

Repeating Sequence Length Estimation in Time Series

Author: Jakub Ševcech

Supervisor: prof. Mária Bielíková

- Representing repeating sequences as symbols
- Incremental transformation
 - Splitting to sequences
 - Sequence clustering
 - Cluster identifiers as symbols
- Symbol (repeating sequence) length estimation
- Autocorrelation vs. Repetitiveness
- Evaluation on synthetic datasets with known symbol lengths
- Comparison on real-world datasets

Symbol number	Symbol lengths	Noise	Repetitiveness accuracy	Autocorrelation accuracy
1	20	-	92%	100%
1	20	0.05	89%	100%
1	20	0.01	95%	100%
3	20	-	76%	95%
3	20	0.05	73%	90%
5	20	-	54%	93%
5	20	0.05	52%	82%
10	20	-	29%	54%
10	20	0.05	20%	55%
1	15,20	-	8%	62%
3	15,20	-	3%	45%
5	15,20	-	4%	22%

