LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND STATISTIK
INSTITUT FÜR INFORMATIK
LEHRSTUHL FÜR DATENBANKSYSTEME UND DATA MINING

Inferring Process Performance Models from Interval Events using the Performance Skyline

Master Thesis, Munich Germany

Andrea Maldonado

30th March 2020

DBS

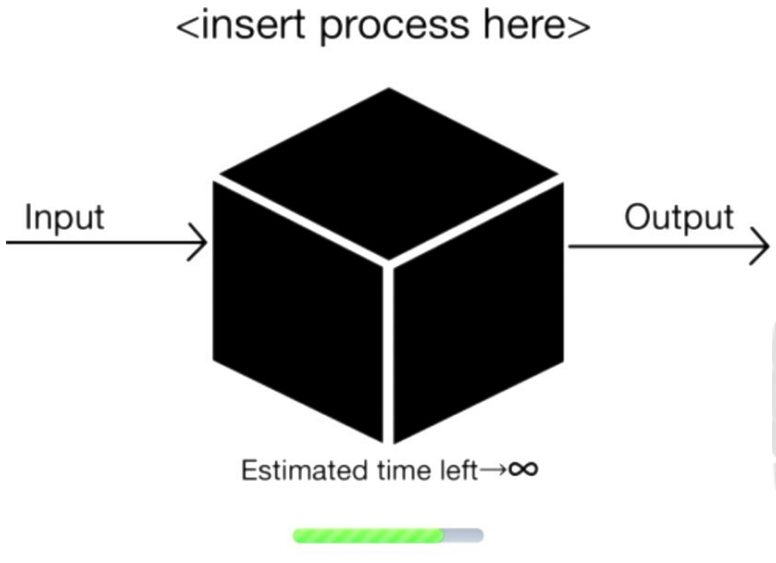
academic advisor:
Janina Sontheim
academic supervisor:
Prof. Dr. Thomas Seidl

Agenda

- Introduction
- Background Knowledge and Related Work
- Performance Models Contribution
  - Interval Events
  - Performance Skyline
- Experiments
  - Dataset Description
  - Anomaly Detection
- Final Thoughts and Conclusion

Introduction: Motivation

- Process Enhancement requires understanding the process
- Deviation detection of described vs. prescribed processes
- Business processes and information systems' alignment requires continuous attention
- Processes are constantly changing



Introduction: Challenge

How to extract knowledge and model performance of processes containing (interval) events?

Related Work

Related Work: Process Mining

```
$ cat nfs/processmining/pm4pyexample/running-example-just-two-cases.csv
Case ID;Event ID;dd-MM-yyyy:HH.mm;Activity;Resource;Costs
1;35654423;30-12-2010:11.02;register request;Pete;50
1;35654424;31-12-2010:10.06;examine thoroughly;Sue;400
1;35654425;05-01-2011:15.12;check ticket;Mike;100
1;35654426;06-01-2011:11.18;decide;Sara;200
1;35654427;07-01-2011:14.24;reject request;Pete;200
4;35654641;06-01-2011:15.02;register request;Pete;50
4;35654643;07-01-2011:12.06;check ticket;Mike;100
4;35654644;08-01-2011:14.43;examine thoroughly;Sean;400
4;35654645;09-01-2011:12.02;decide;Sara;200
4;35654647;12-01-2011:15.44;reject request;Ellen;200
```

Related Work: Process Mining

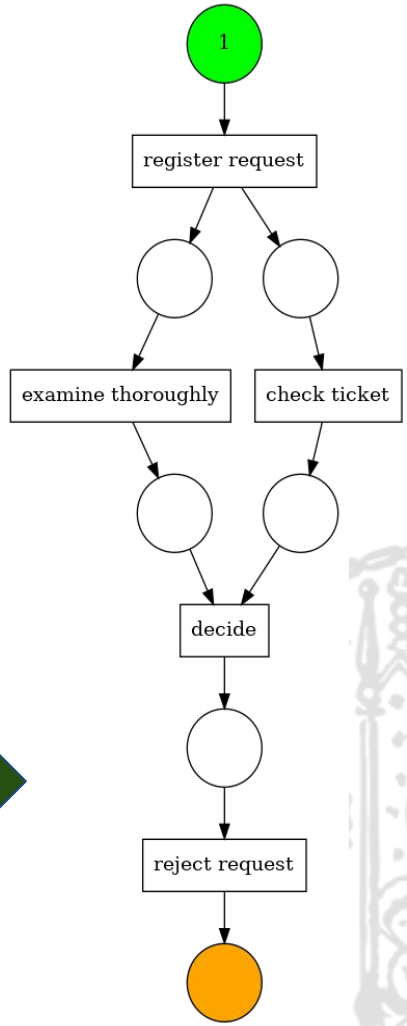
```
$ cat nfs/processmining/pm4pyexample/running-example-just-two-cases.csv
Case ID;Event ID;dd-MM-yyyy:HH.mm;Activity;Resource;Costs
1;35654423;30-12-2010:11.02;register request;Pete;50
1;35654424;31-12-2010:10.06;examine thoroughly;Sue;400
1;35654425;05-01-2011:15.12;check ticket;Mike;100
1;35654426;06-01-2011:11.18;decide;Sara;200
1;35654427;07-01-2011:14.24;reject request;Pete;200
4;35654641;06-01-2011:15.02;register request;Pete;50
4;35654643;07-01-2011:12.06;check ticket;Mike;100
4;35654644;08-01-2011:14.43;examine thoroughly;Sean;400
4;35654645;09-01-2011:12.02;decide;Sara;200
4;35654647;12-01-2011:15.44;reject request;Ellen;200
```

Case id	Activity id	Timestamp
1	register request	30-12-2010:11.02
1	examine thoroughly	31-12-2010:10.06
1	check ticket	05-01-2011:15.12
1	decide	06-01-2011:11.18
1	reject request	07-01-2011:14.24
⋮	⋮	⋮

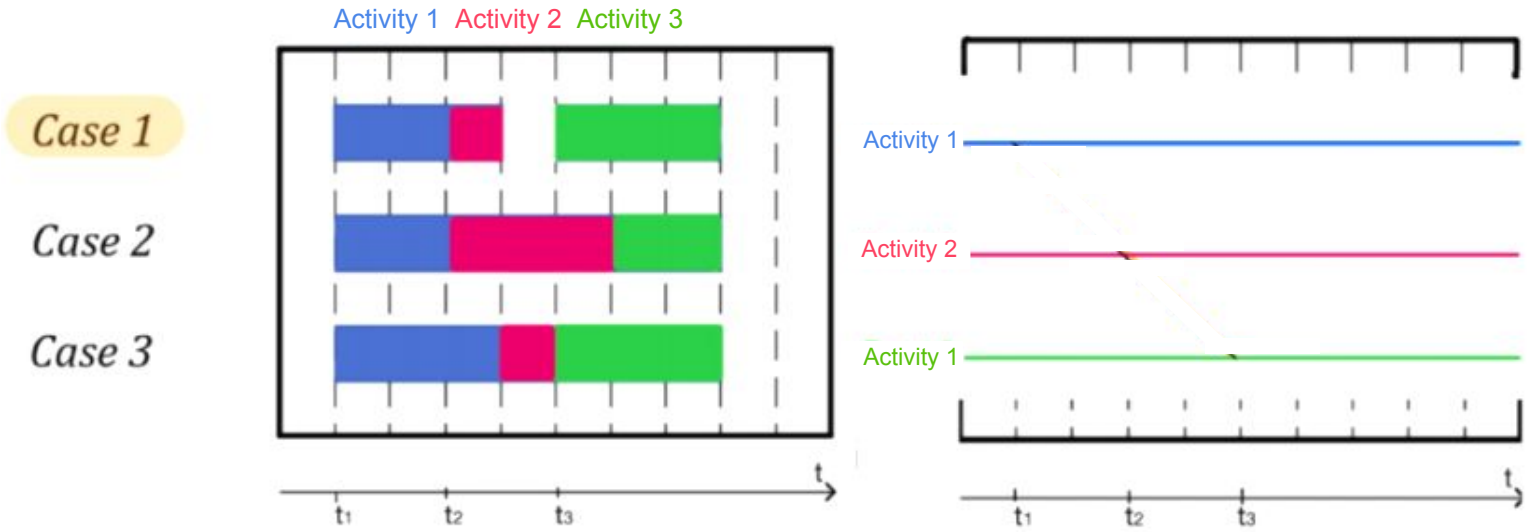
Related Work: Process Mining

```
$ cat nfs/processmining/pm4pyexample/running-example-just-two-cases.csv
Case ID;Event ID;dd-MM-yyyy:HH.mm;Activity;Resource;Costs
1;35654423;30-12-2010:11.02;register request;Pete;50
1;35654424;31-12-2010:10.06;examine thoroughly;Sue;400
1;35654425;05-01-2011:15.12;check ticket;Mike;100
1;35654426;06-01-2011:11.18;decide;Sara;200
1;35654427;07-01-2011:14.24;reject request;Pete;200
4;35654641;06-01-2011:15.02;register request;Pete;50
4;35654643;07-01-2011:12.06;check ticket;Mike;100
4;35654644;08-01-2011:14.43;examine thoroughly;Sean;400
4;35654645;09-01-2011:12.02;decide;Sara;200
4;35654647;12-01-2011:15.44;reject request;Ellen;200
```

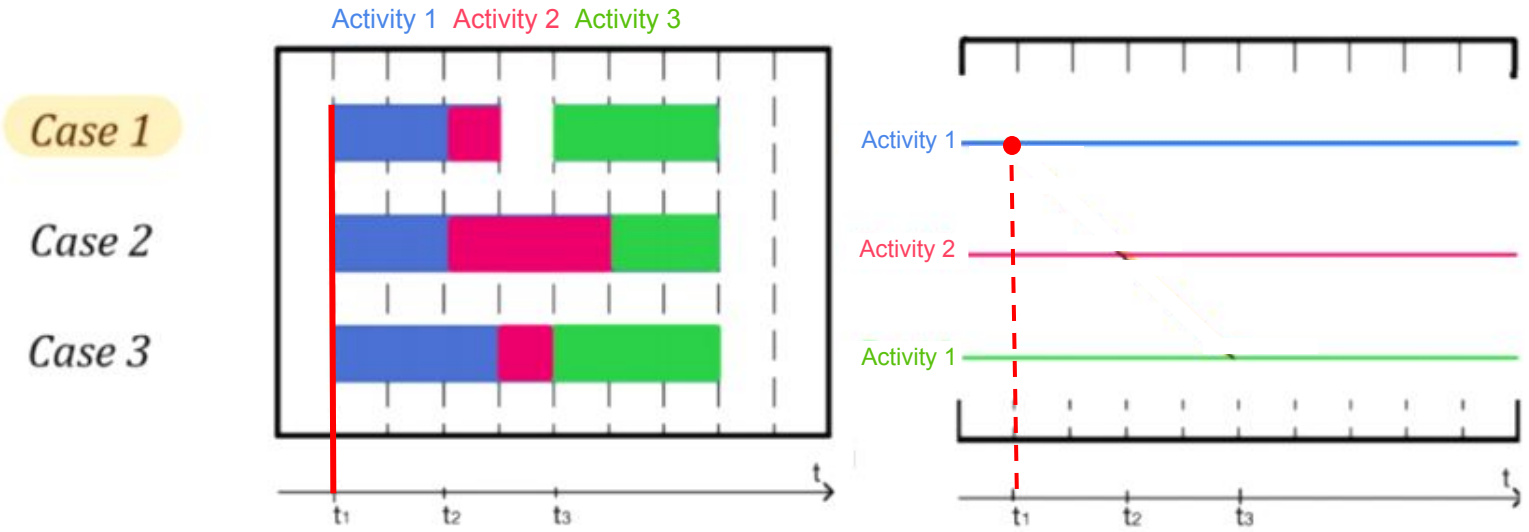
Case id	Activity id	Timestamp
1	register request	30-12-2010:11.02
1	examine thoroughly	31-12-2010:10.06
1	check ticket	05-01-2011:15.12
1	decide	06-01-2011:11.18
1	reject request	07-01-2011:14.24
⋮	⋮	⋮



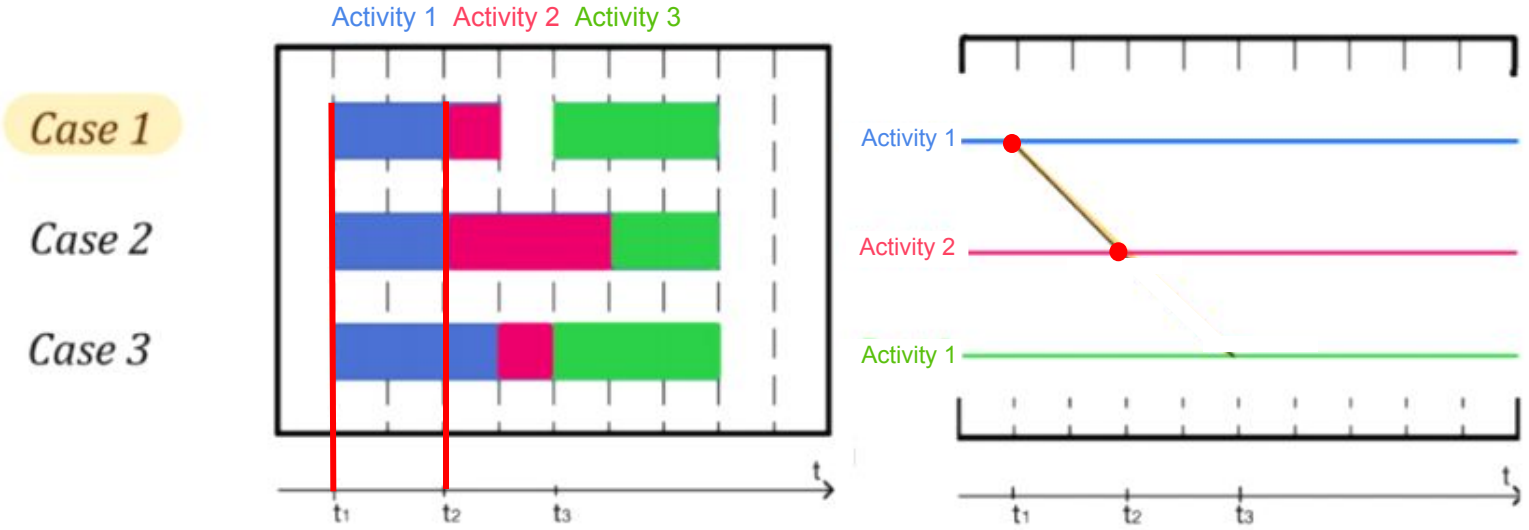
Related Work: Performance Spectrum



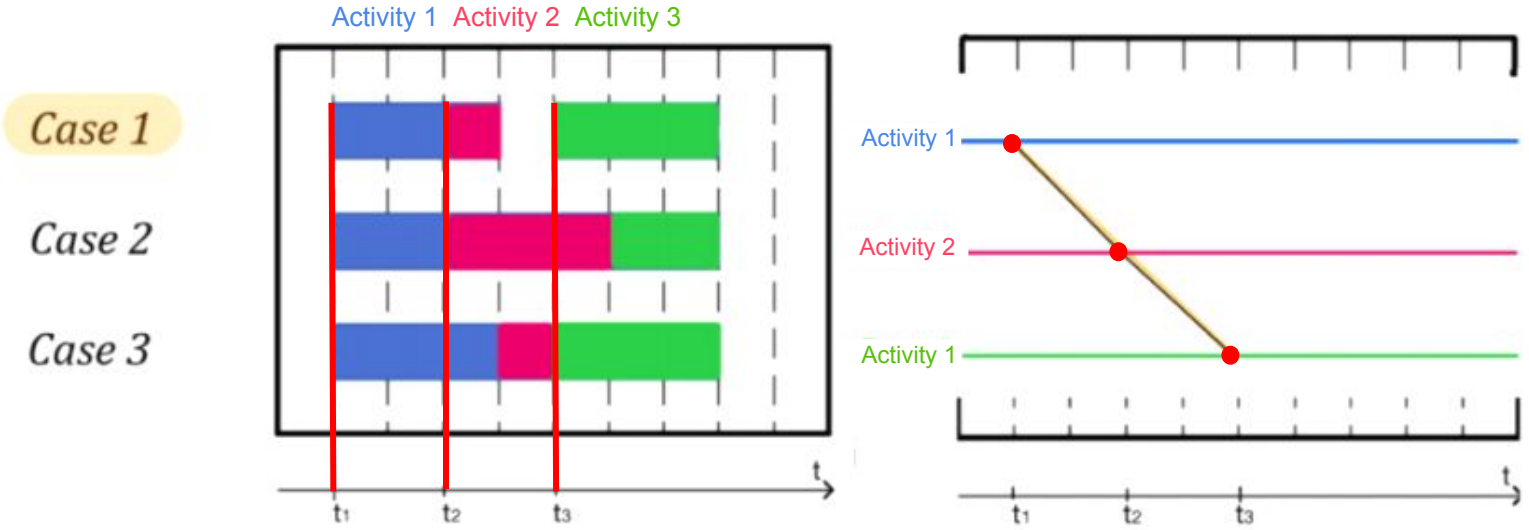
Related Work: Performance Spectrum



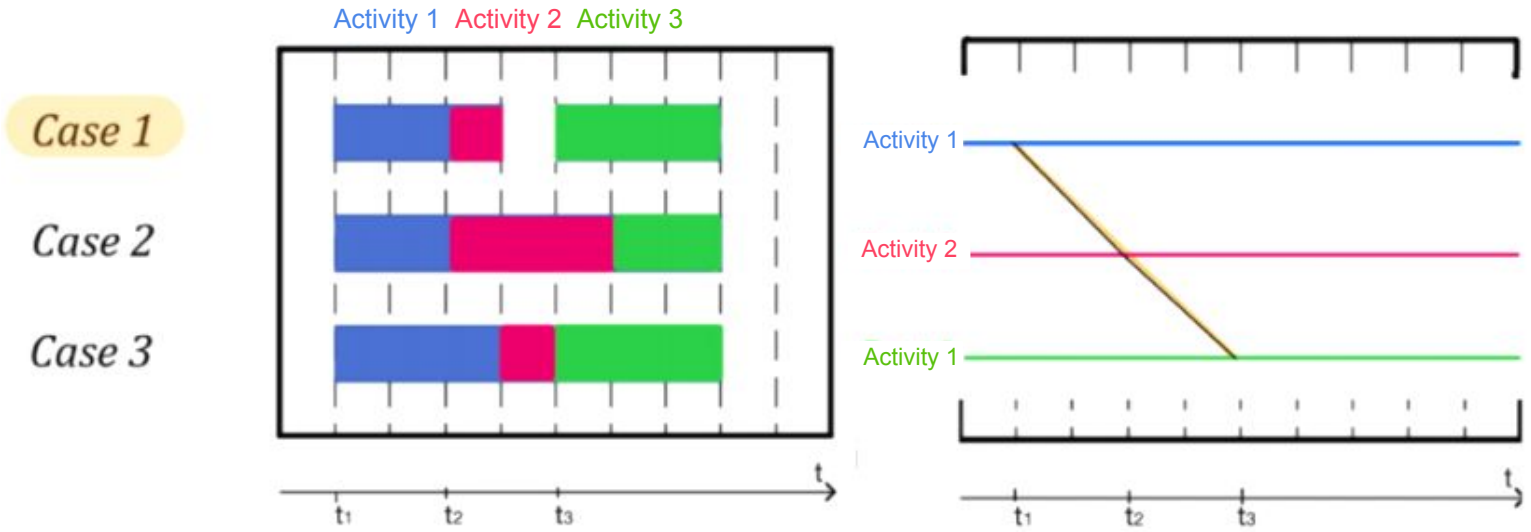
Related Work: Performance Spectrum



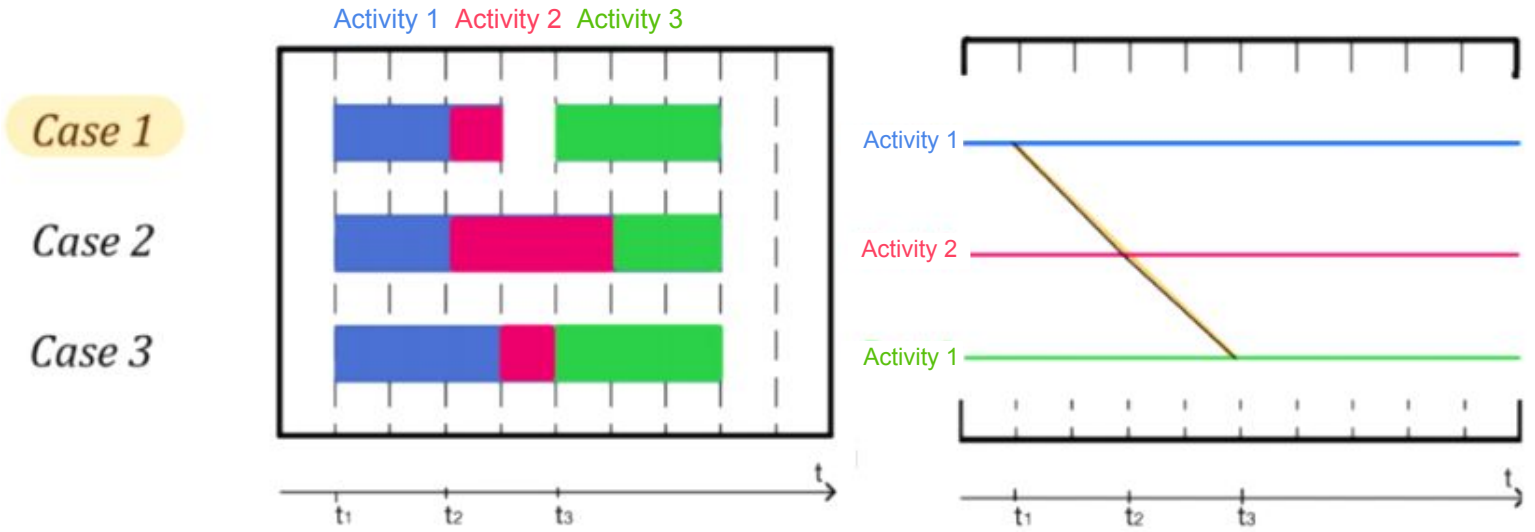
Related Work: Performance Spectrum

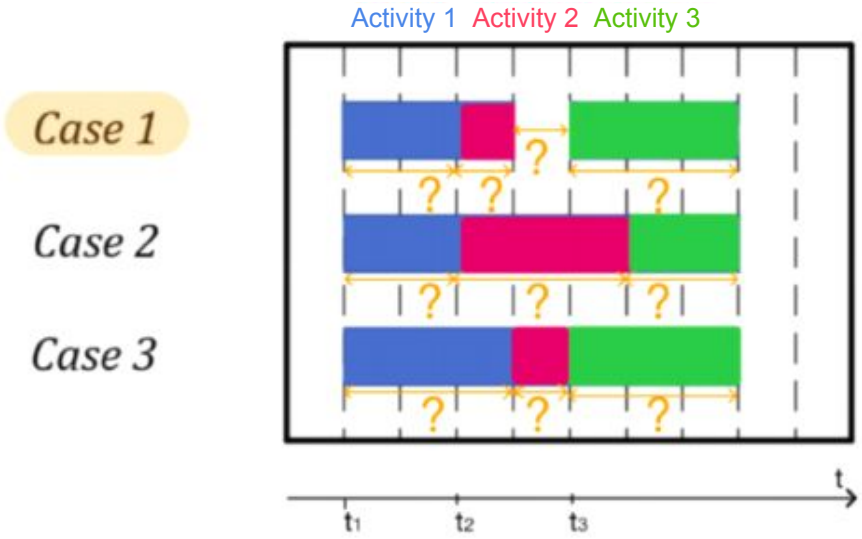


Related Work: Performance Spectrum



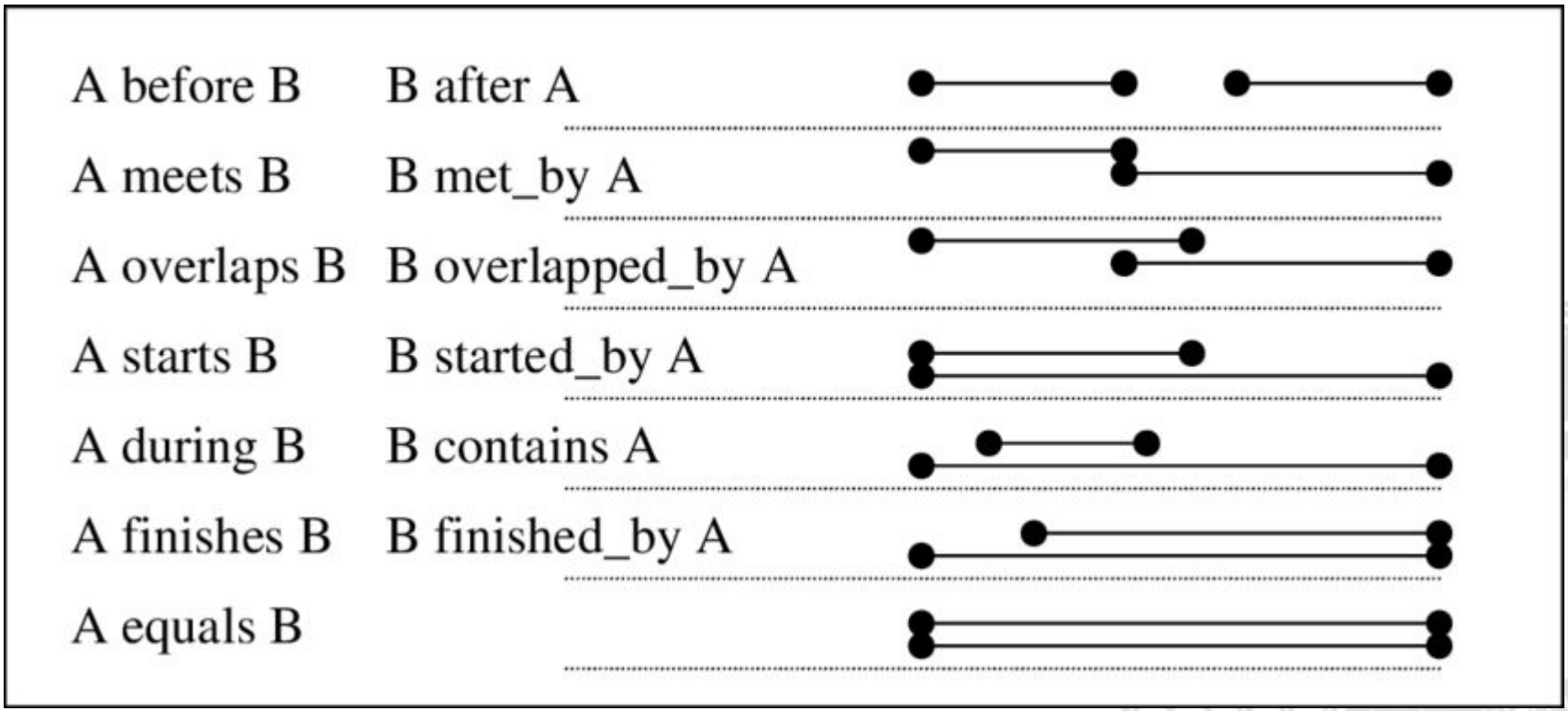
Related Work: Performance Spectrum



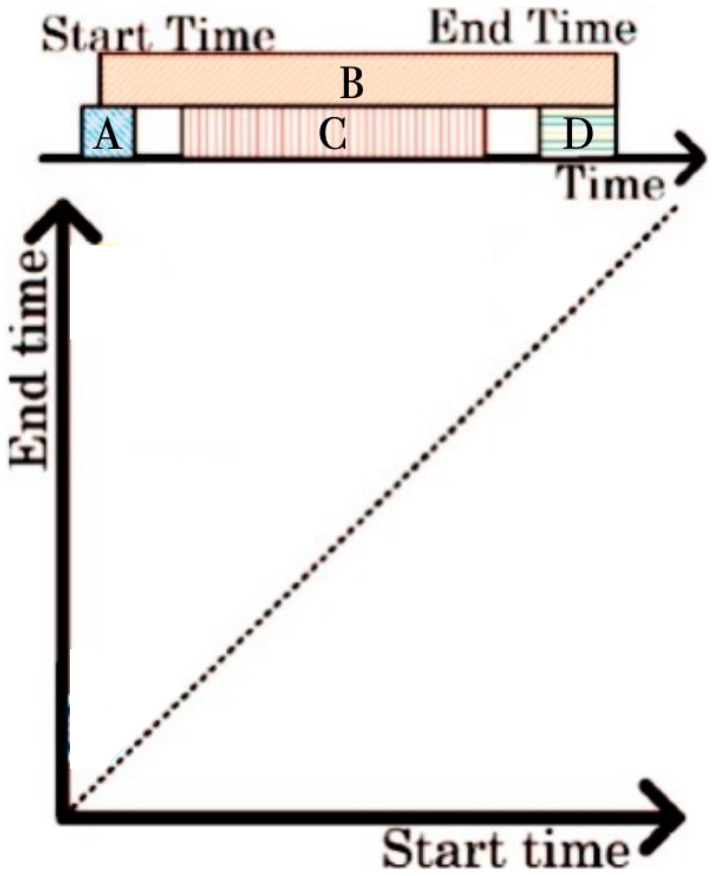


Related Work: Allen's Interval Algebra

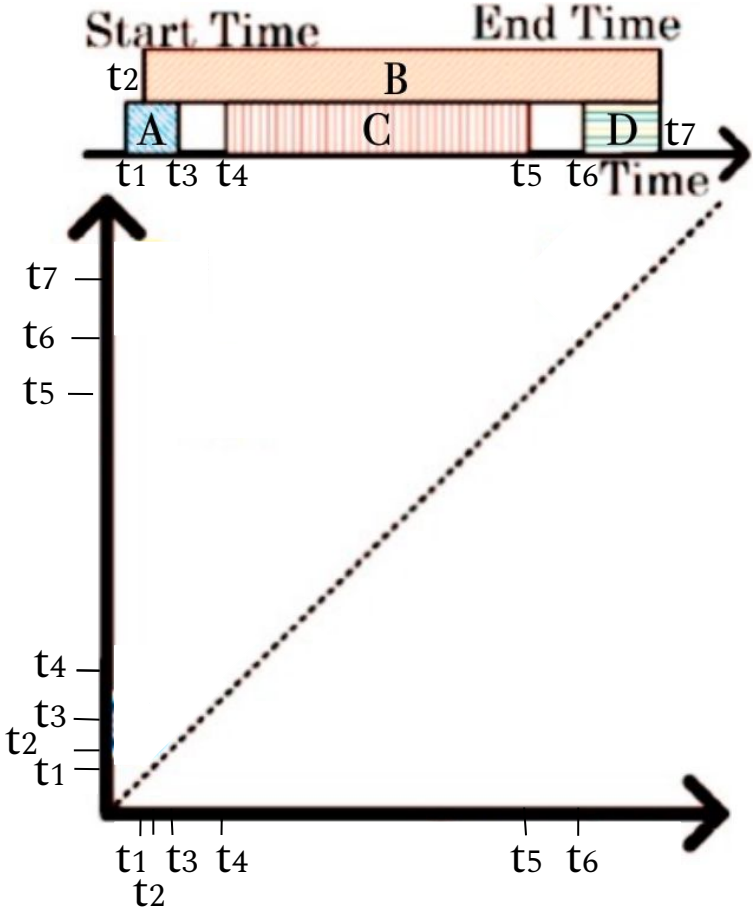
- Deals with temporal intervals *{A},{B}*:
  with starting at *{A+},{B+}* and ending at *{A-},{B-}*



Related Work: Sequence Pattern Mining

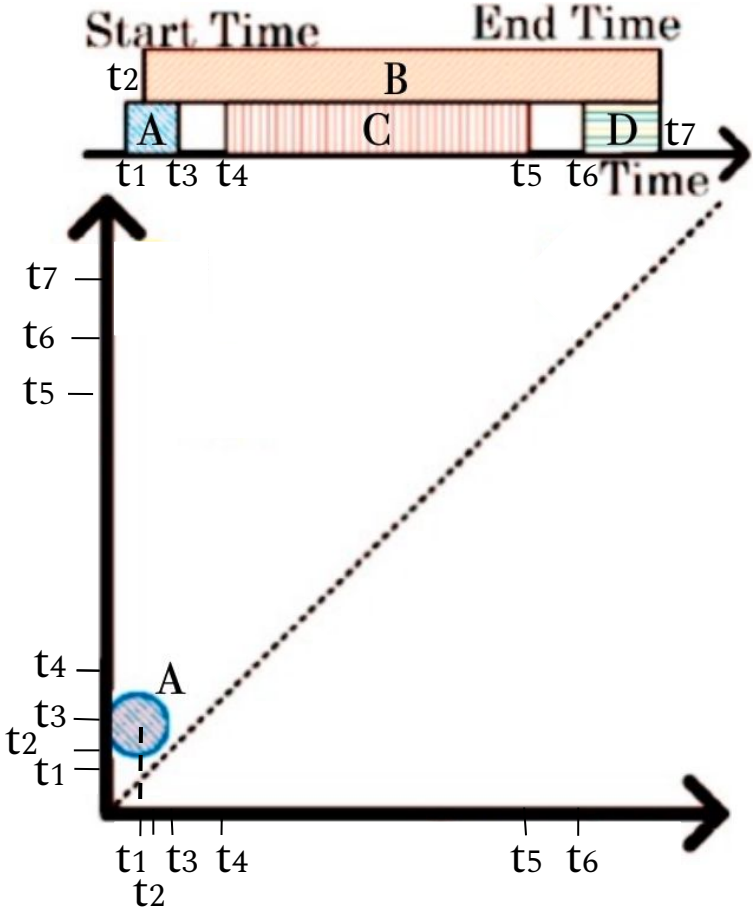


Related Work: Sequence Pattern Mining



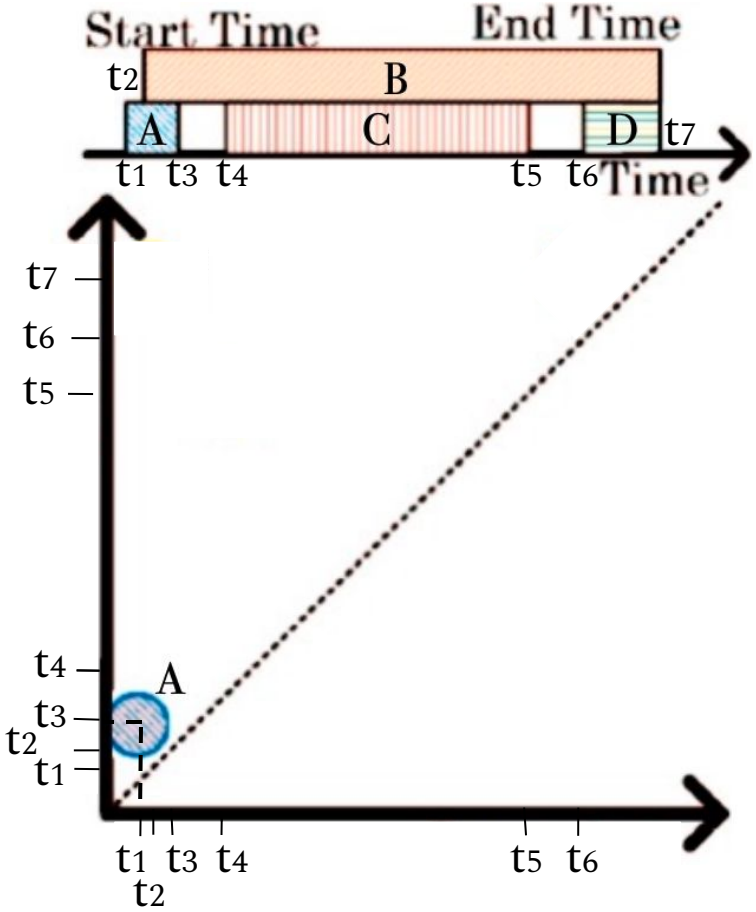
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3

Related Work: Sequence Pattern Mining



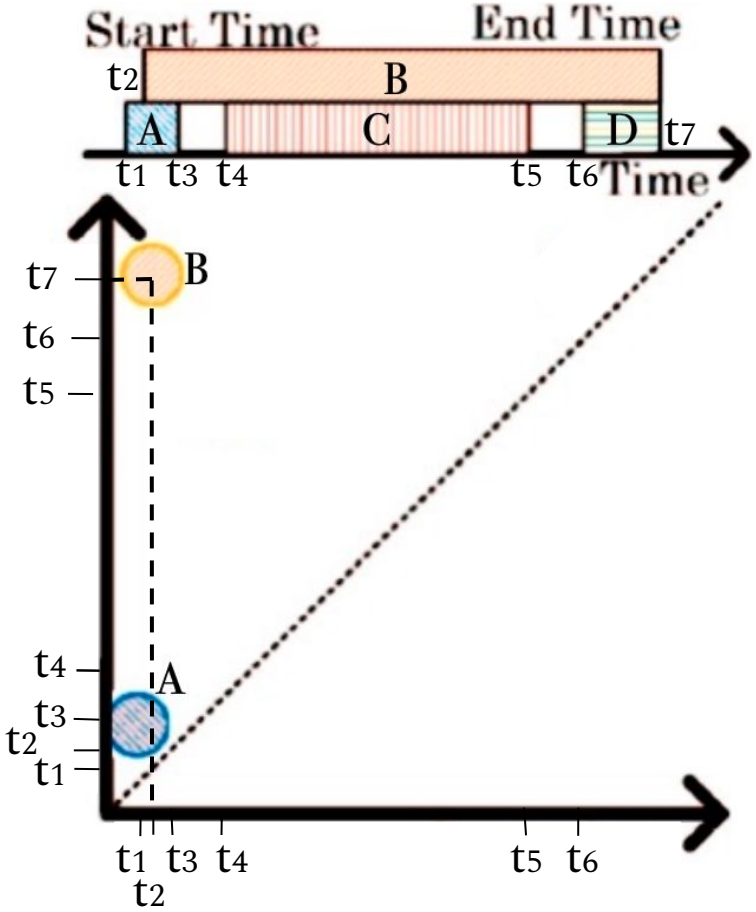
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3

Related Work: Sequence Pattern Mining



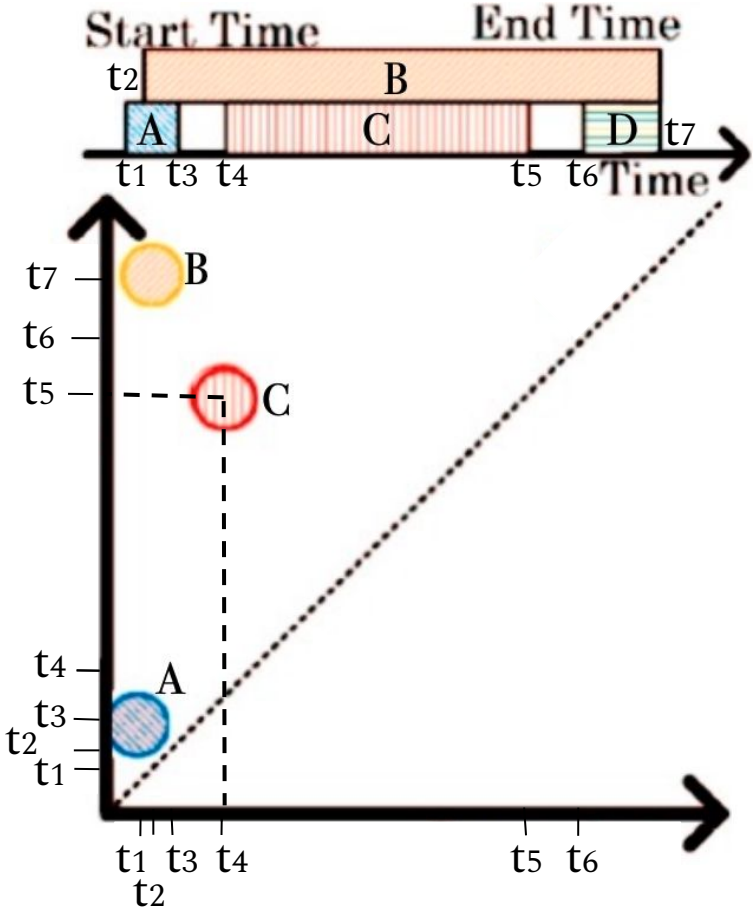
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3

Related Work: Sequence Pattern Mining



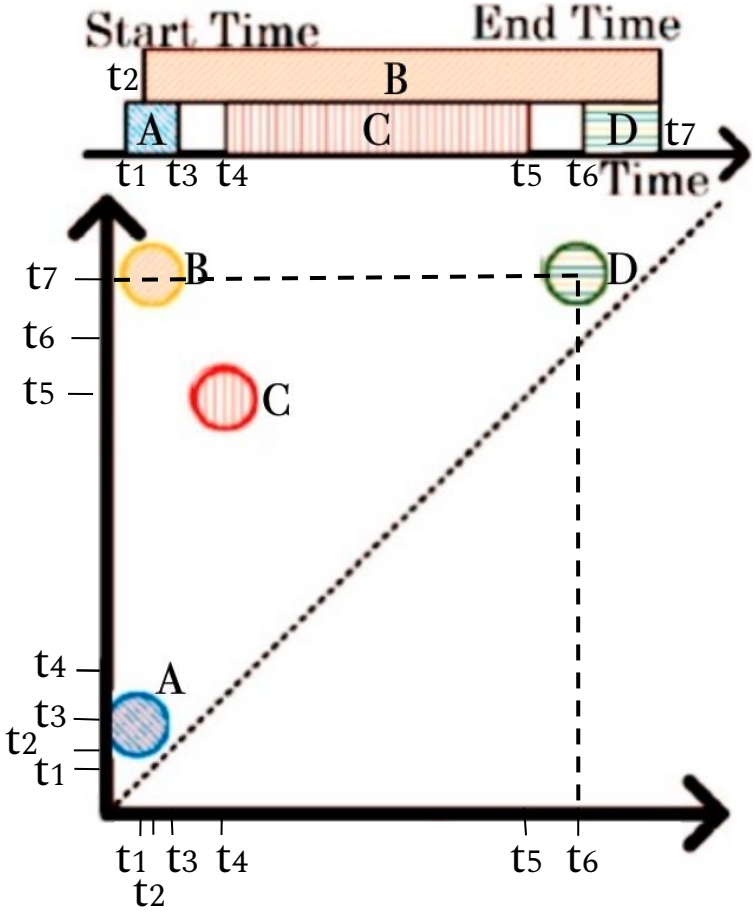
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7

Related Work: Sequence Pattern Mining



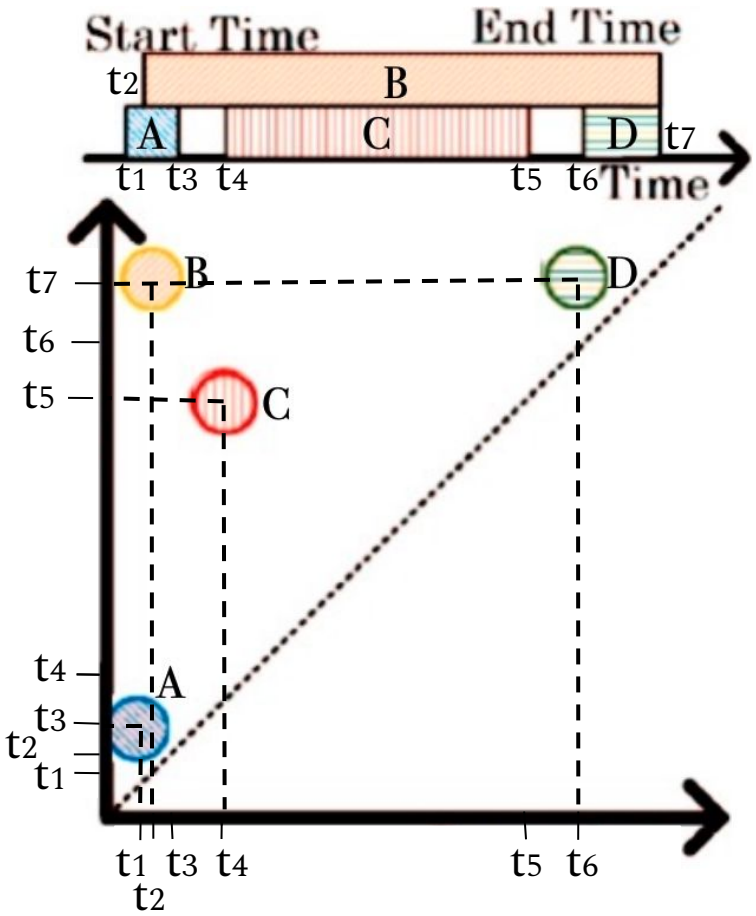
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7
  {C+}=t4; {C-}=t5

Related Work: Sequence Pattern Mining



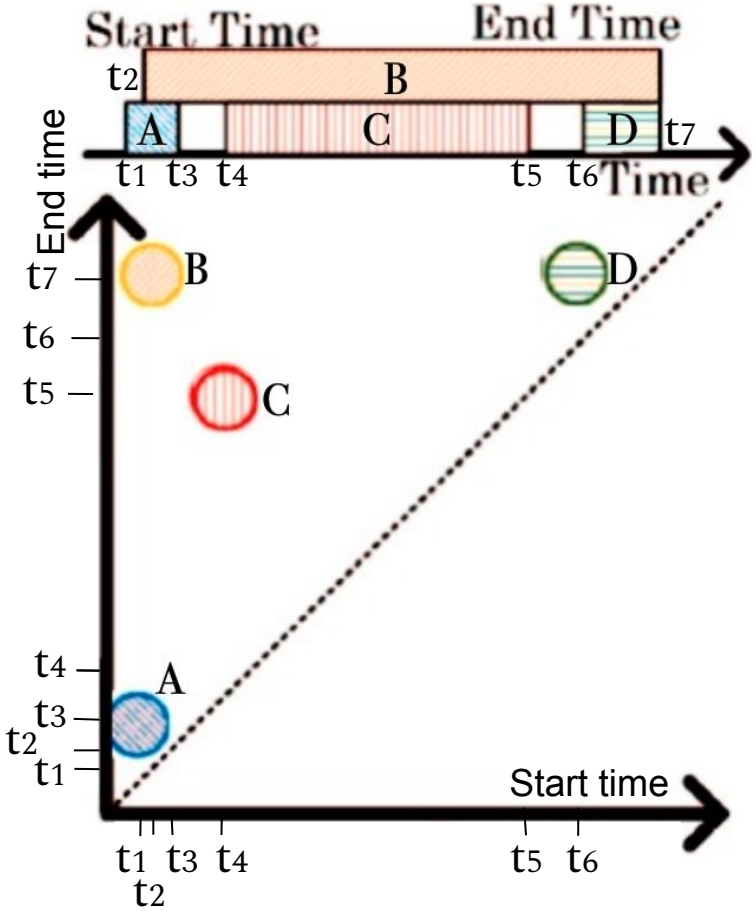
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7
  {C+}=t4; {C-}=t5
  {D+}=t6; {D-}=t7

Related Work: Sequence Pattern Mining



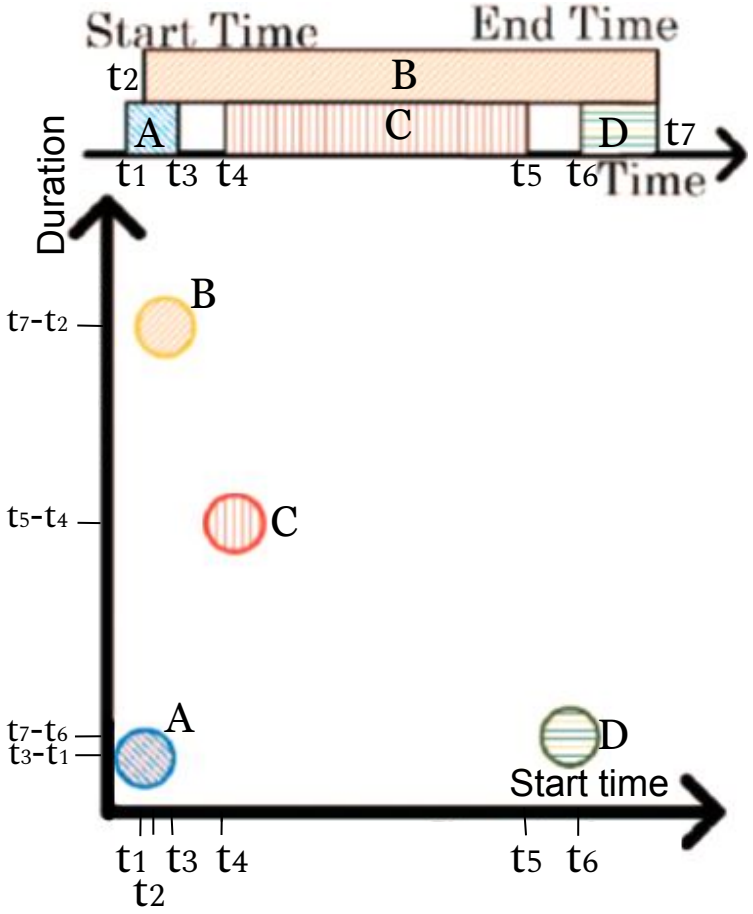
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7
  {C+}=t4; {C-}=t5
  {D+}=t6; {D-}=t7

Related Work: Sequence Pattern Mining



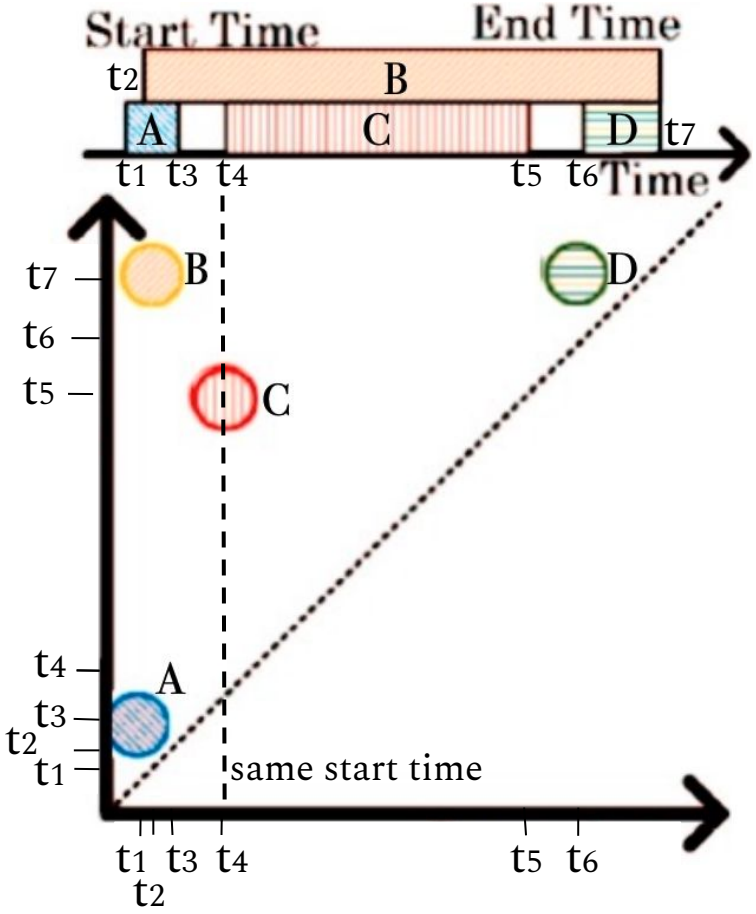
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7
  {C+}=t4; {C-}=t5
  {D+}=t6; {D-}=t7

Related Work: Sequence Pattern Mining



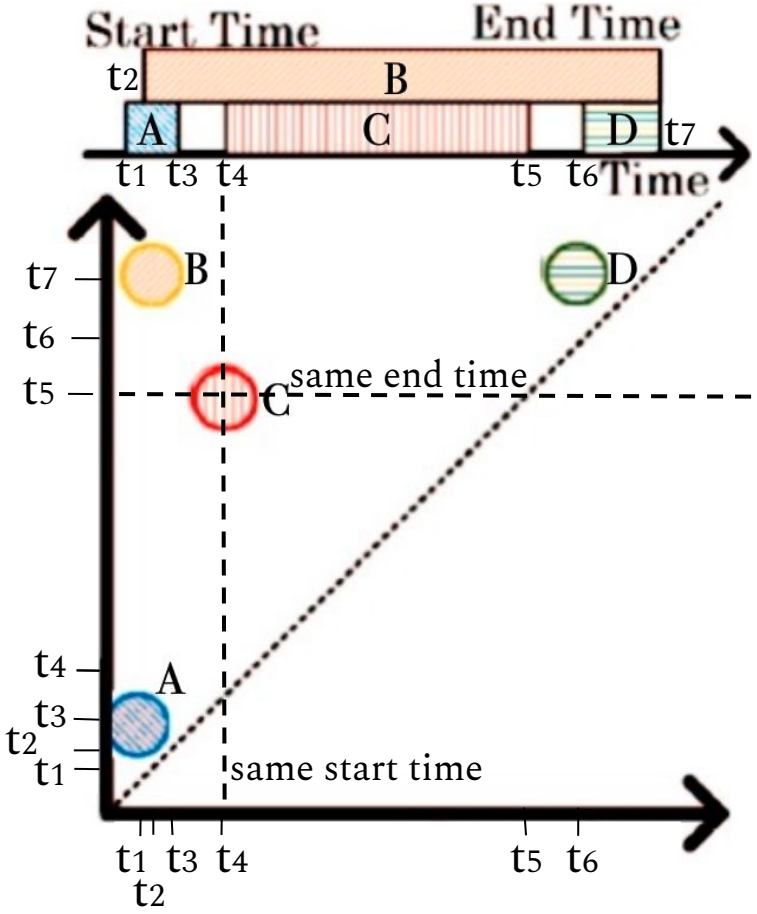
- Interval Geometric Representation:
  {A+}=t1; {A-}=t3
  {B+}=t2; {B-}=t7
  {C+}=t4; {C-}=t5
  {D+}=t6; {D-}=t7

Related Work: Sequence Pattern Mining



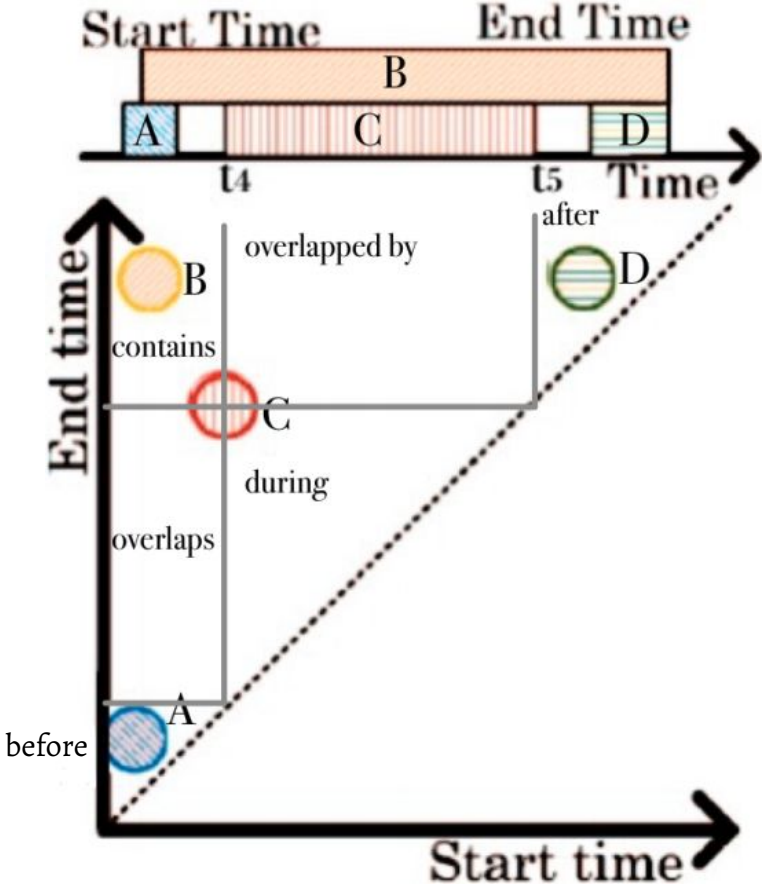
- Interval Geometric Representation for **C**: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra

Related Work: Sequence Pattern Mining



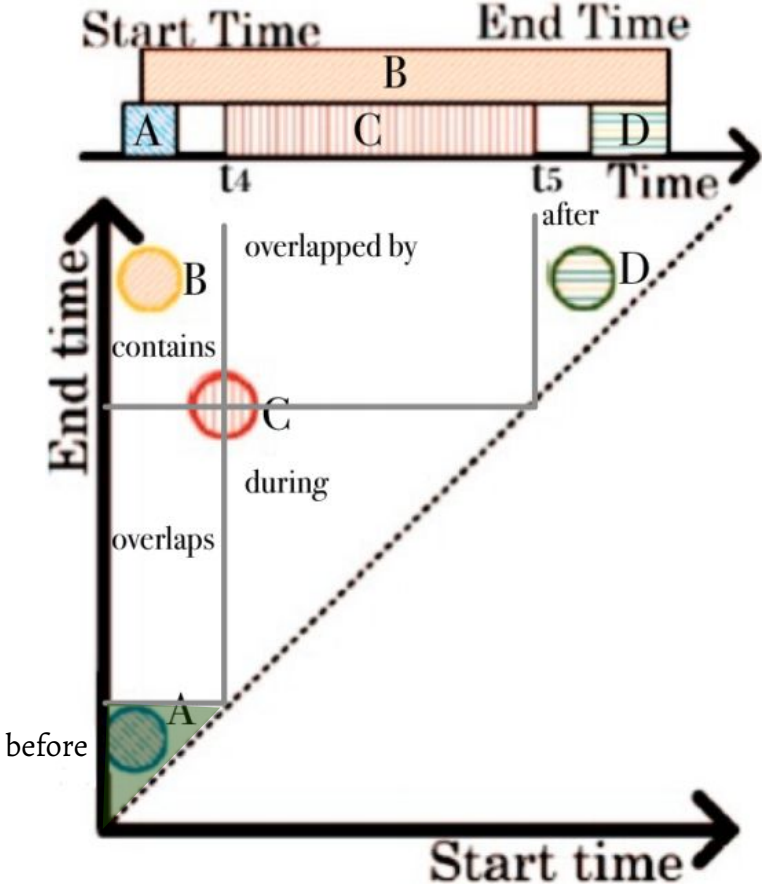
- Interval Geometric Representation for **C**: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra

Related Work: Sequence Pattern Mining



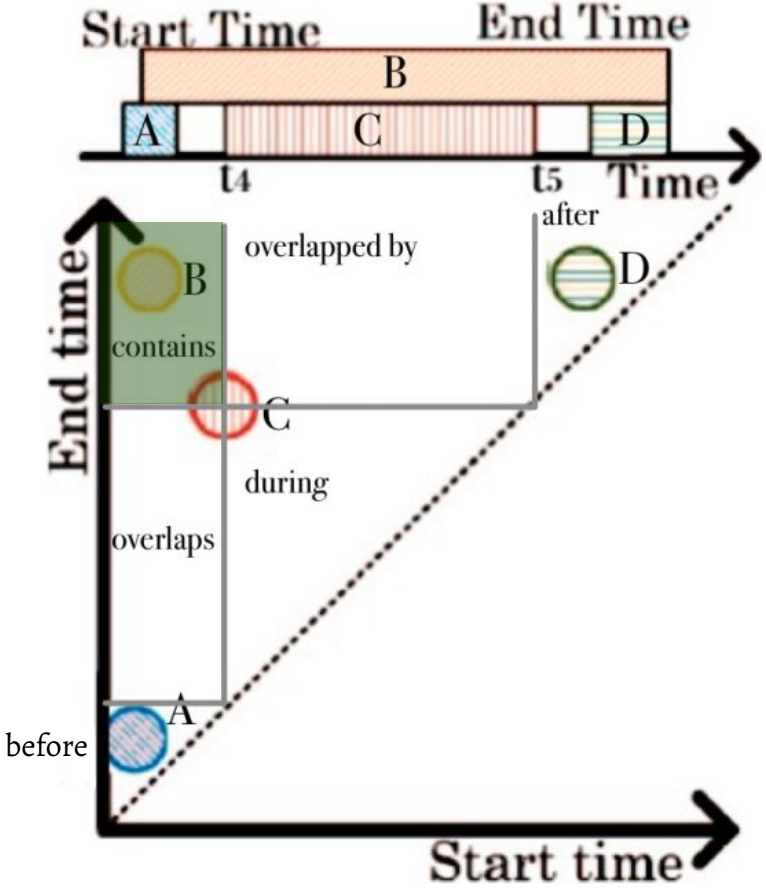
- Interval Geometric Representation for *C*: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra

Related Work: Sequence Pattern Mining



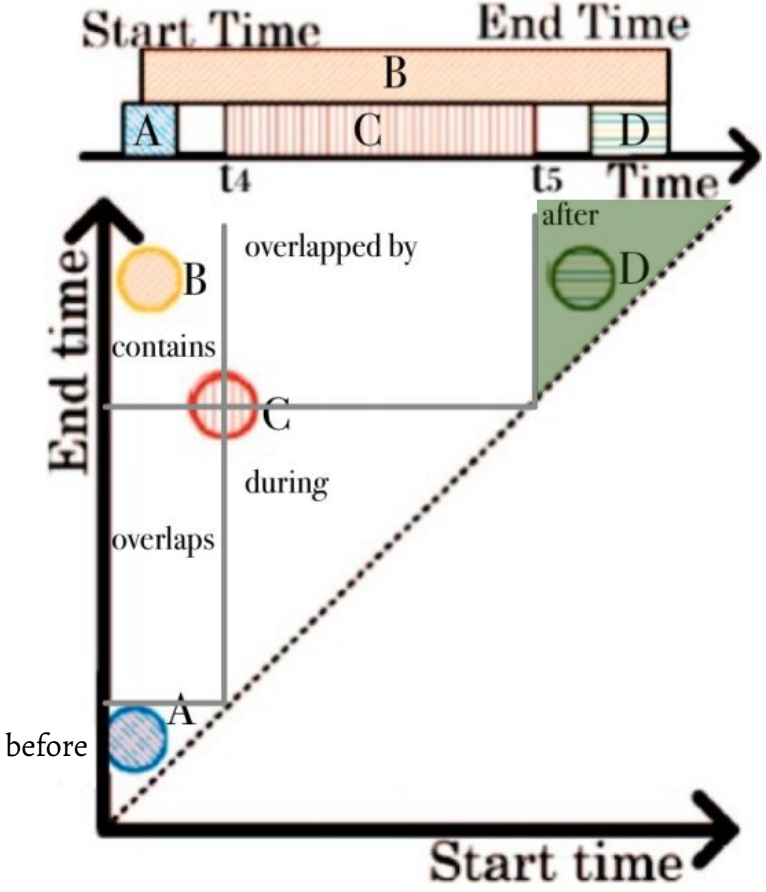
- Interval Geometric Representation for *C*: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra
  - *A* before *C*

Related Work: Sequence Pattern Mining



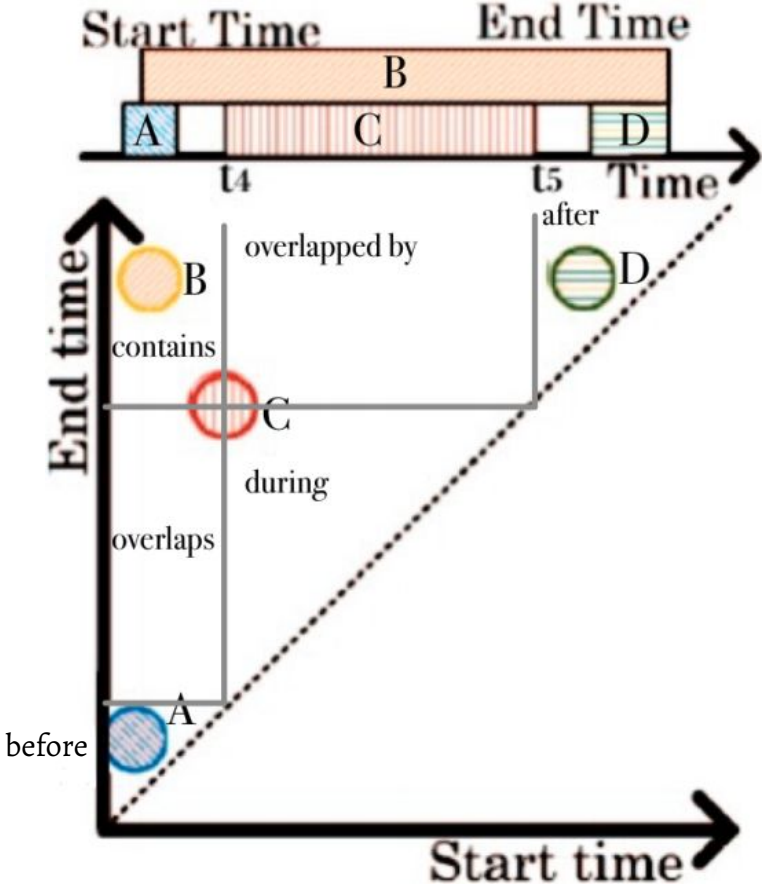
- Interval Geometric Representation for *C*: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra
  - *A* before *C*
  - *B* contains *C*

Related Work: Sequence Pattern Mining



- Interval Geometric Representation for *C*: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra
  - *A* before *C*
  - *B* contains *C*
  - *D* after *C*

Related Work: Sequence Pattern Mining



- Interval Geometric Representation for *C*: *{C+}=t4; {C-}=t5*
- Allen's Interval Algebra
  - *A* before *C*
  - *B* contains *C*
  - *D* after *C*

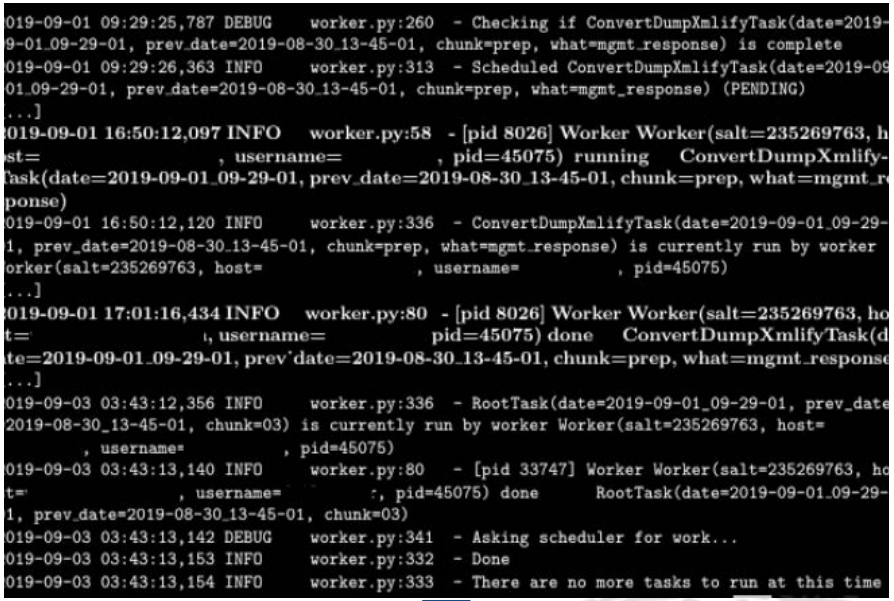
Performance Models Contribution

- Interval Events
- Performance Skyline

Interval Events

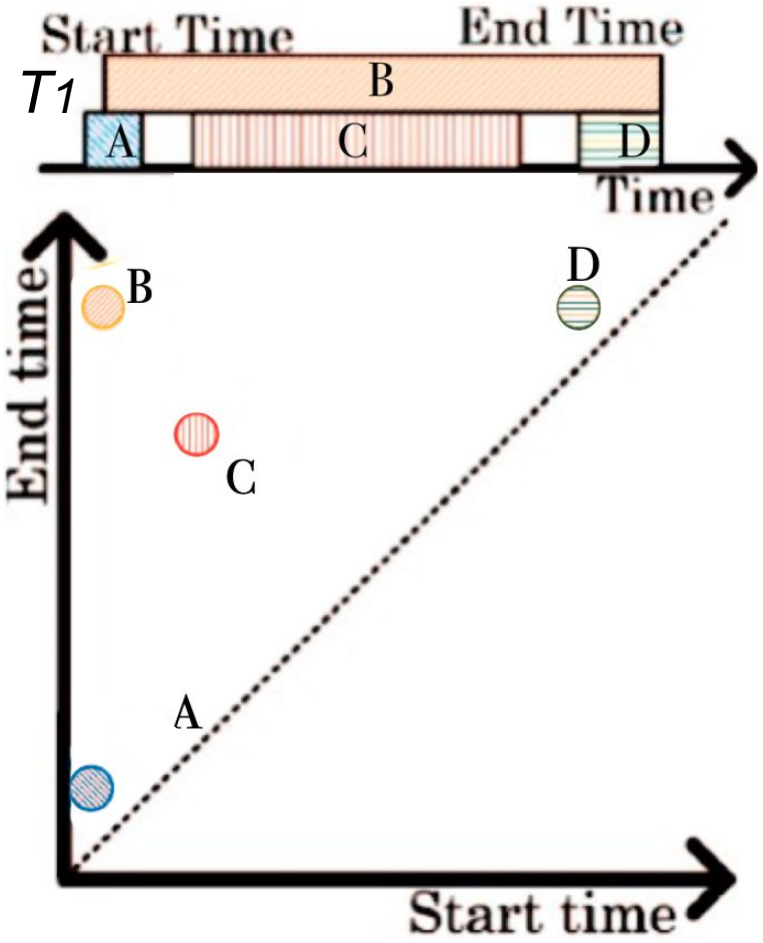
$$e = (c, a, t^+, t^-) \epsilon \mathbb{N} \times A \times \mathbb{N} \times \mathbb{N}$$

- Case id
- Activity id
- Timestamp (start)
- Timestamp (end)



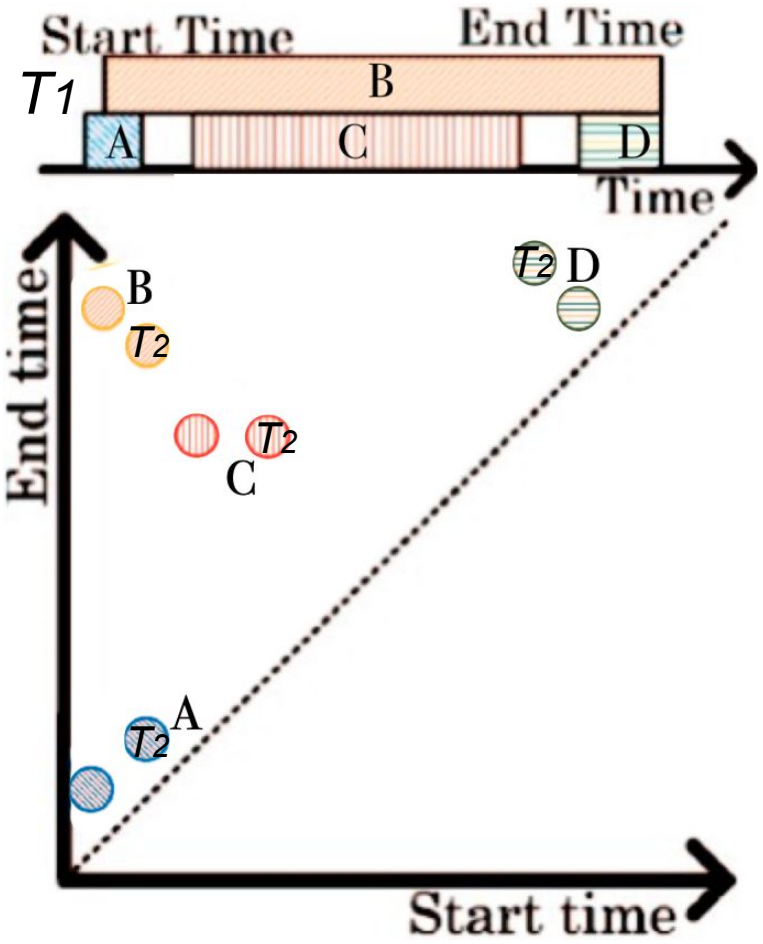
case id	activity id	start time	end time
example-process. 2019-09-01	ConvertDumpXmlifyTask (what=mgmt_response)	2019-09-01 16:50:12	2019-09-01 17:01:16
example-process. 2019-09-01	ConvertDumpXmlifyTask (what=source)	2019-09-01 14:13:56	2019-09-01 14:17:03
example-process. 2019-09-03	ConvertDumpXmlifyTask (what=source)	2019-09-03 07:33:56	2019-09-03 07:37:24

Performance Skyline



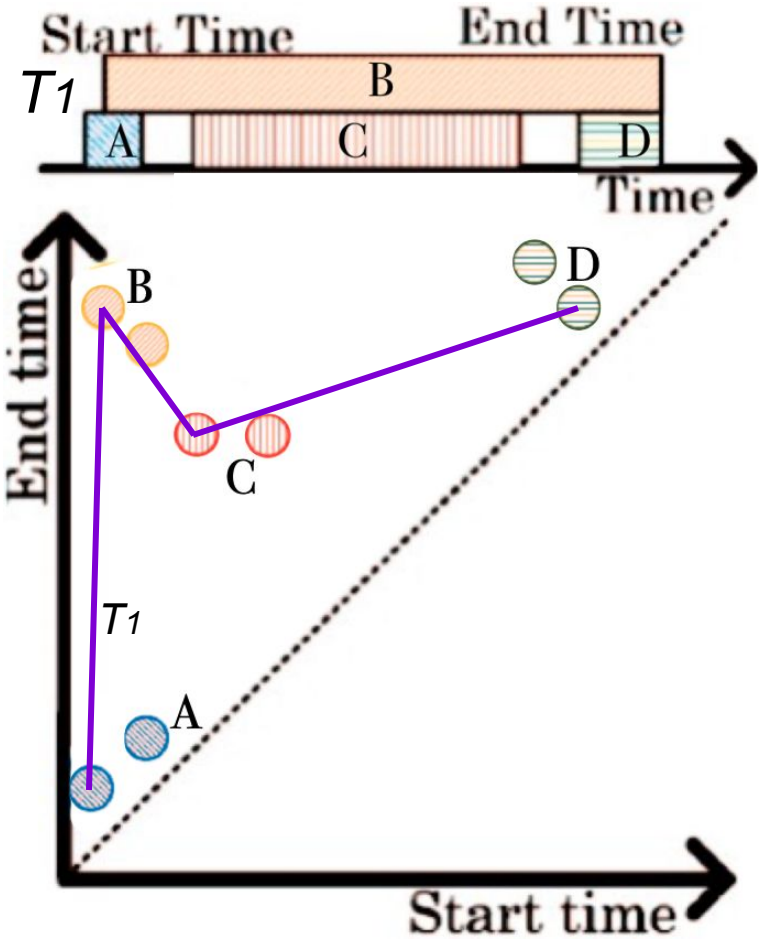
- Process Geometrical Representation
  - Traces *T1* has 4 events
  - Activity ids: *A*, *B*, *C*, *D*

Performance Skyline



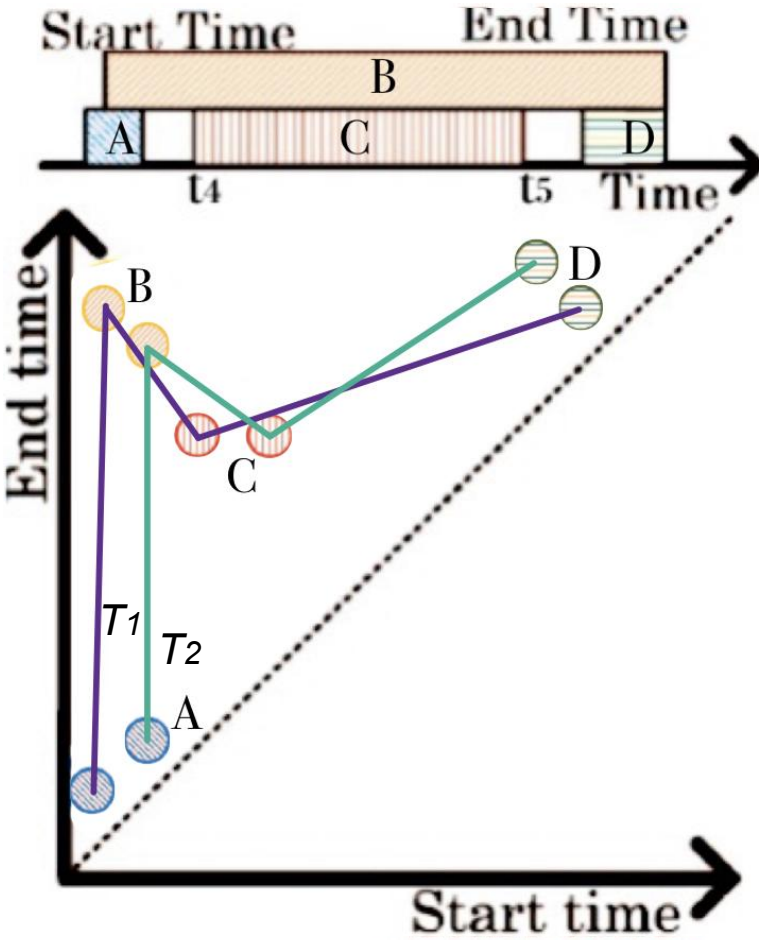
- Process Geometrical Representation
  - Traces *T1* and *T2* have four events each
  - Events of activity ids: *A*, *B*, *C*, *D*

Performance Skyline



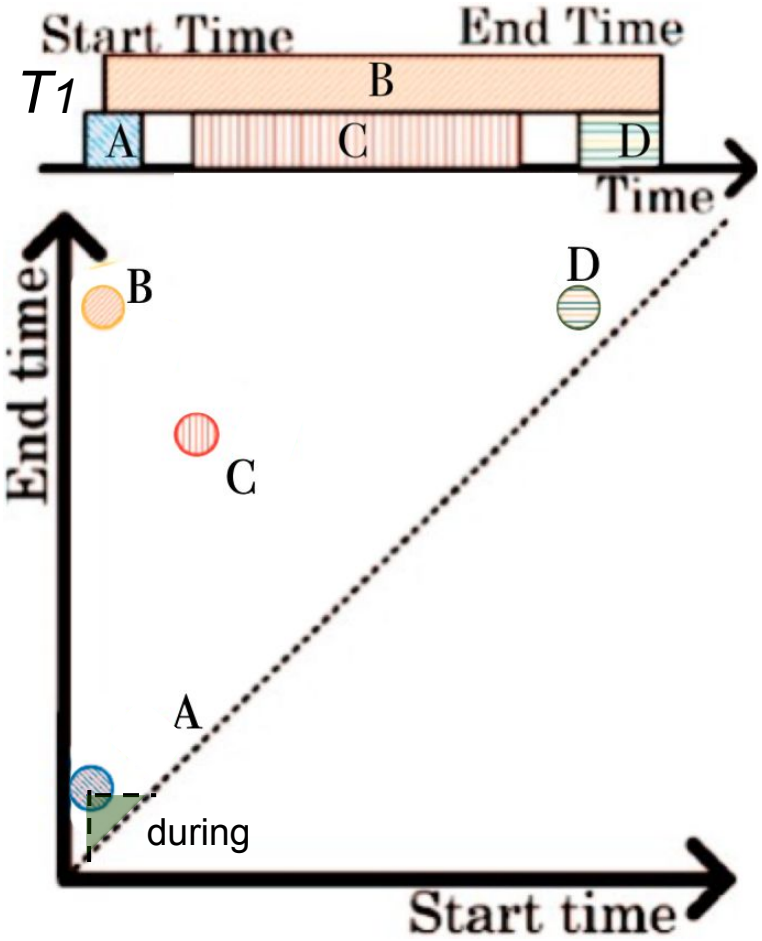
- Process Geometrical Representation
  - Traces *T1* and *T2* have four events each
  - Events of activity ids: *A*, *B*, *C*, *D*

Performance Skyline

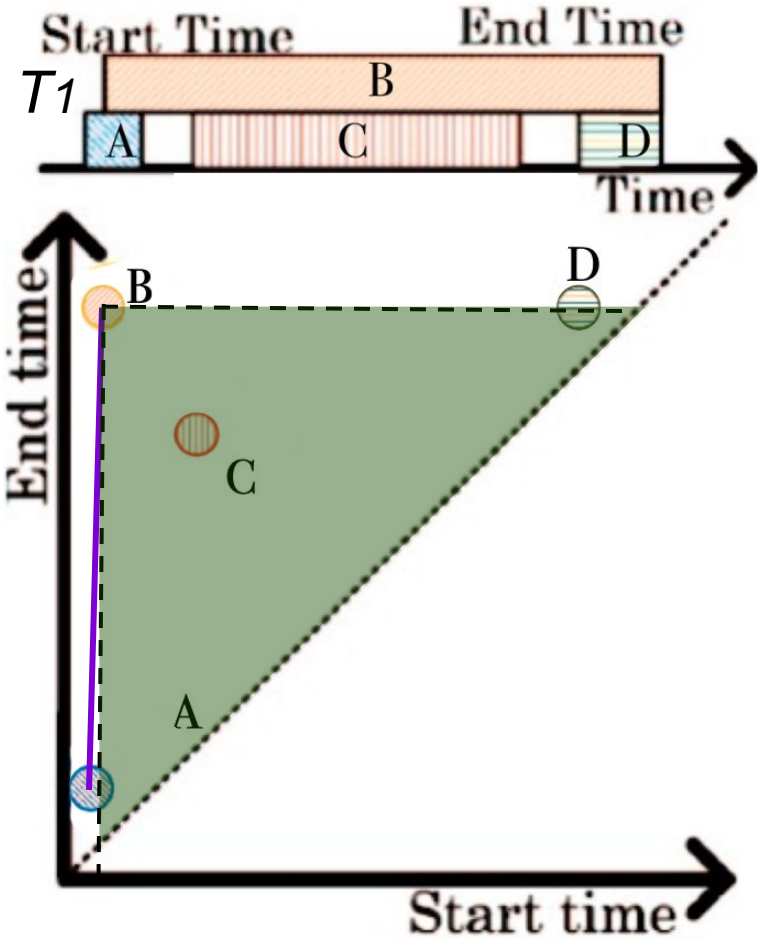


- Process Geometrical Representation
  - Traces *T1* and *T2* have four events each
  - Events of activity ids: *A*, *B*, *C*, *D*

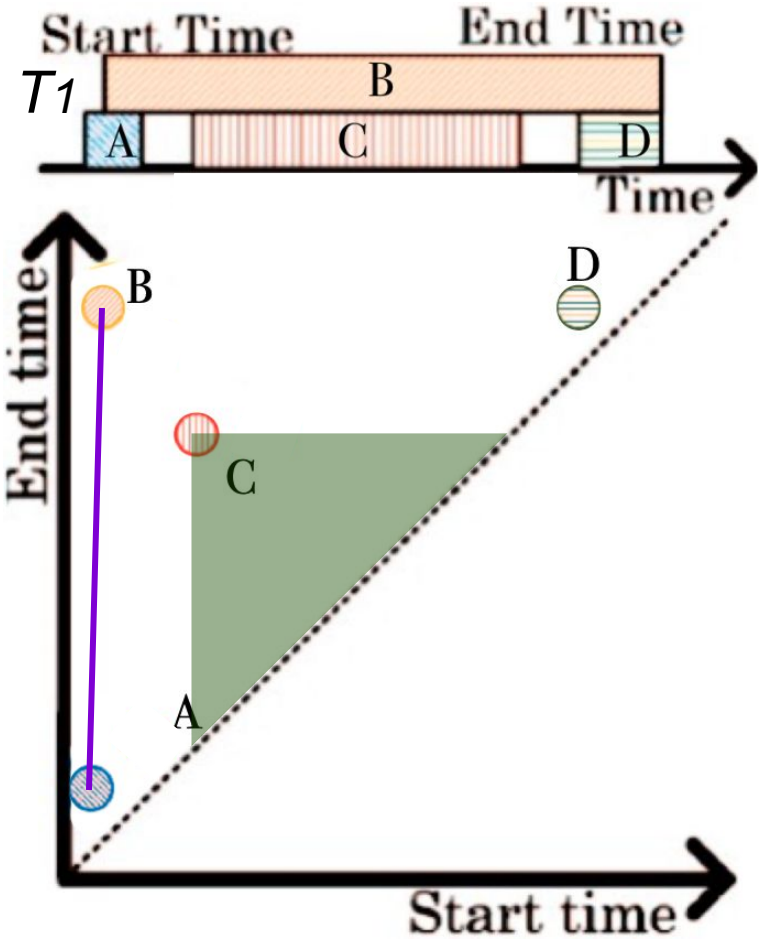
Performance Skyline



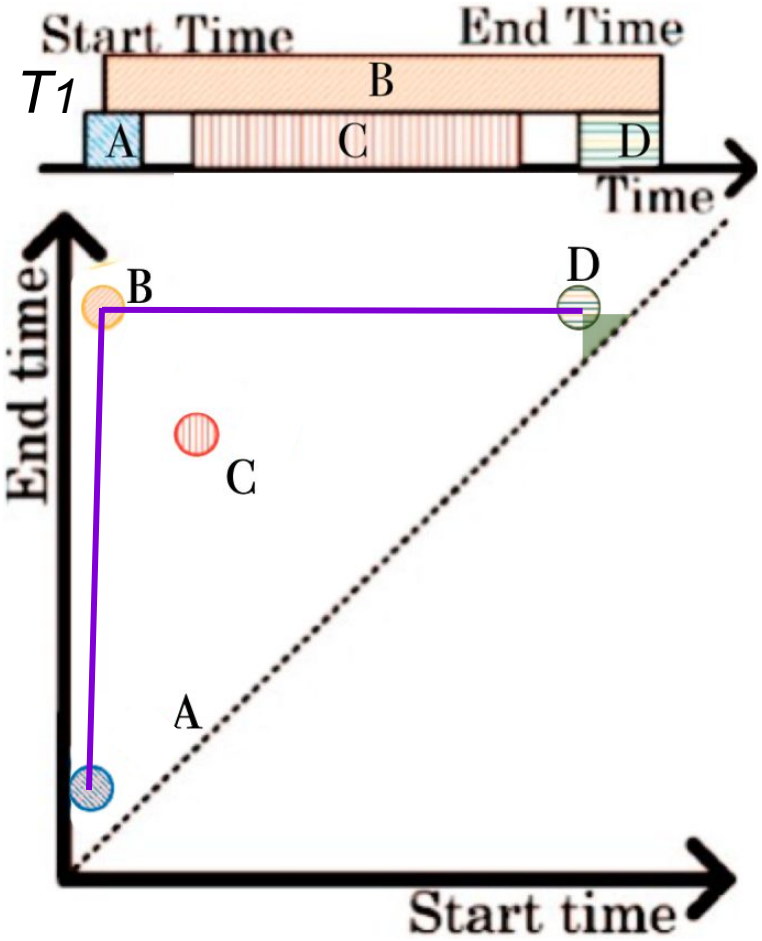
Performance Skyline



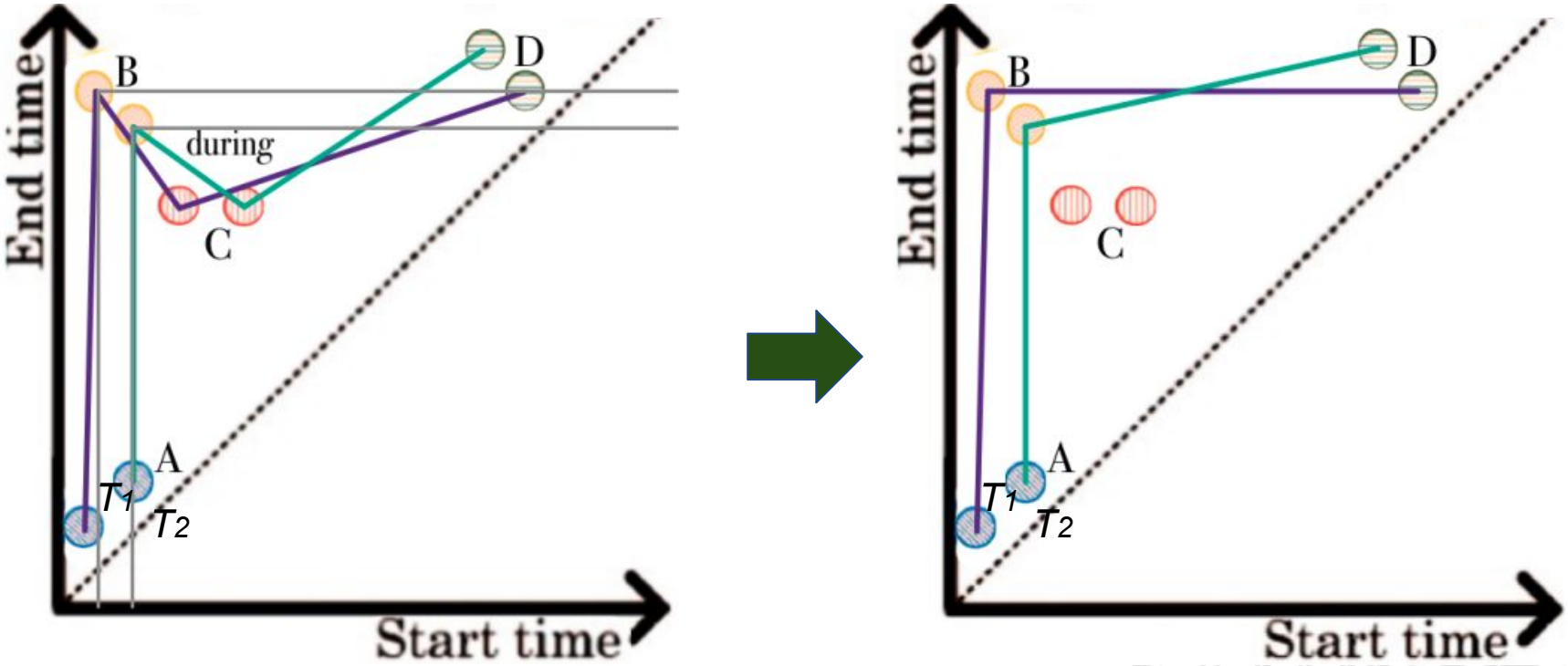
Performance Skyline



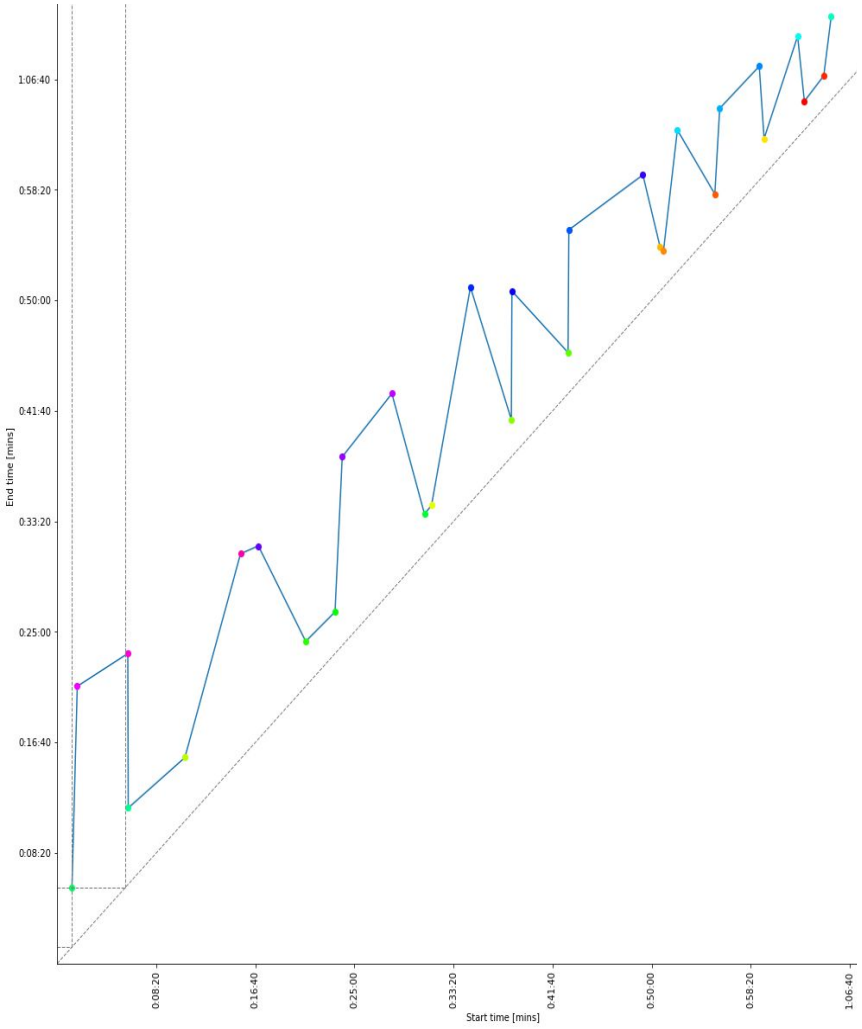
Performance Skyline

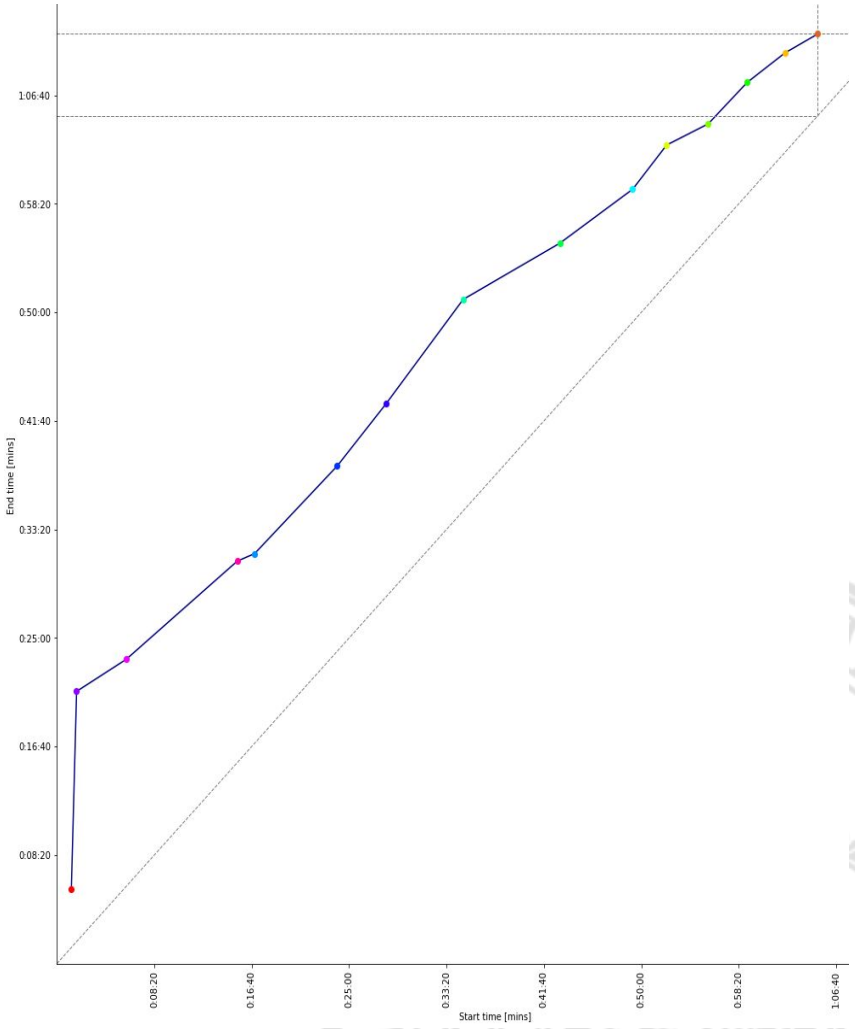


Performance Skyline

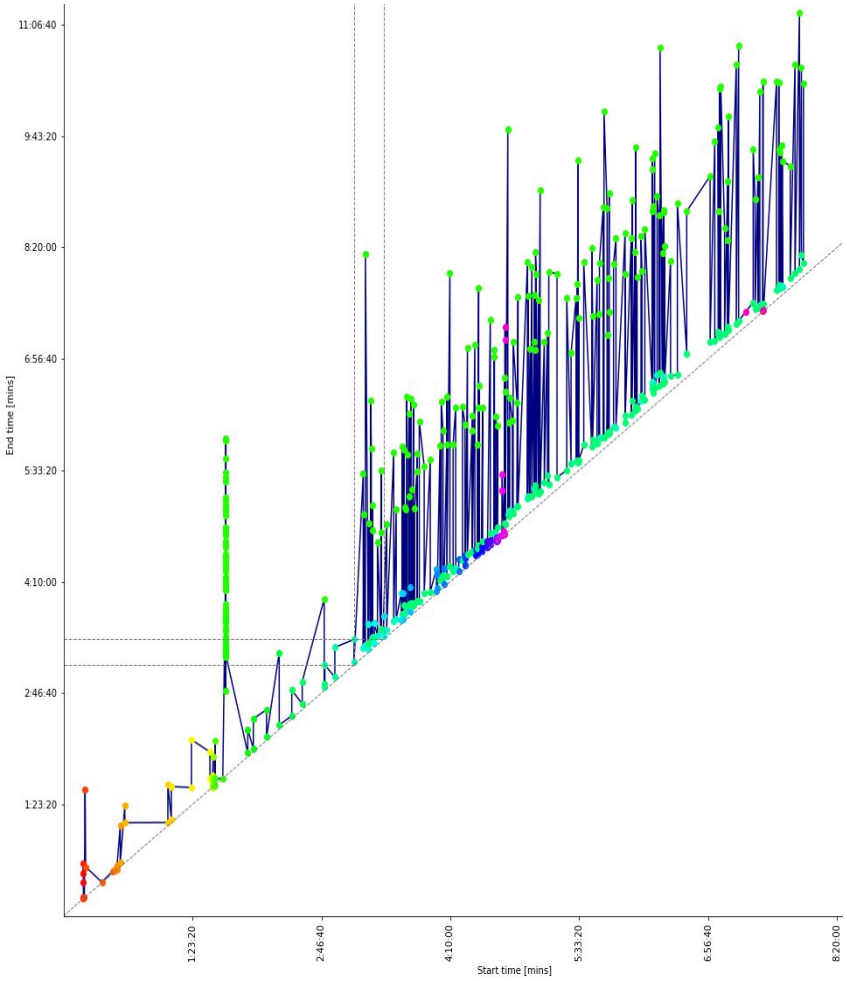


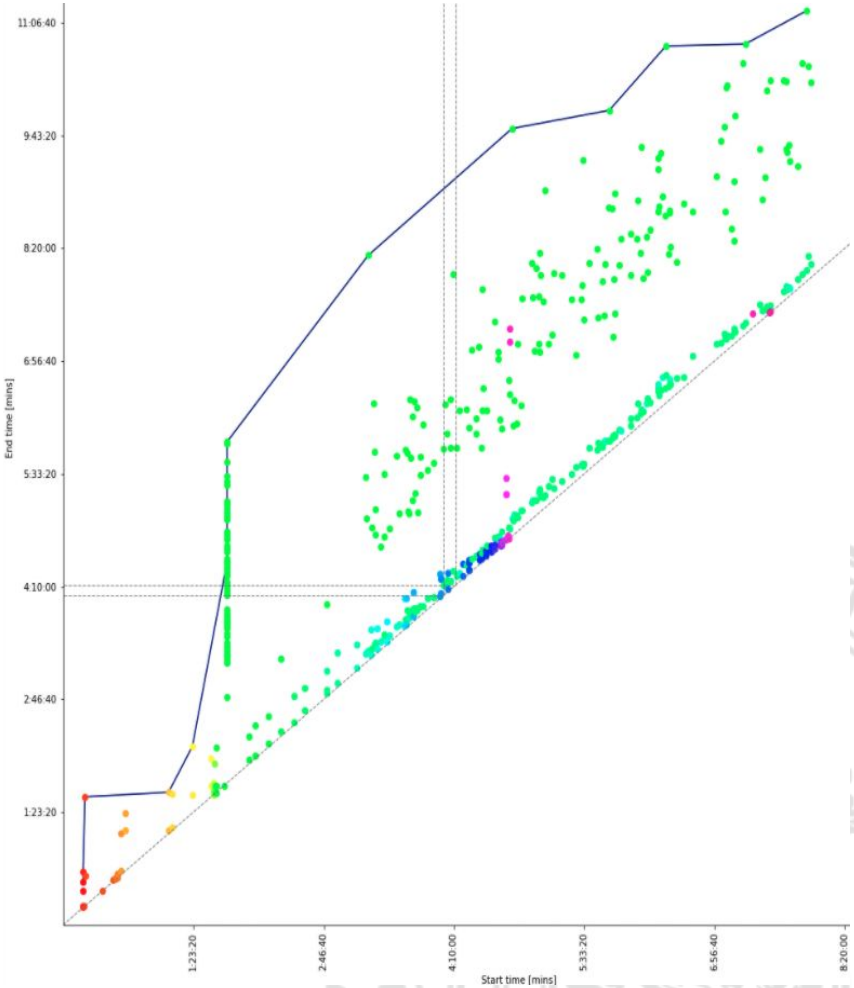
Performance Skyline



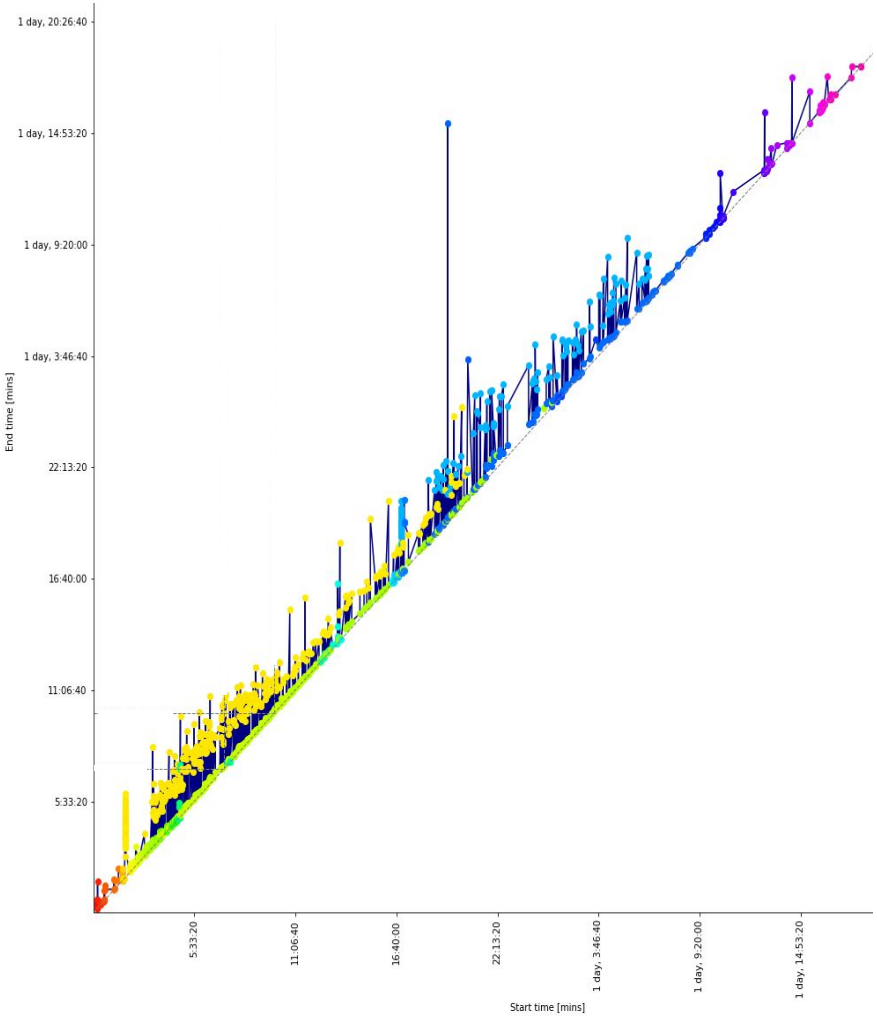


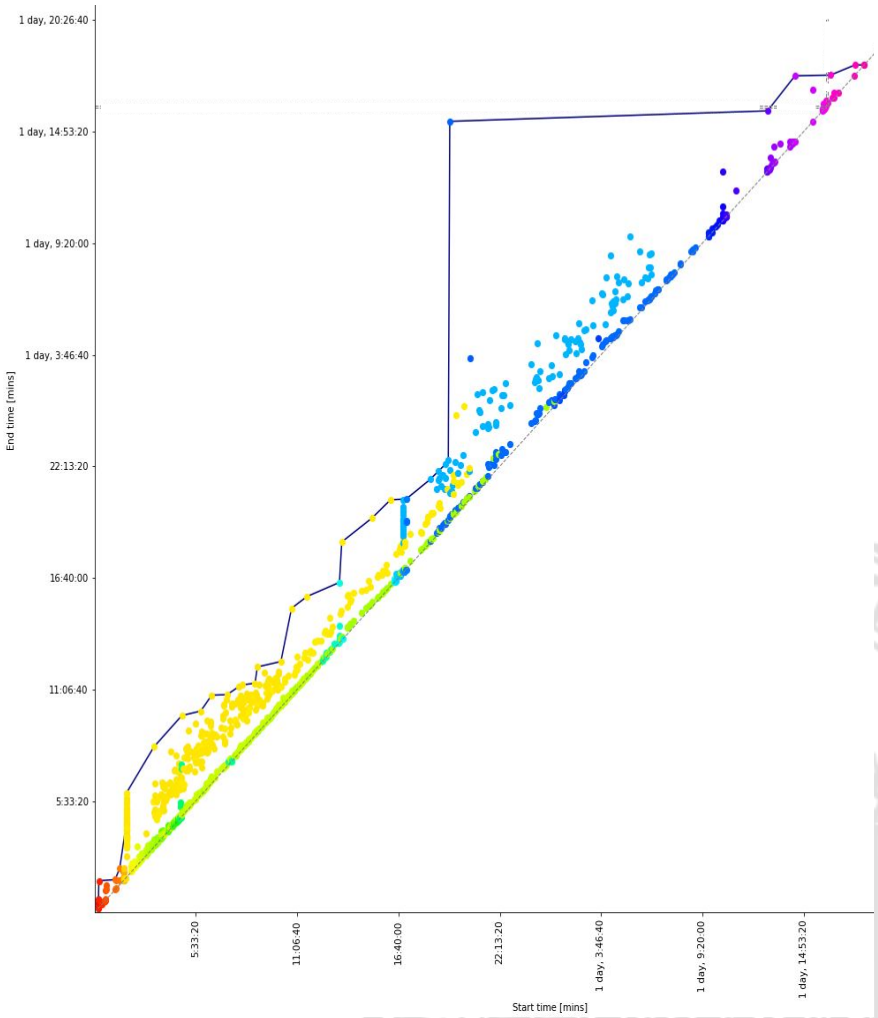
Performance Skyline





Performance Skyline



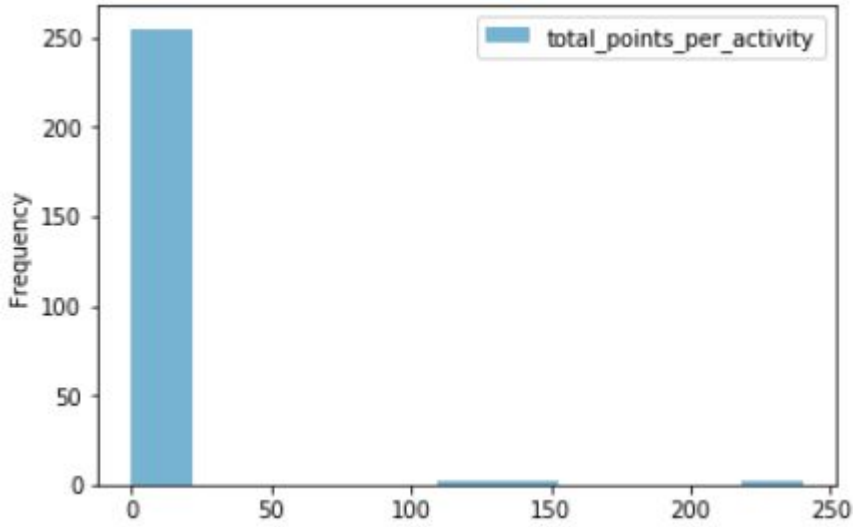


Computational Experiments:

- Trace Set Description
- Anomaly Detection

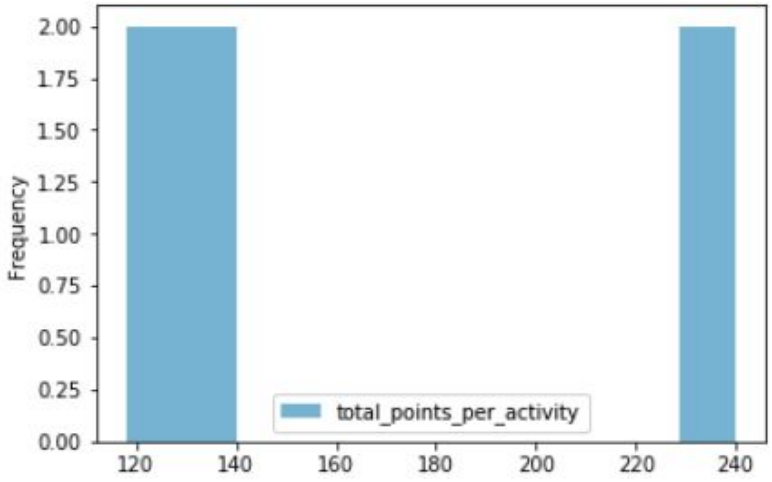
Experiments: Data Description

- 61864 events (~1238 on average)
- 50 traces
- 261 activities



Experiments: Data Description

- 61864 events (~1238 on average)
- 50 traces
- 261 activities



	activity	total_points_per_activity
72	CrawlTask(crawler=creepy-crawly)(chunk=01)	240
133	ExtractTask(crawler=creepy-crawly)(chunk=01)	240
135	ExtractTask(crawler=creepy-crawly)(chunk=03)	139
74	CrawlTask(crawler=creepy-crawly)(chunk=03)	138
134	ExtractTask(crawler=creepy-crawly)(chunk=02)	118
73	CrawlTask(crawler=creepy-crawly)(chunk=02)	118

Anomaly Detection

- Number of Events in Trace
- Percentage of Events on Skyline
- Skyline Activity Set
- Number of Events per Activity on Skyline
- Number of Events in Skyline of Unexpected Duration

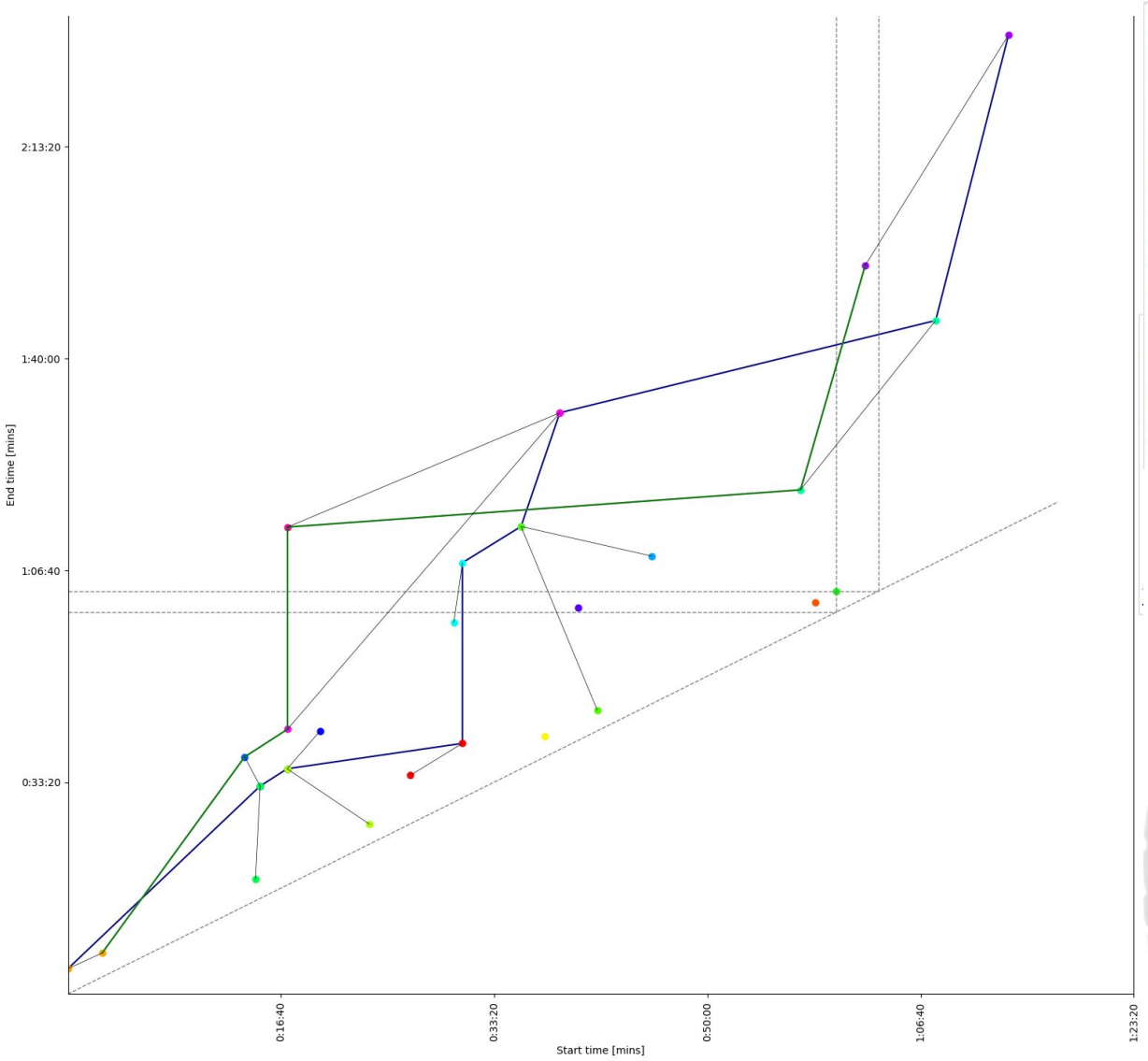
Anomaly Detection

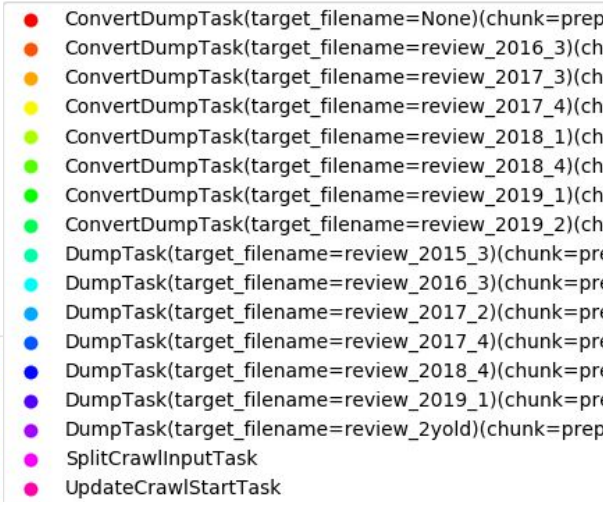
- Number of Events in Trace
- Percentage of Events on Skyline
- Skyline Activity Set
- Number of Events per Activity on Skyline
- Number of Events in Skyline of Unexpected Duration

Anomaly Detection: Results

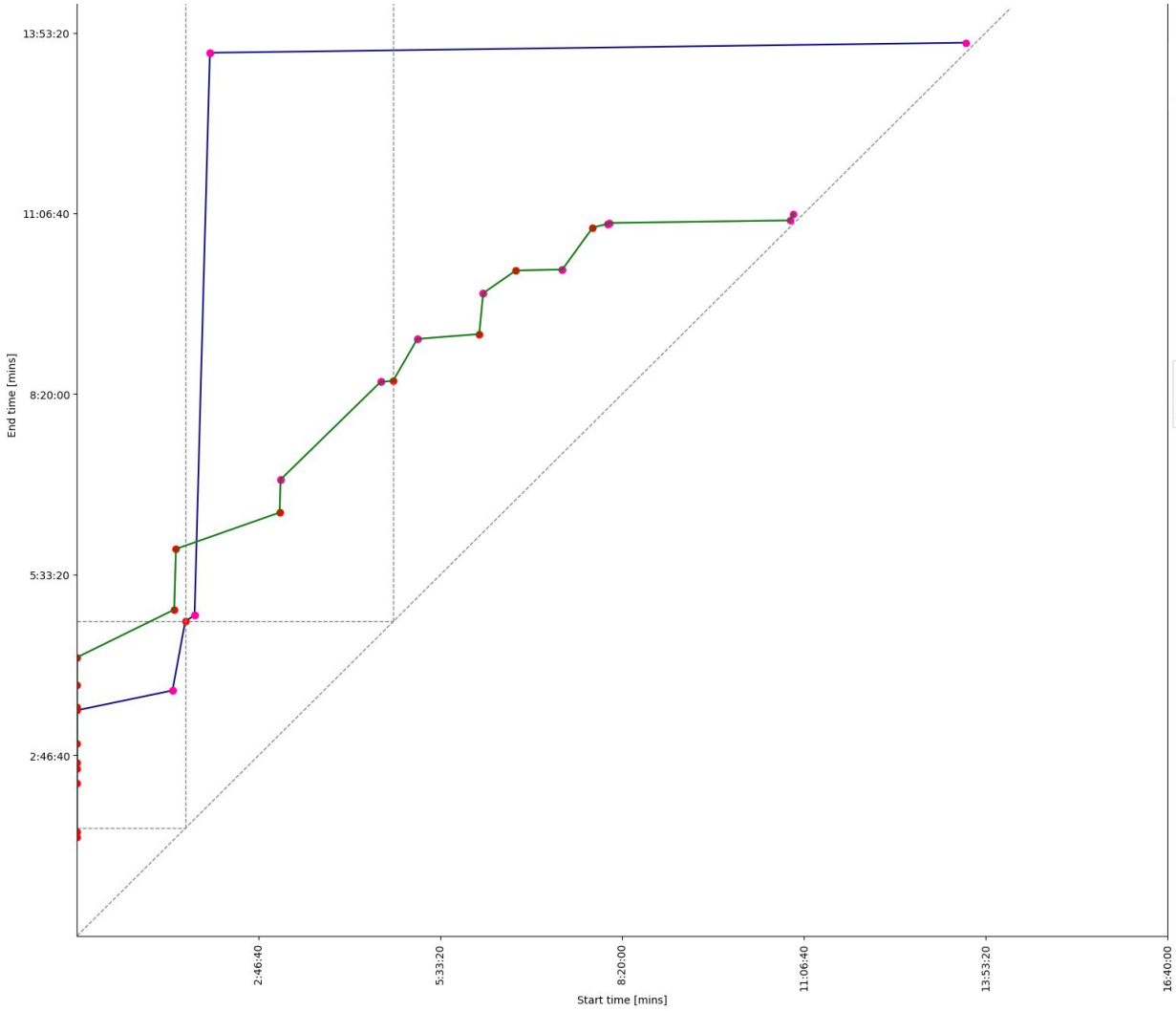
Anomaly Trace Type	**Tolerance Threshold**		**Number of Anomaly Traces**
	Average	Deviation	
Number of Events in Trace	1238	**±1*71.62**	**1**
		±1.5*71.62	1
		±2.25*71.62	1
		±3.375*71.62	1
Percentage of Events in Skyline	5.22	±1*1.38	11
		±1.5*1.38	**4**
		±2.25*1.38	3
		±3.375*1.38	0
Skyline Activity Set	>70% activities of overall set are found		33
	>75%		**17**
	>80%		11
	>85%		8
	>90%		5
Number of Events per Activity on Skyline	>90% anomalous activities are found		10
	>80%		**3**
	>70%		2
	>60%		2
	>40%		1

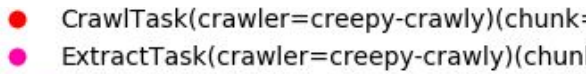
Percentage of Events on Skyline Anomaly



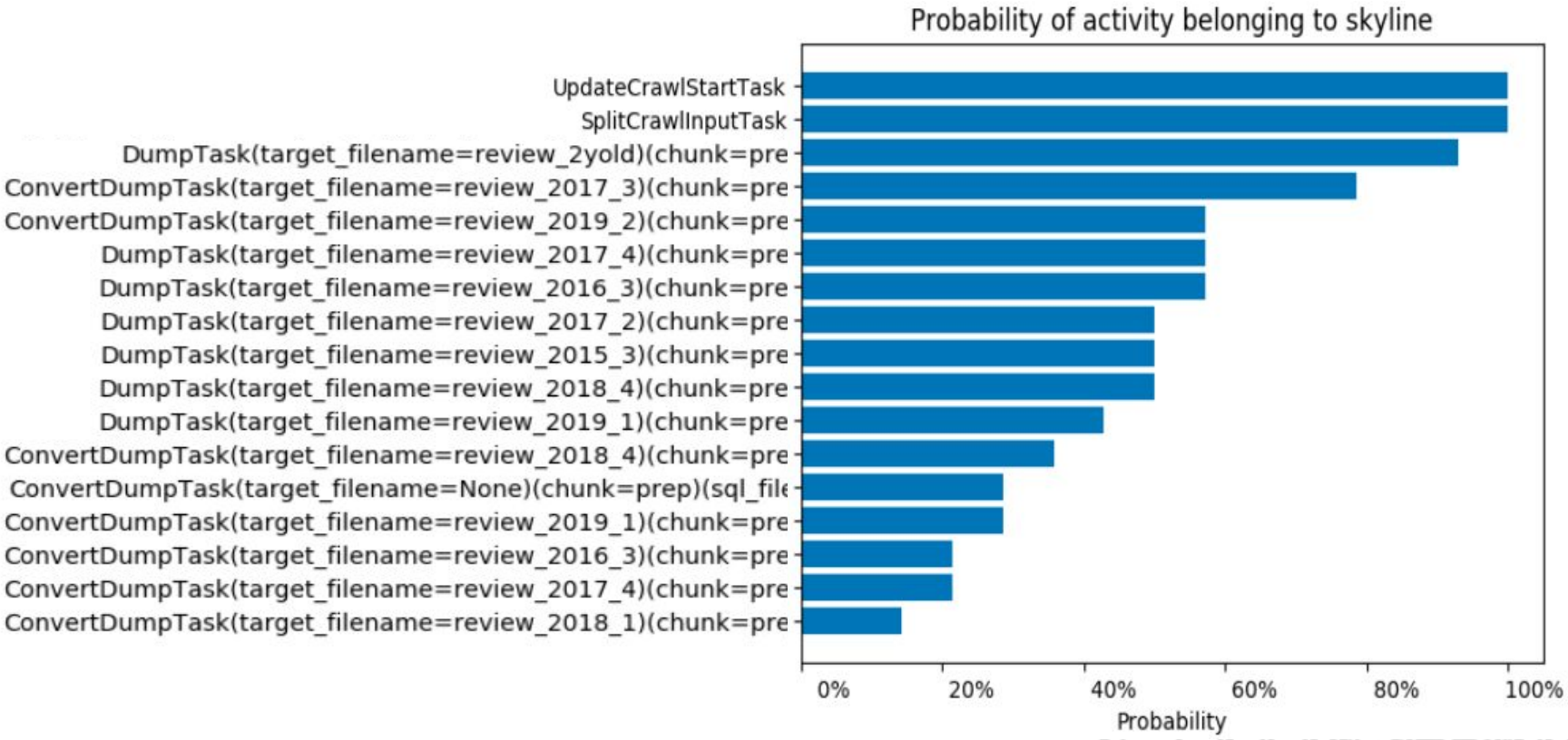


Anomaly Detection: Number of Events per Activity

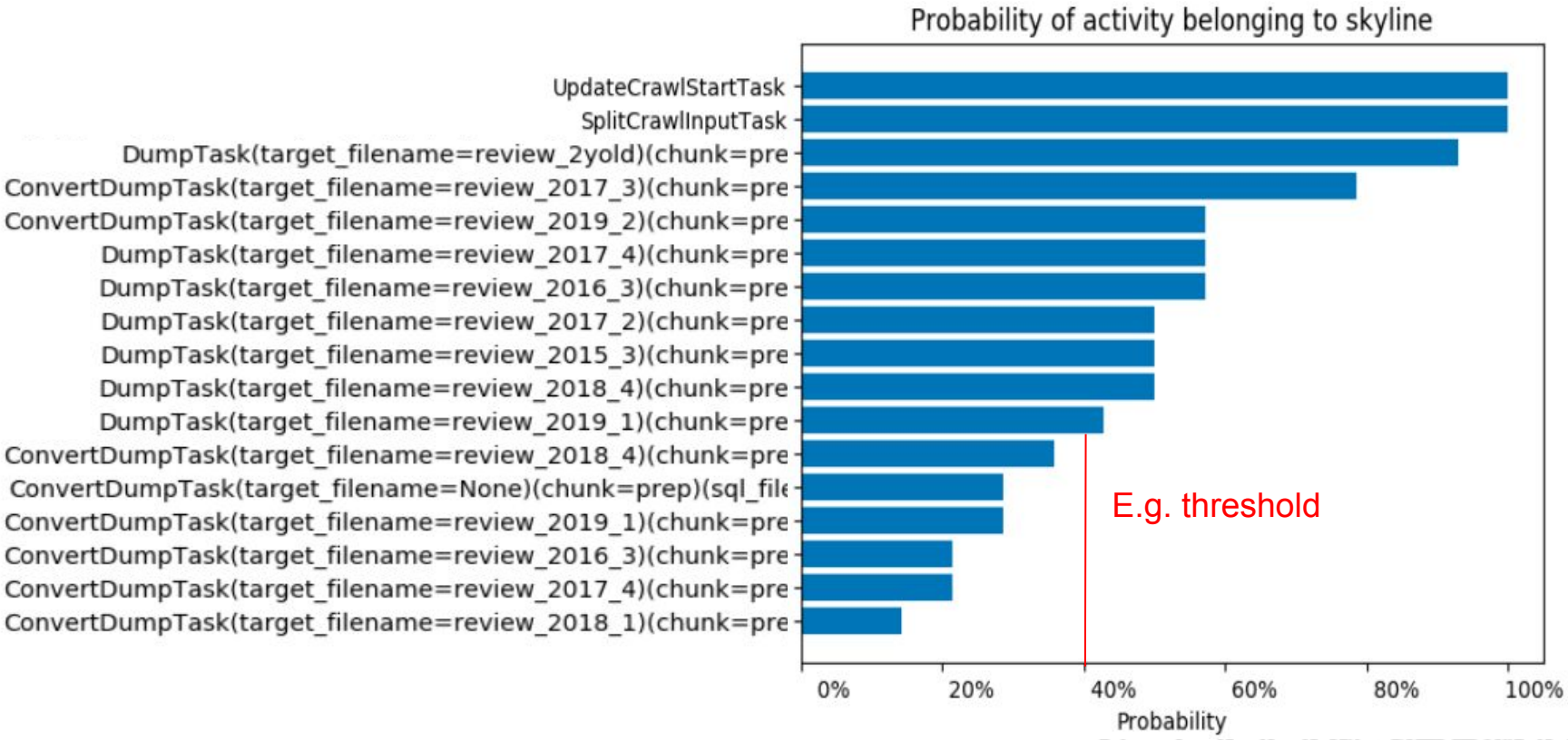




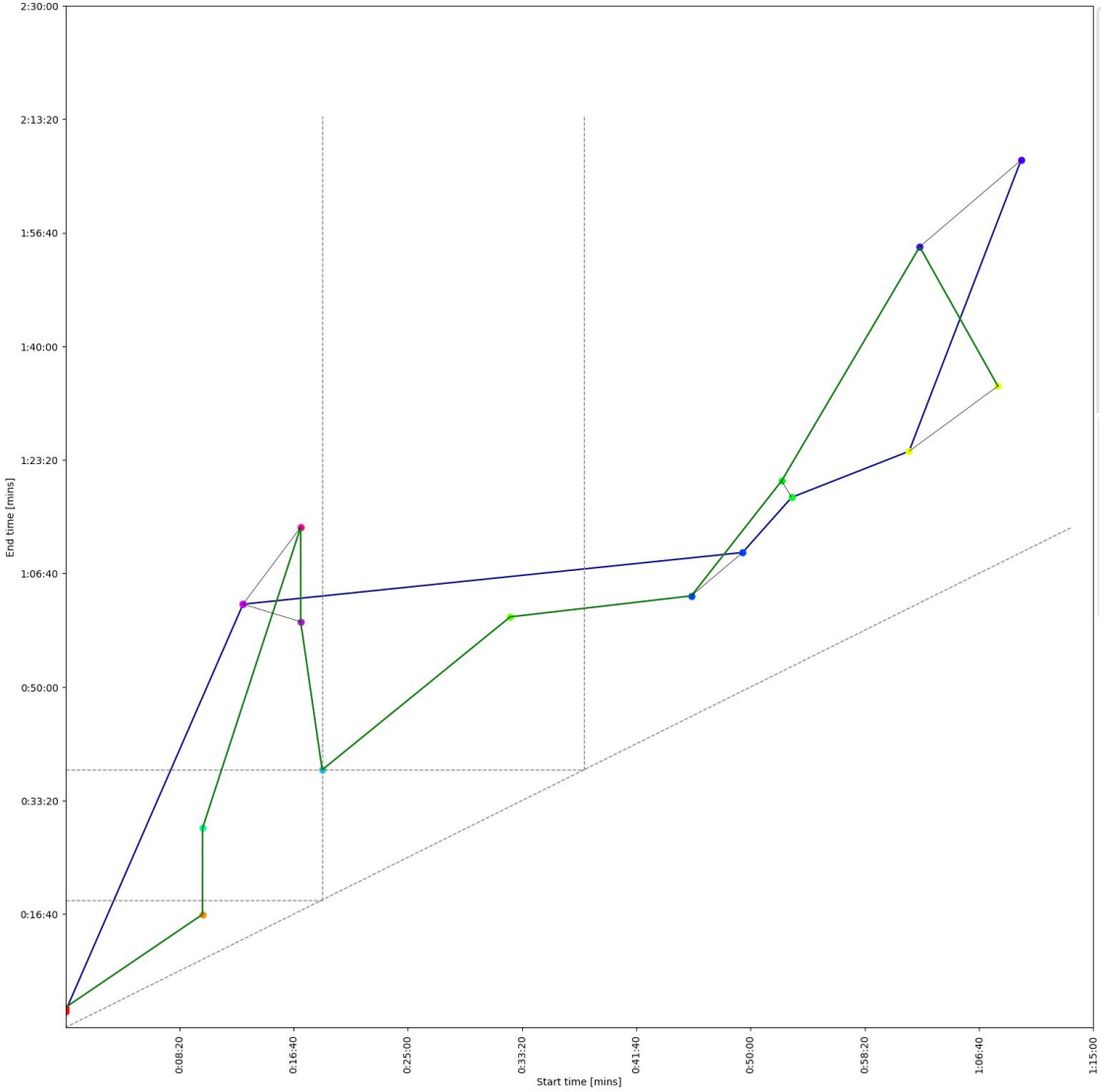
Activity Set Anomaly

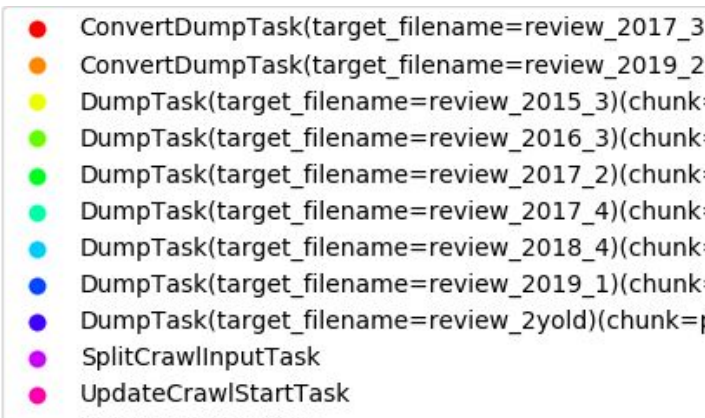


Activity Set Anomaly



Activity Set Anomaly





Final Thoughts & Conclusion

- Achievements:
  - Process performance notation for interval events
  - Visualization of real process event logs
  - Pattern Recognition: Time Variance Anomaly Detection
- Further work:
  - Process discovery: Extensions, alignment, clustering, etc.
  - Conformance checking: Predictions, drift recognition, skyline expectation maximization, etc.
  - Process Enhancement: Bottlenecks recognition, resource allocation, etc.

Bibliography

[1] Wil Aalst and A.K.A. Medeiros. Process mining and security: Detecting anomalous process executions and checking process conformance. *Electronic Notes in Theoretical Computer Science*, 121:3–21, 02 2005.

[2] J.F. Allen. Maintaining Knowledge About Temporal Intervals. volume 26, pages 832–843, New York, NY, USA, 1983. ACM.

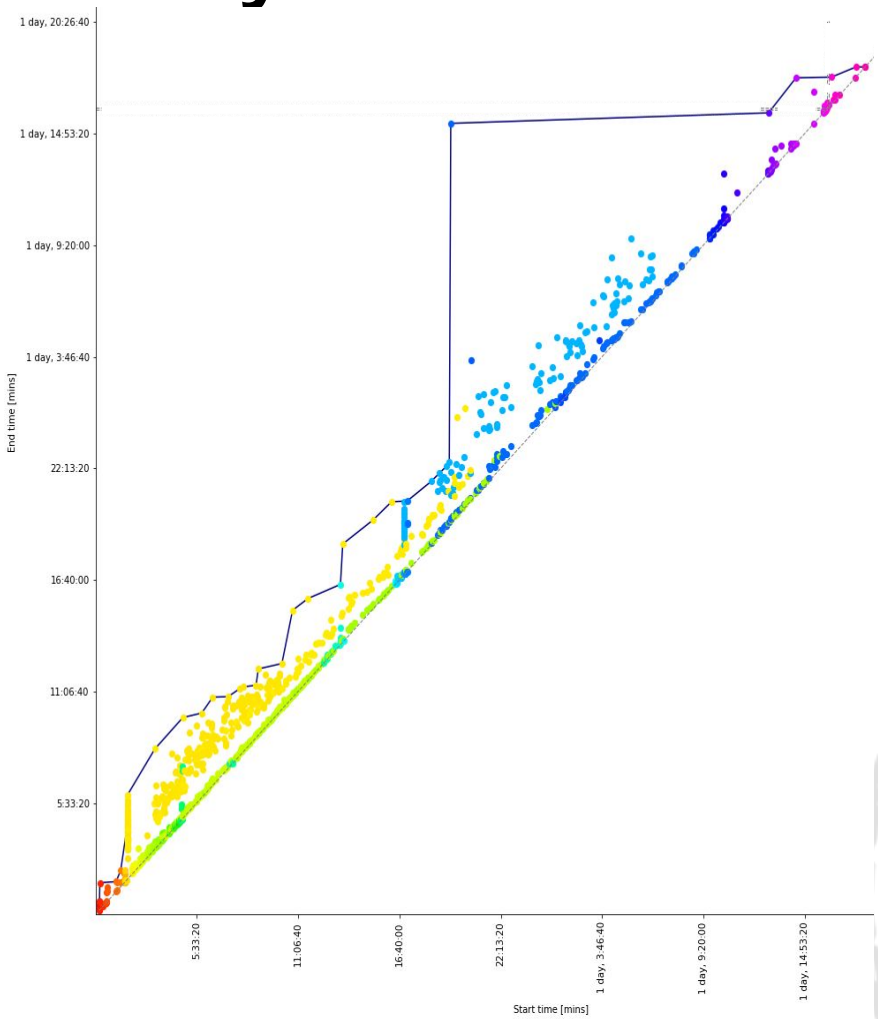
[3] C. Atkins. Prescription or description: some observations on the conceptual modelling process. In *Proceedings 1996 International Conference Software Engineering: Education and Practice*, pages 34–41, 1996.

[4] Andrea Burattin. *Introduction*, pages 1–7. Springer International Publishing, Cham, 2015.

[5] Andrea Burattin and Alessandro Sperduti. Heuristics miner for time intervals. volume 207, 01 2015.

[6] Andrea Burattin, Alessandro Sperduti, and Marco Veluscek. Business models enhancement through discovery of roles. In *IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013, Singapore, 16-19 April, 2013*, pages 103–110, 2013.

In summary…



Thank you!

Bibliography

[1] Wil Aalst and A.K.A. Medeiros. Process mining and security: Detecting anomalous process executions and checking process conformance. *Electronic Notes in Theoretical Computer Science*, 121:3–21, 02 2005.

[2] J.F. Allen. Maintaining Knowledge About Temporal Intervals. volume 26, pages 832–843, New York, NY, USA, 1983. ACM.

[3] C. Atkins. Prescription or description: some observations on the conceptual modelling process. In *Proceedings 1996 International Conference Software Engineering: Education and Practice*, pages 34–41, 1996.

[4] Andrea Burattin. *Introduction*, pages 1–7. Springer International Publishing, Cham, 2015.

[5] Andrea Burattin and Alessandro Sperduti. Heuristics miner for time intervals. volume 207, 01 2015.

[6] Andrea Burattin, Alessandro Sperduti, and Marco Veluscek. Business models enhancement through discovery of roles. In *IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013, Singapore, 16-19 April, 2013*, pages 103–110, 2013.

[7] Vadim Denisov, Elena Belkina, Dirk Fahland, and Wil M.P. van der Aalst. The performance spectrum miner: visual analytics for fine-grained performance analysis of processes. In F. Casati, editor, *BPMTracks 2018*, CEUR Workshop Proceedings, pages 96–100. CEUR-WS.org, 9 2018.

Bibliography

[8] Lucantonio Ghionna, Gianluigi Greco, Antonella Guzzo, and Luigi Pontieri. Outlier detection techniques for process mining applications. volume 4994, pages 150–159, 05 2008.

[9] T. Guyet and R. Quiniou. Mining temporal patterns with quantitative intervals. In *2008 IEEE International Conference on Data Mining Workshops*, pages 218–227, 2008.

[10] Eva L. Klijn and Dirk Fahland. Performance mining for batch processing using the performance spectrum. In *15th International Workshop on Business Process Intelligence*, 2019.

[11] Jens Wischnewsky Marwan Hassani, Yifeng Lu and Thomas Seidl. A geometric approach for mining sequential patterns in interval-based data streams. *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 2128–2135, 2016.

[12] David M. W. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[13] Florian Richter and Thomas Seidl. Tesseract: Time-drifts in event streams using series of evolving rolling averages of completion times. In Josep Carmona, Gregor Engels, and Akhil Kumar, editors, *Business Process Management*, pages 289–305, Cham, 2017. Springer International Publishing.

Bibliography

[14] Pedro Robledo. Process mining plays an essential role in digital transformation. In *https://medium.com/@pedrorobledobpm/process-mining-plays-an-essential-role-in-digital-transformation-384839236bbe*.

[15] Andreas Rogge-Solti and Gjergji Kasneci. Temporal anomaly detection in business processes. In Shazia Sadiq, Pnina Soffer, and Hagen Völzer, editors, *Business Process Management*, pages 234–249, Cham, 2014. Springer International Publishing.

[16] Guangchen Ruan and Beth Plale. Parallel and quantitative sequential pattern mining for large-scale interval-based temporal data. pages 32–39, 01 2015.

[17] Jesse Santiago and Desirae Magallon. Critical path method. *CEE 320 – VDC SEMINAR*, feb 2009.

[18] Ioannis Tsamardinos. *Constraint-Based Temporal Reasoning Algorithms with Applications to Planning*. PhD thesis, 01 2001.

[19] Wil M. P. van der Aalst. *Process Mining: Data Science in Action*. Springer, Heidelberg, 2 edition, 2016.