Hong Yin, Shu-qiang Yang, Xiao-qian Zhu, Shao-dong Ma, Lu-min Zhang, 2015. Symbolic representation based on trend features for knowledge discovery in long time series. *Frontiers of Information Technology & Electronic Engineering*, **16**(9):744-758. [doi:10.1631/FITEE.1400376]

# Symbolic representation based on trend features for knowledge discovery in long time series

**Key words:** Long time series, Segmentation, Trend features, Symbolic, Knowledge discovery

Contact: Hong Yin

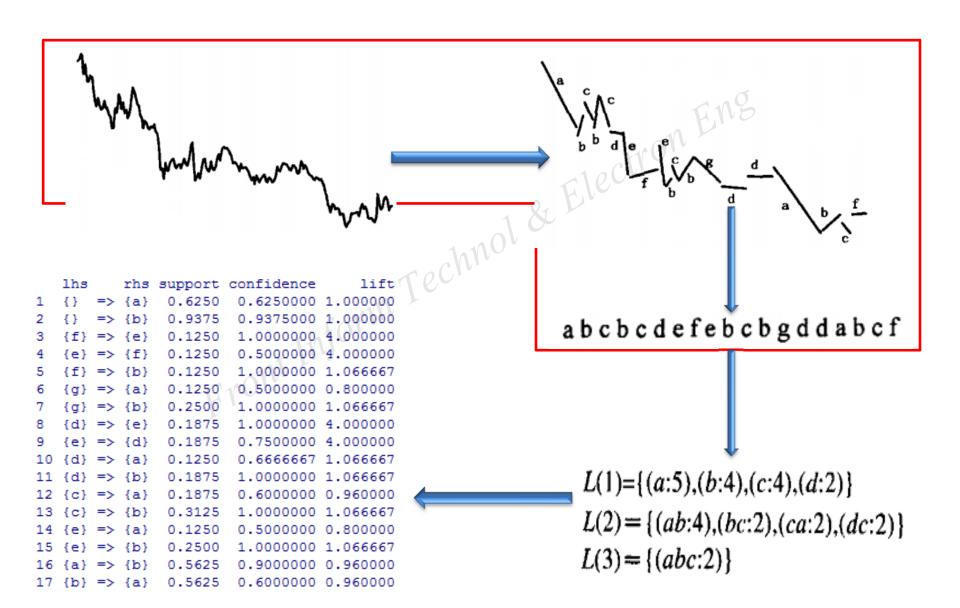
E-mail: yinhonggfkd@aliyun.com

ORCID: http://orcid.org/0000-0002-0682-6781

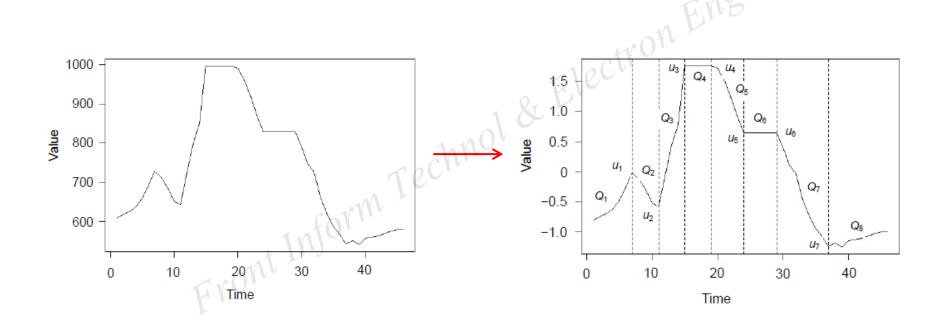
#### Introduction

- The symbolic representation of time series has attracted much research interest recently. The high dimensionality typical of the data is challenging, especially as the time series becomes longer.
- However, most time series merely explicate the potential information of using themselves, and many more underlying rules are derived by mining algorithms.
- The original contributions of this paper are an improvement in the segmentation efficiency of long time series, and the TFSA method which focuses on preserving the trend features of the original time series and interprets the rules obtained from mining time series.

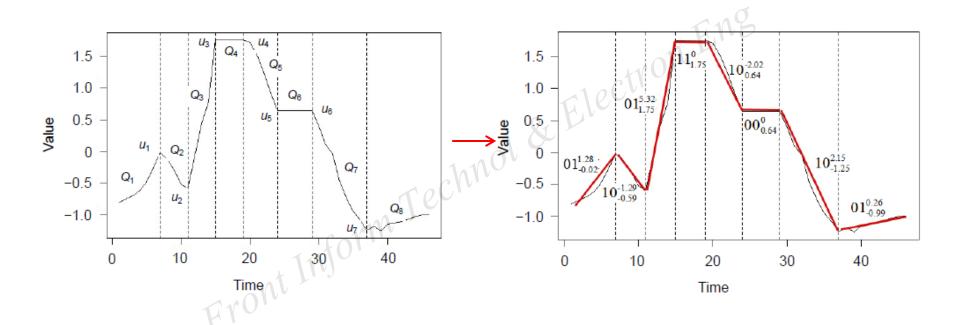
#### The basic idea on this work



### Two-step segmentation mechanism



#### The TFSA algorithm



 $Q_1Q_2Q_3Q_4Q_5Q_6Q_7Q_8 = 01_{-0.02}^{1.28}10_{-0.59}^{-1.29}01_{1.75}^{5.32}11_{1.75}^{0}10_{-0.64}^{-2.02}00_{0.64}^{0}10_{-1.25}^{2.15}01_{-0.99}^{0.26}$ 

# Measurement result: the efficiency of the two-step segmentation mechanism

Table 4 A comparison of segmentation efficiency with different time series lengths

Alcorithm	Execution time (s)					
Algorithm	@1000	10000	100000	1000000		
TD form	1.80	10.38	120.12	1570.35		
SW ( <i>fit</i> =20, step=1)	1.13	3.01	70.93	648.51		
Two-step segmentation						
k=50	2.29	4.49	10.52	21.86		
k=100	2.18	3.72	7.75	15.68		

## The results of mining time series

Table 7 Association rules obtained from time series

ID	Rule	Support	Confidence
1	$\{IT240-10J1\} \Longrightarrow \{IT267-01F1\}$	0.0430107	1.0000000
2	${IT240 \text{ offset fault}} \Longrightarrow {E240-01G1}$	0.0537634	0.7142857
3	{ISH236-10I1} => {E281-10D1}	0.0430107	0.8000000
4	$\{ISH236-10F1\} \Rightarrow \{IT267-01F1\}$	0.0645161	0.7500000
5	{IT240-01B1} => {E242-01I1}	0.0752688	0.7000000
6	$\{E281-01J1\} \Longrightarrow \{ESH244A-01I1\}$	0.0860215	0.6666667
7	$\{E265-10B1\} => \{E240-01G1\}$	0.1182795	0.6111111
8	{IT240-00C0, ISH236-01I1} => {ESH244A-01I1}	0.0430107	1.0000000
9	$\{IT240-01I1, E265-10B1, E240-01G1\} \Longrightarrow \{IT267-01F1\}$	0.0430107	1.0000000

#### **Conclusions**

- In this paper, we propose a two-step segmentation mechanism which can be used in parallel to improve the segmentation efficiency for long time series.
- Experimental results show that, especially for long time series, the segmentation efficiency and classification accuracy of TFSA are better than those of other methods.