.6	12530.2	2786.8	5716.5
DGW(sD)	ResNet	121.1	935.4	221.6	426.0
ISM	ResNet	62.3	478.6	134.2	225.0

Performance comparison of ISM with state-of-the-art methods on typical datasets. Left: Accuracy and standard error using the FCN and ResNet architecture; Right: The time consumption (Minute).