



Improved Multi-way Time Series Join on Multi-length Patterns

17/11/2022

Group 1:

Yue Deng

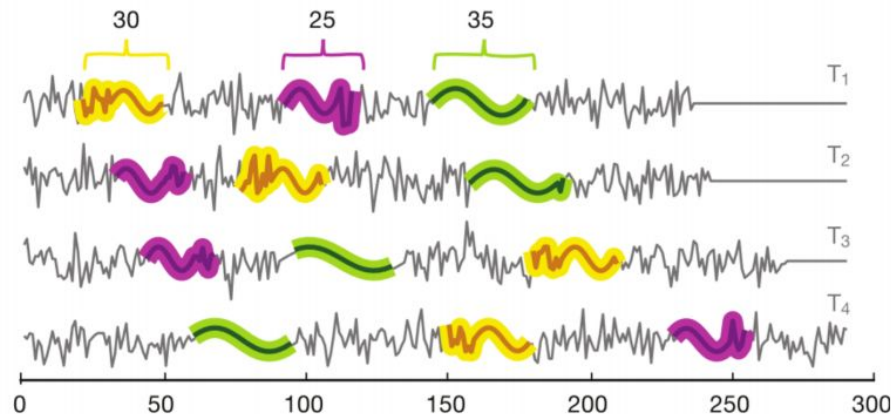
Jiaxi Zhang

Junze Li

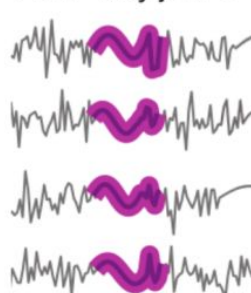
Zihan Li

Background

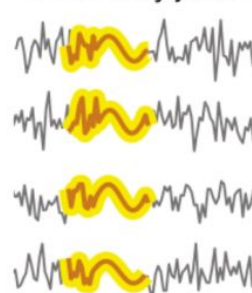
- Time series
- Multiple time series
- Subsequence
- Patterns
- Time series join (two-way and multi-way join)
- Non-trivial multi-way join



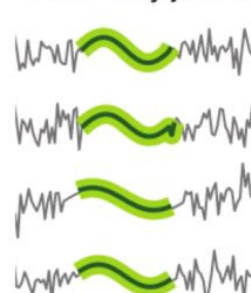
Four-way join 1



Four-way join 2



Four-way join 3





Related Work

Time series join

number of time series

single time series join: HIME and MOEN

two-way time series join: Jocor and LCS-Jocor

multi-way time series join : **mSTAMP**

Cannot satisfy multi way and variable length simultaneously!

fixed-length time series join: **mSTAMP**

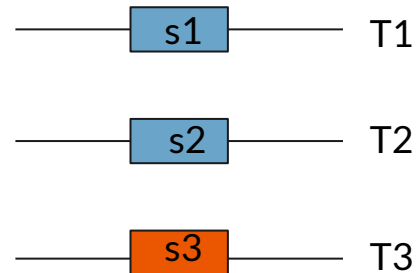
length of subsequence

variable-length time series join: HIME, Jocor and LCS-Jocor



MultiPAL(multi-way and multi-length)

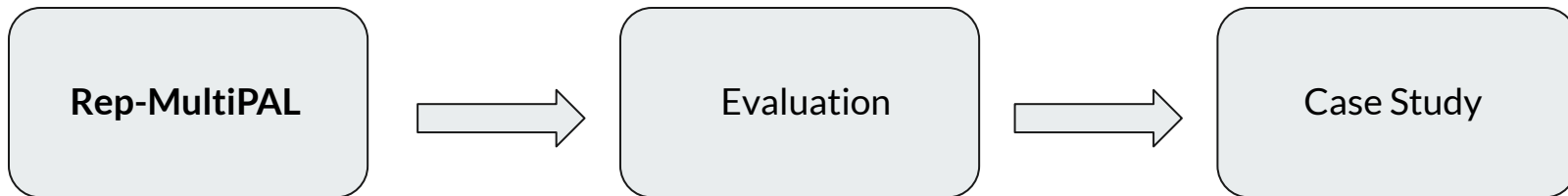
- Naive algorithm: the time complexity of multiway join on time series is exponential to the number of time series
- MultiPAL has three improvements:
 - only considers the similarity between the candidate subsequence and the last added one
 - applies BFS to better control memory
 - exploits Matrix Profile to develop the algorithm





Project Pipeline

Problem Definition: There is a collection of n time series. The goal is to discover at most k non-trivial multi-way joins ($N \leq n, L, C$) where k and subsequence length L is given by users to **maximize** the number of subsequences **N** and the minimum correlation coefficient **C** .

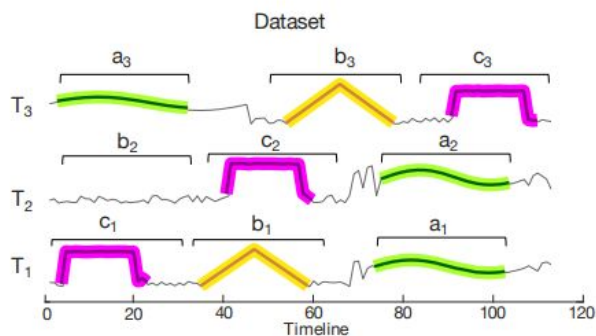




Methodology

- ❖ MultiPAL
- ❖ Rep-MultiPAL
- ❖ Similarity Evaluation

MultiPAL



An Intuitive Example

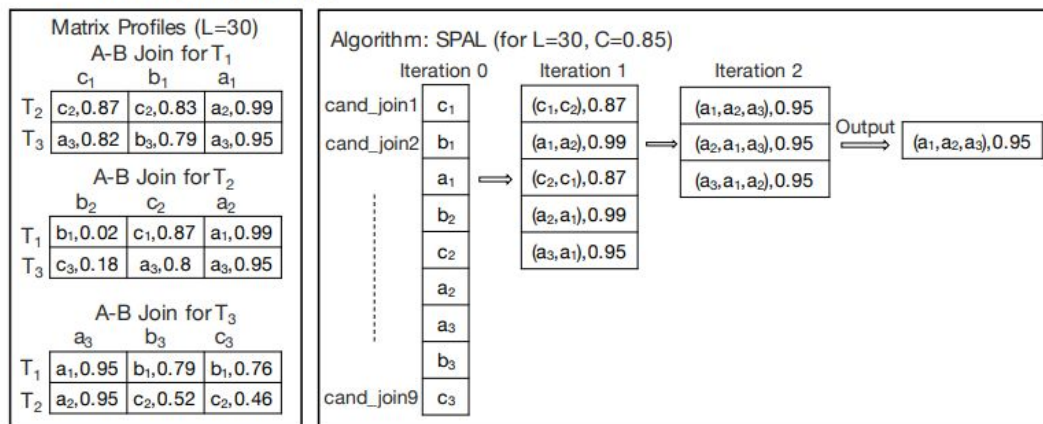


Fig. 3. An illustration of the proposed MultiPAL algorithm generating the best multi-way join for a length $L = 30$.

Rep-MultiPAL

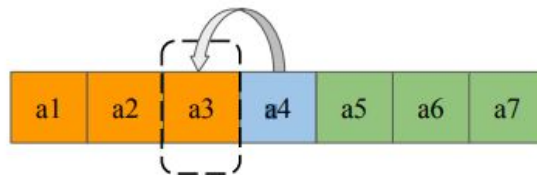
Parameter p : only when # of subsequences $< p$, compute the similarity matrix and reelect a representative one.

MultiPAL:

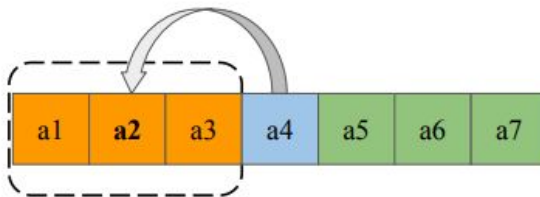
$$O((nm)^2 + n^2Mt + Mt \log(Mt))$$

Rep-MultiPAL:

$$O((nm)^2 + (n^2 + p^2)Mt + Mt \log(Mt))$$



(a) the schematic diagram of the original method



(b1) the schematic diagram of our proposed method

	a1	a2	a3
a1	1		
a2	0.92	1	
a3	0.87	0.9	1

(b2) the similarity matrix



Similarity Metrics



We also considered other metrics to evaluate the multi-way joins.

- Double Correlation Coefficient (DCO)
- k-Derivative Dynamic Time Warping (k-DDTW)
- Improved Periodogram-based Distance (IPD)

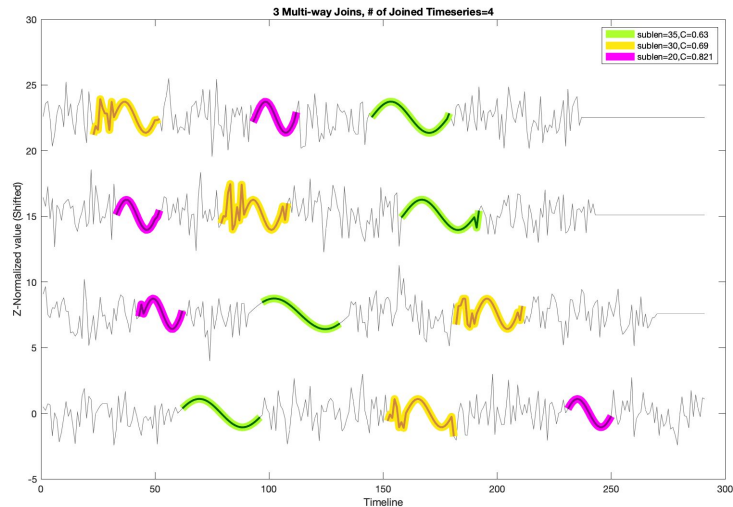


Experiment Settings

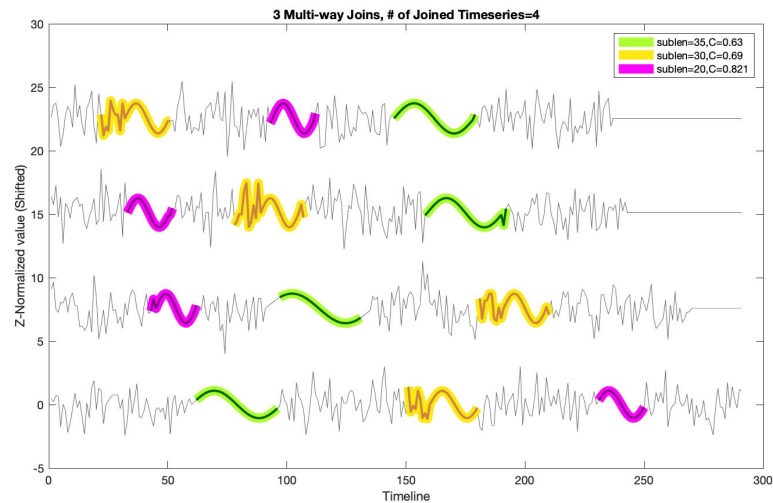
(Adaption: Choose different length of patterns (L) and combine the results together manually.)

- 
- 
- **SUSh**: The unsupervised shapelet discovery algorithm to find fixed-length patterns.
 - **mSTAMP**: The multivariate motif discovery algorithm to find fixed-length patterns.
 - **Exact**: The naive version algorithm.
 - **MultiPAL**: The MultiPAL algorithm can discover multi-length patterns in multiple time series.
 - **Rep-MultiPAL**: The **improved** version of MultiPAL algorithm.

Sanity Check



Rep-MultiPAL



MultiPAL



Performance Evaluation

Adjusted Rand Index (ARI)

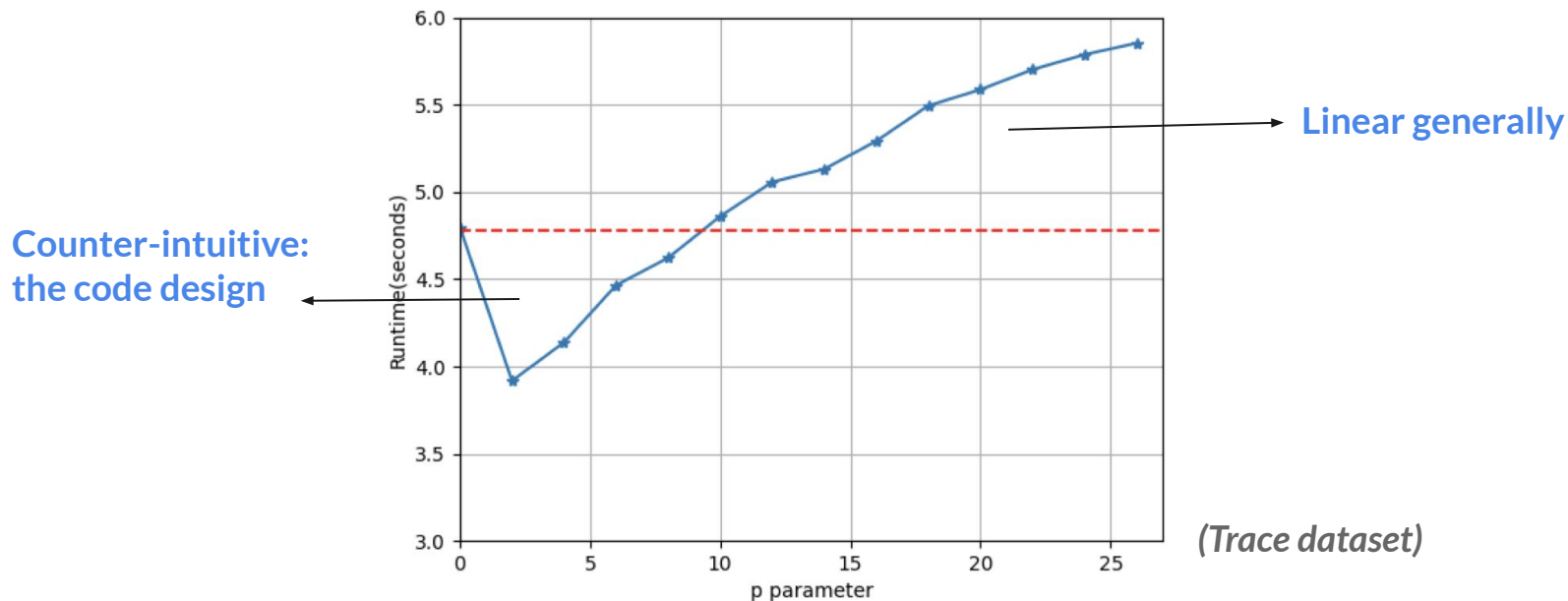
	Rep-MultiPAL	MultiPAL	Exact	mSTAMP	SUSh
Sinusoidal	0.97(12)	0.97(12)	1.0(12)	0.64(24)	0.73(6)
Birds	1.0(15)	0.93(15)	1.0(15)	0.61(24)	0.79(8)
Trace	0.94(200)	0.88(200)	-	0.63(400)	0.68(189)
PAMAP	0.94(120)	0.90(120)	-	0.73(240)	0.81(105)



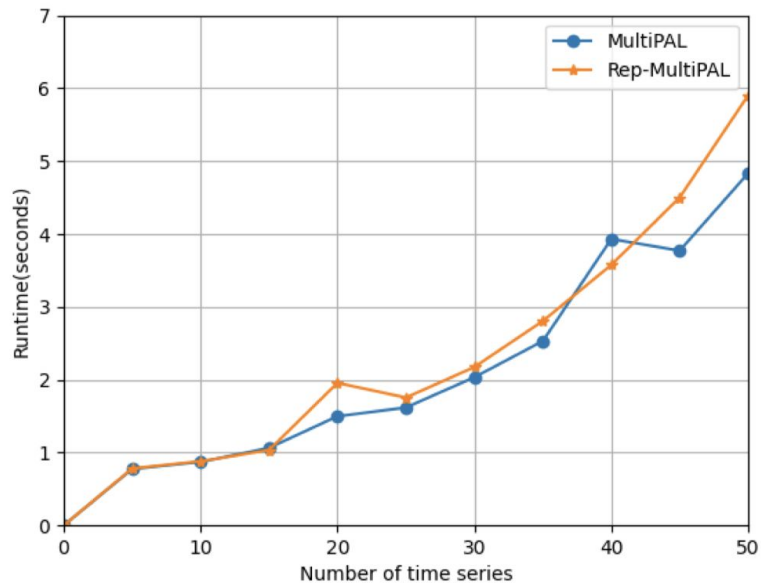
Performance Evaluation

	DCO		1-DDTW		IPD	
	MultiPAL	Rep-MultiPAL	MultiPAL	Rep-MultiPAL	MultiPAL	Rep-MultiPAL
Sinusoidal	0.5890	0.5643	0.0979	0.0875	0.3603	0.3959
Birds	0.5378	0.5470	0.0141	0.0081	0.2627	0.3040
Trace	0.6613	0.6645	0.0019	0.0018	0.0574	0.0252
PAMAP	0.5930	0.6349	0.1177	0.0970	0.0459	0.0685

Parameter Sensitivity



Scalability



(Trace dataset)



Case Study - Real World Datasets

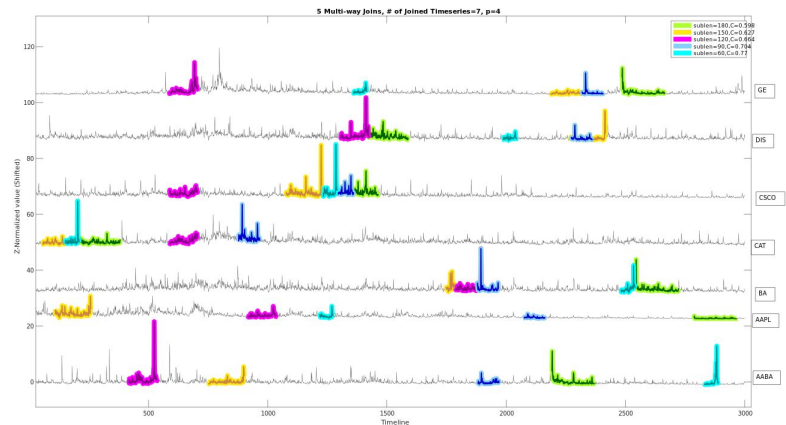
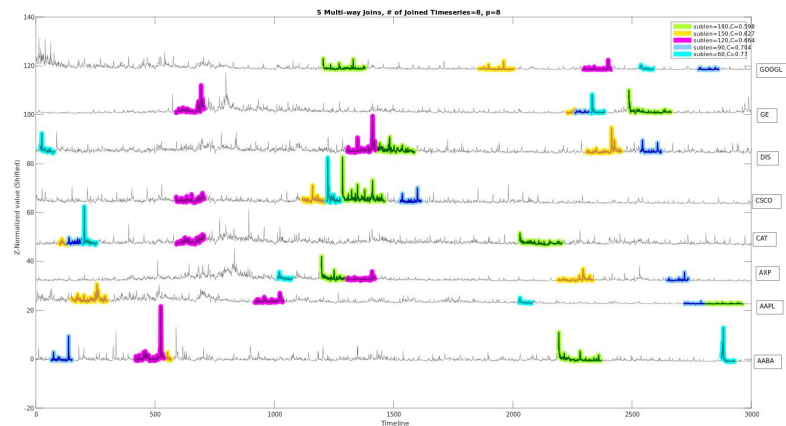
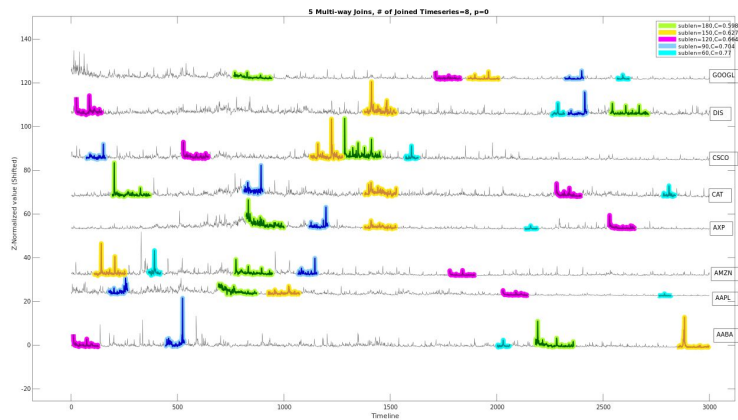
Large Company
Stock Dataset

Reported Death of
COVID19 Dataset

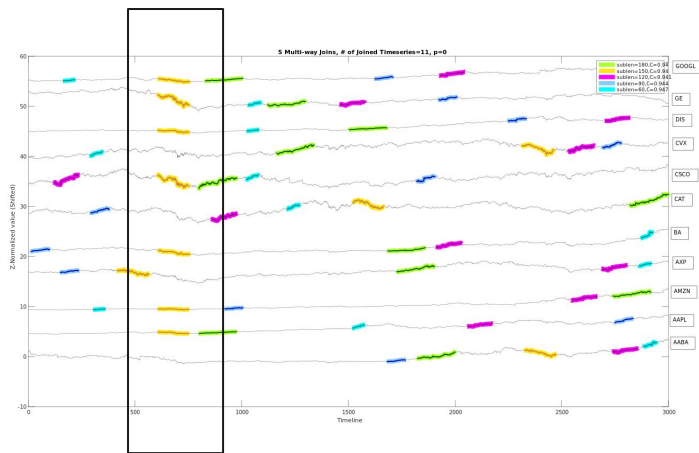
MultiPAL

REP-MultiPAL

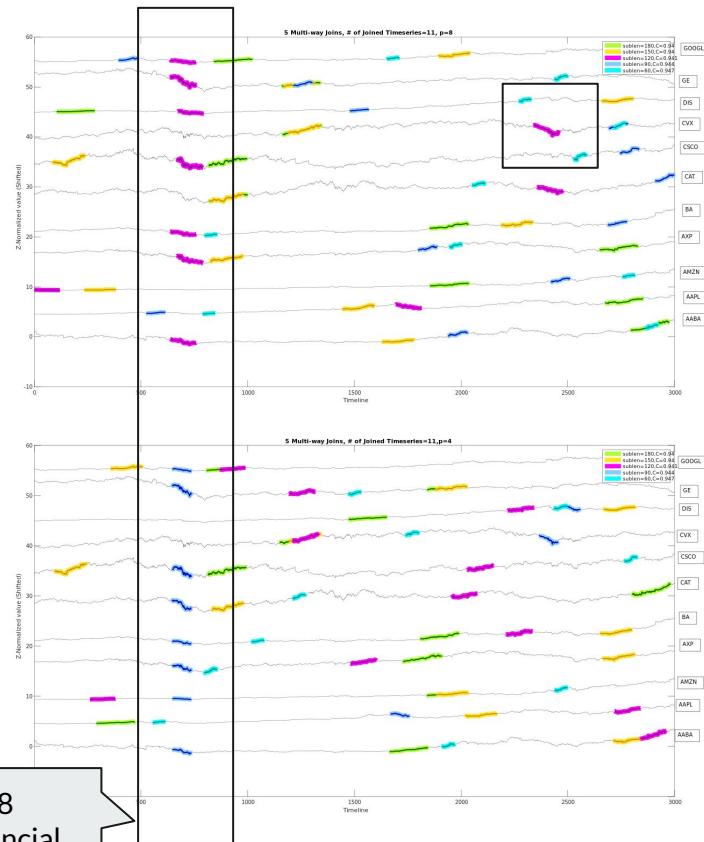
Case Study - Trading Volume of Stock



Case Study - Closing Price of Stock



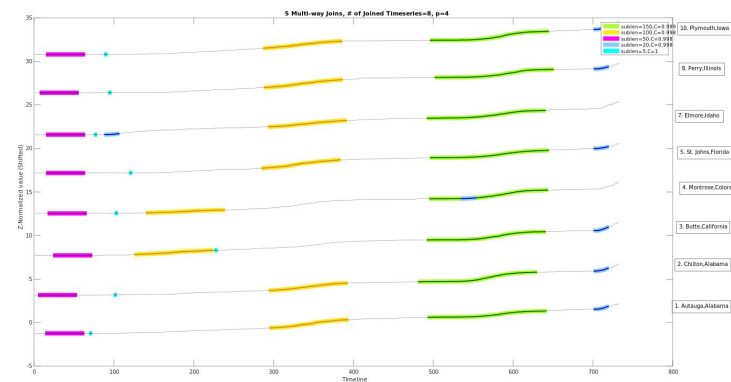
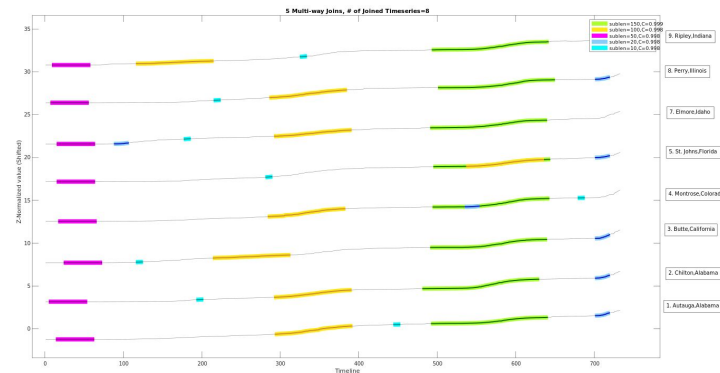
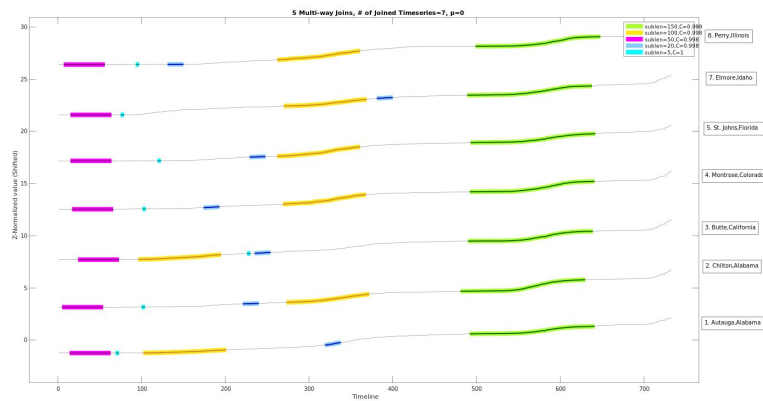
MultiPAL



2008
Financial
Crisis

REP-MultiPAL

Case Study - Reported Death with COVID-19





Conclusion

- Propose a novel multi-way time series join algorithm, Rep-MultiPAL
- Evaluate the performance difference between Rep-MultiPAL and other baseline algorithms
- Most of the experiment results indicate that Rep-MultiPAL can achieve a better performance than MultiPAL with a little decrease in time efficiency.
- Further explore some real-world time series datasets