DTW-based voting for multivariate time-series classification

Dec 02, 2015

Je Hyuk Lee
Dept of Industrial Engineering, SNU

Contents

- 1. Introduction
- 2. Background
- 3. Experiment
- 4. Results
- 5. Conclusion

Section1

INTRODUCTION

Introduction

- **Time-series data**
  - a sequence of data points, typically consisting of successive measurements made over a time interval
  - These days, these kinds of data are widely used in many different area
    - Medicine (Tormene et al., 2009)
    - Finance (Rada, 2008)
    - Bioinformatics (Aach & Church, 2001)

- Univariate time series data have been well-studied
  - Distance measure: Euclidean, DTW,…
  - Representation: DWT, DFT, SAX, …
  - 1NN-DTW method is difficult to defeat

Introduction

- **Multivariate Time-series data**
  - A kind of time series data that consists of two or more variables

- But, MTS(Multivariate Time Series) is not well-studied
  - It is very different from univariate time series
  - The main difference is a correlation among variables

- Two approaches of MTS similarity measure
  - Compare the TS variable by variable
  - Compare the TS as a whole

Introduction

- **In this research,**
  - We conducted a classification problem by using DTW and voting method
  - 3 voting method
    - Voting from each attribute
    - Voting from projected sequence on hyperplane which is spanned by principal components(PCs)
    - Voting based on hyperplane similarity spanned by PCs.

Section2

BACKGROUND

Background

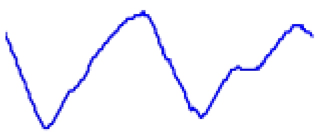
- Background Contents
  - DTW
  - PCA
  - History of MTS classification problem

Section3

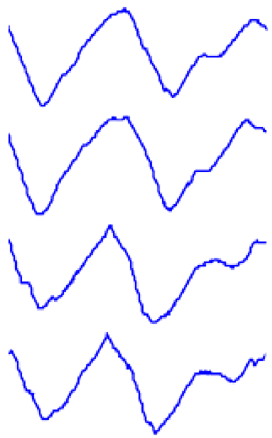
EXPERIMENTS

Proposed Methods

- **1. DTW + 1NN classifier for each variable and vote**

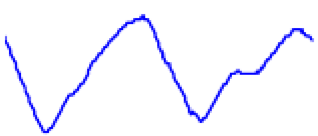


Test data(var 1)

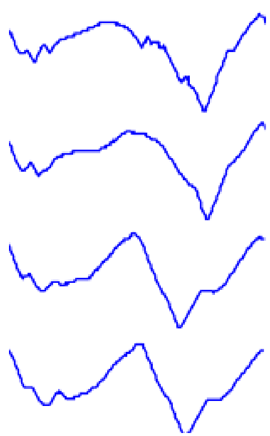


Training data(var 1)

DTW
1(A)
2(B)
3(C)
4(A)



Test data(var 2)



Training data(var 2)

DTW
4(D)
3(B)
5(A)
2(C)

Class	# of var
A	**5**
B	3
C	2
D	0

Test data is classified to A class

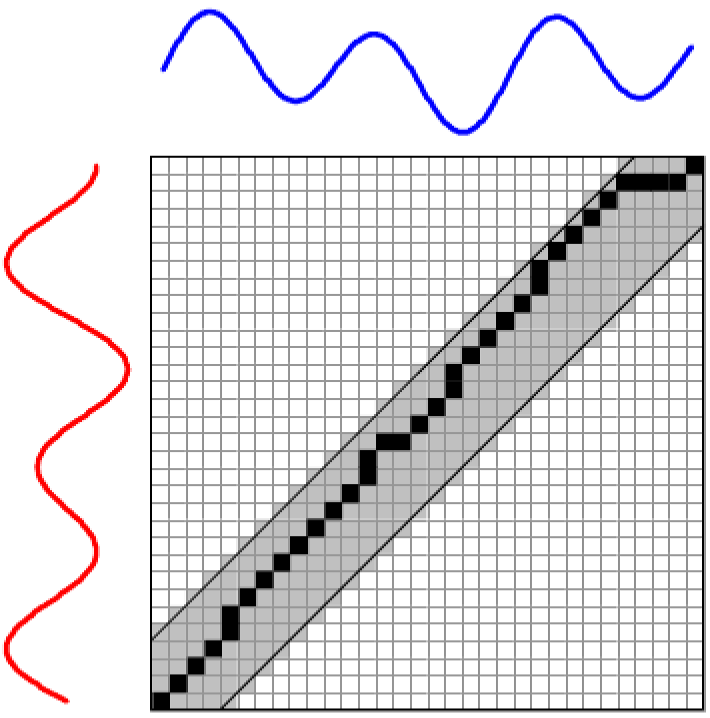
Proposed Methods

- Method (1) is simple and easy to understand
  - But it does not include anything about correlation structure
  - Also, if each variable have correlation structure
    - Some variables can overly cause influence to vote results

- We need two constraints
  - Sequences need to include correlation structure
  - Variable for voting should be nearly independent
  - How about using PCA?

Proposed Method

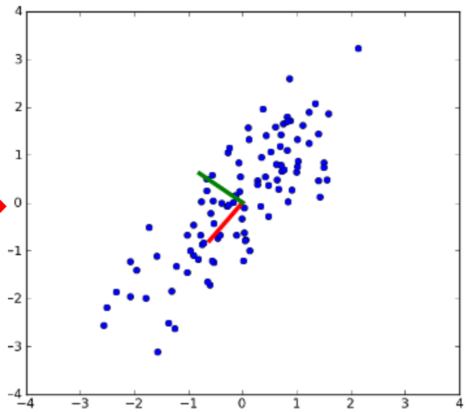
- Also, to avoid 'the longer sequence, the longer distance'
  - We divided the DTW distance by sequence length



Proposed Methods

- 2. Use PCA projected sequence. Then, DTW+1NN+voting classifier

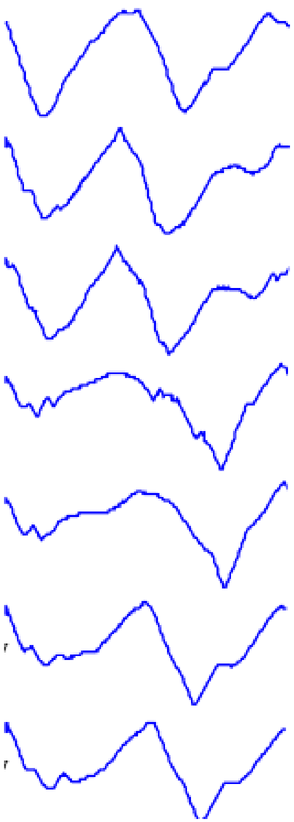
	v1	v2	…	vM
t1	a11	a12		a1M
t2	a21	a22		a2M
t3	a31	a32		a3M
t4	a41	a42		a4M
…				
tN	aN1	a2N		aMN



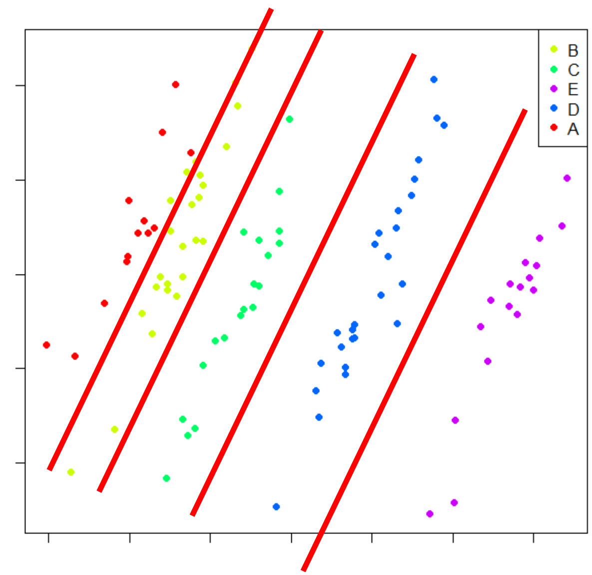
	v1	v2	…	vM
t1	a11'	a12'		a1M'
t2	a21'	a22'		a2M'
t3	a31'	a32'		a3M'
t4	a41'	a42'		a4M'
…				
tN	aN1'	a2N'		aMN'

Proposed Methods

- **3. Project to the coordinates and use conventional classifier (Not yet)**
  - First, calculate the DTW distance matrix for training data
  - Projected to the coordinate space (How? MDS??)
  - Use conventional classifier

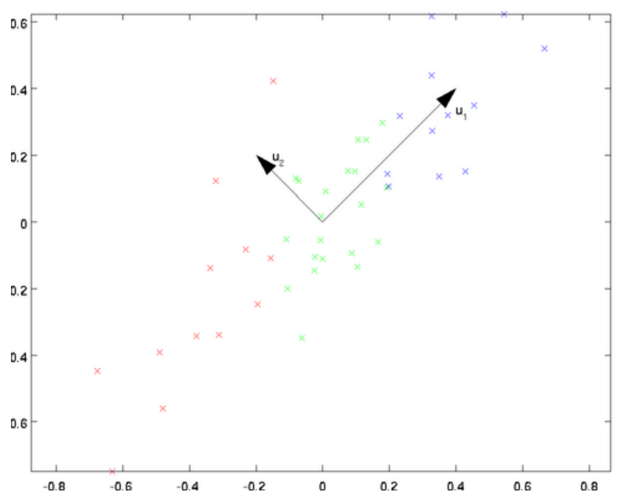


	s1	s2	s3	s4
s1				
s2				
s3				
s4				

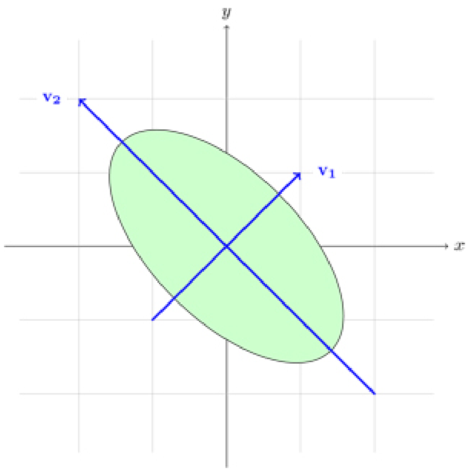


Proposed Method

- (Sub) How about using Krzanowski distance(1-SPCA)?
  - The distance of two hyper plane which is made by PCs?



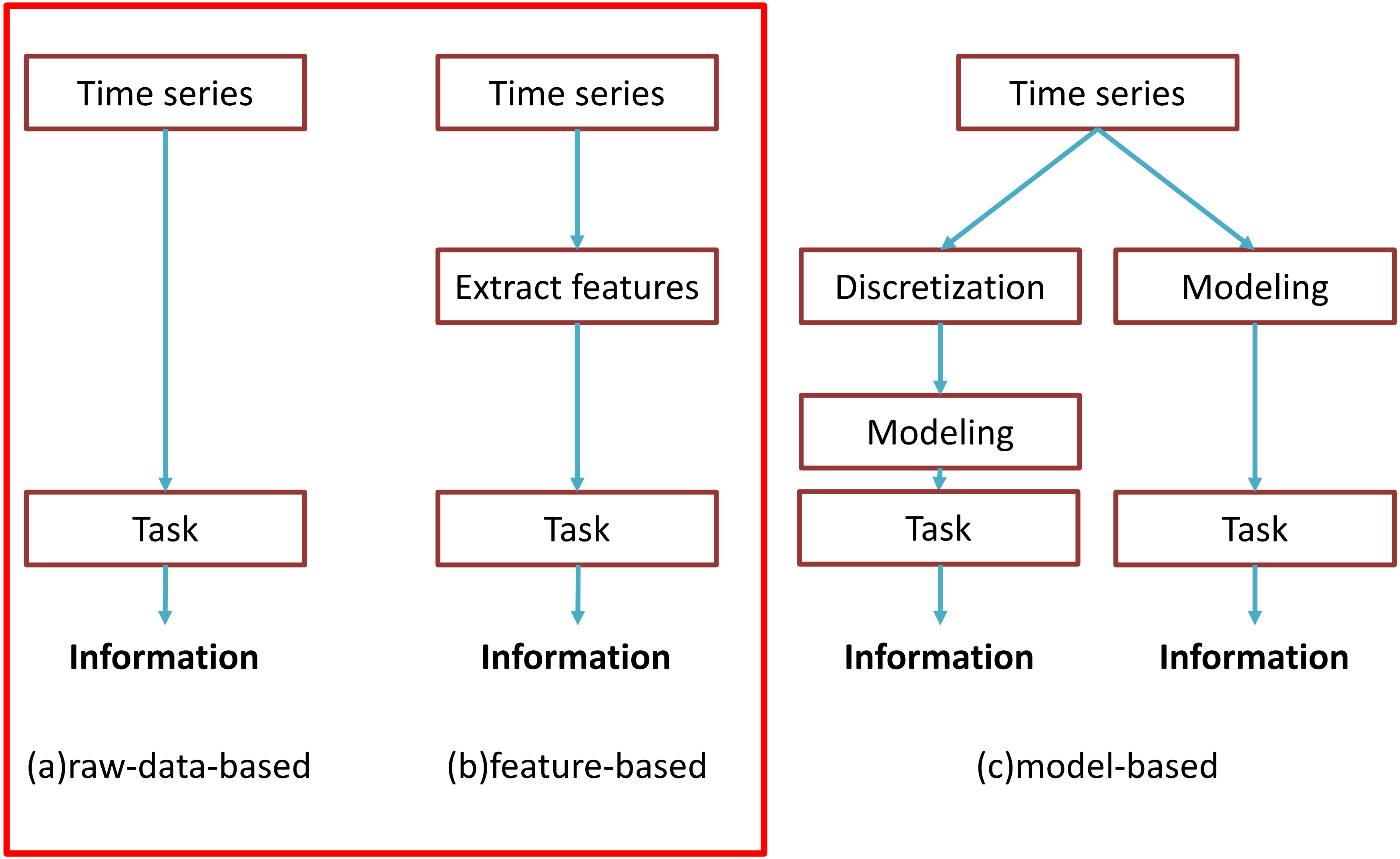
PCs from data 1



PCs from data 2

Subspace1: subspace spanned by PCs from data1
Subspace2: subspace spanned by PCs from data2
Krzanowski distance: Distance between Subspace1 and Subsapce2

Time series data mining



Experiment

- **Dataset**

	Name	# of classes	# of Variables	Length	Training size	Test size
UCI	AUSLAN	95	22	45~136	(1140)	(1425)
	Pendigits	10	2	8	300	10692
	Japanese Vowels	9	12	7-29	270	370
	Arabic Digits	10	13	4~93	6600	2200
	Character Trajectories	20	3	109~205	(2058)	(800)
	ECG	2	2	39~152	100	100
	Wafer	2	6	104~198	298	896

Experiment

- Comparing the experimental results for each data set

- 2-class classification results
  - Select 2 classes randomly (10times) and averaging the accuracy

Section4

RESULTS

Experiment

- **Accuracy Results**

Name	DTW+1NN (All)	DTW+PCA (All)	PCA coeff (All)	DTW+1NN (2-class)	DTW+PCA (2-class)	PCA coeff (2-class)
AUSLAN	40.35%	11.37% (2PCs)	43.58% (2PCs)	73.33%	71% (3PCs)	73% (4PCs)
Japanese Vowels	73.78%	28.92% (1 PCs)	42.7% (2PCs)	76.17%	83.96% (2PCs)	75.99% (4PCs)
Arabic Digits			17.36% (1PC)	99.55%	96.59% (1PC)	98.41% (2PCs)
Character Trajectories	85.97%	86.16% (3PCs)	30.61% (2PCs)	89.93%	85.78% (3PCs)	67.89% (1PC)
ECG	73.00%	75.00% (2PCs)	67% (1PC)	-	-	-
Wafer	93.97%	94.08% (1PC)	89.4% (4PCs)	-	-	-

Experiment

- **Classification Time Results**

Name	DTW+1NN (All)	DTW+PCA (All)	PCA coeff (All)	DTW+1NN (2-class)	DTW+PCA (2-class)	PCA coeff (2-class)
AUSLAN	5041.41s	939.15s	22.64s	4.26s	6.43s	0.37s
Japanese Vowels	160.60s	11.02s	2.20s	8.37s	1.46s	0.24s
Arabic Digits	(>2.5days)		487.40s		379.12s	8.40s
Character Trajectories	8121.11s	7406.07s	16.83s	91.72s	57.93s	0.43s
ECG	18.03s	16.33s	0.40s	-	-	-
Wafer	1524.99s	361.52s	4.54s	-	-	-

Section5

CONCLUSION

Conclusion

- For multi-class problem, the proposed method's performance is poor
  - For attribute wise voting method, its performance is not so bad
  - But, for PCA-based voting method, its performance is similar to random guessing

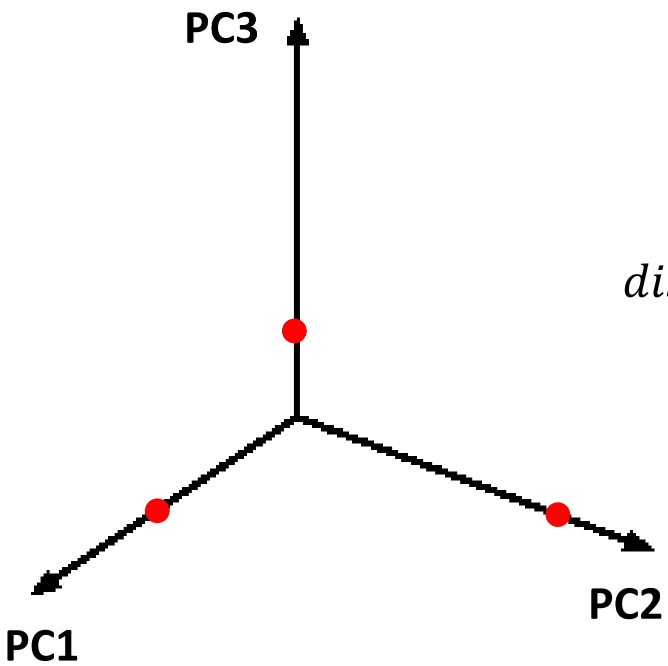
- But, for 2-class problem, their performance is almost same

- For multi-class problem,
  - MTS correlation structure might affect the performance difference
  - How might be?

- When they are applied to real data, how might their performance be?

Further Research

- To consider the correlation structure, the PCA-DTW method would be proper
  - Instead of PCA-DTW voting, how about calculating weighted DTW distance?
  - Weight is determined by each PCs variance



$$dist = \sqrt{(w_1 DTW_1)^2 + (w_2 DTW_2)^2 + (w_3 DTW_3)^2}$$

References

- Aach, John, and George M. Church. "Aligning gene expression time series with time warping algorithms." *Bioinformatics* 17.6 (2001): 495-508.
- Abonyi, Janos, et al. "Principal component analysis based time series segmentation—application to hierarchical clustering for multivariate process data." *Proc, of the IEEE Int. Conf. on Computational Cybernetics*. 2003.
- Bankó, Zoltán, and János Abonyi. "Correlation based dynamic time warping of multivariate time series." *Expert Systems with Applications* 39.17 (2012): 12814-12823.
- Baydogan, Mustafa Gokce, George Runger, and Eugene Tuv. "A bag-of-features framework to classify time series." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.11 (2013): 2796-2802.
- Baydogan, Mustafa Gokce, and George Runger. "Learning a symbolic representation for multivariate time series classification." *Data Mining and Knowledge Discovery* 29.2 (2014): 400-422.
- Krzanowski, W. J. "Between-groups comparison of principal components." *Journal of the American Statistical Association* 74.367 (1979): 703-707.
- Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
- Olszewski RT (2012) http://www.cs.cmu.edu/~bobski/
- Rada, Roy. "Expert systems and evolutionary computing for financial investing: A review." *Expert systems with applications* 34.4 (2008): 2232-2240.
- Tormene, Paolo, et al. "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation." *Artificial intelligence in medicine* 45.1 (2009): 11-34.
- Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen and Gustavo Batista (2015). *The UCR Time Series Classification Archive.* URLwww.cs.ucr.edu/~eamonn/time_series_data/