8-11 September 2019, Bled, Slovenia

ADBIS

Accurate and Transparent Path Prediction Using Process Mining

Gaël BERNARD

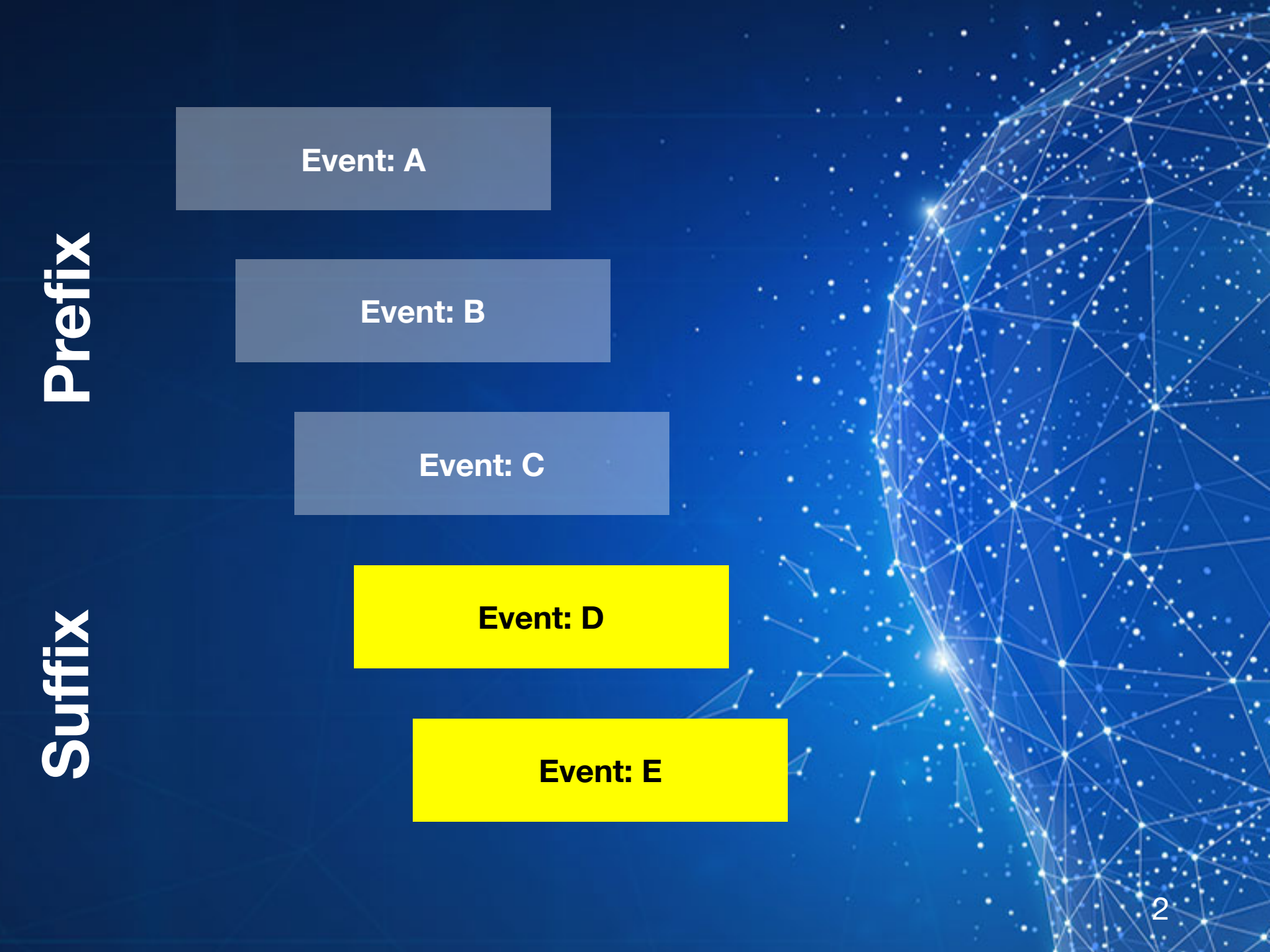
University of Lausanne, Faculty of Business and Economics (HEC), Switzerland

Periklis ANDRITSOS

University of Toronto, Faculty of Information, Canada

UNIL | Université de Lausanne

UNIVERSITY OF TORONTO


Prefix
Suffix
Event: A
Event: B
Event: C
Event: D
Event: E

Use Case: Call Center

Use Case: Healthcare

Use Case: Online Retail

Definitions

Input

Event logs

Trace

Trace

Trace

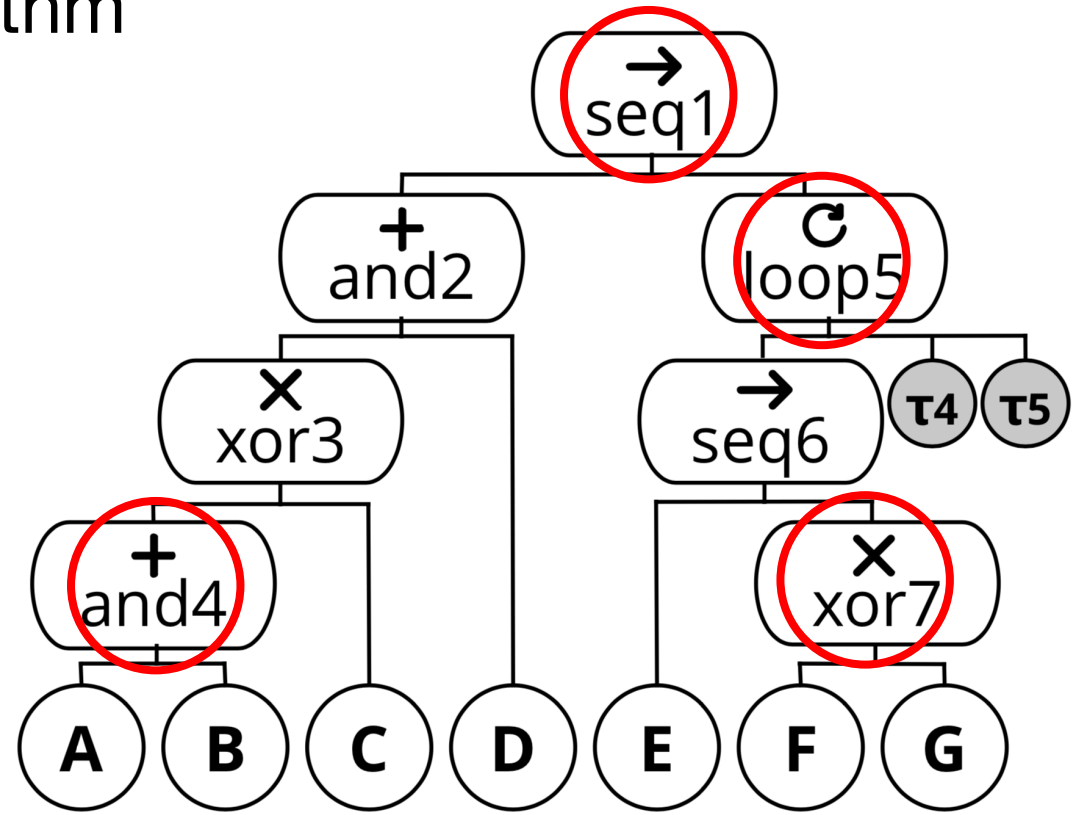
Events:

09/09/2019 - 16:35:37: Open the ticket
09/09/2019 - 16:37:39: Transfer the ticket
20/09/2019 - 13:12:31: Update the information
21/09/2019 - 09:14:32: Inform the customer
21/09/2019 - 09:14:32: Close the ticket

Process Mining

Event Logs:
<abdef> <bdaegef>
<dcefeg> <cdeg>

- Process Mining Discovery Algorithm
- Inductive Miner
- Process Tree



Related Works

- LSTM [1]
- Process Mining based approach [2]

[1] Tax, N., Verenich, I., La Rosa, M., Dumas, M.: Predictive business process mon- itoring with lstm neural networks. In: International Conference on Advanced In- formation Systems Engineering. pp. 477–492. Springer (2017)
[2] Polato, M., Sperduti, A., Burattin, A., de Leoni, M.: Time and activity sequence prediction of business process instances. arXiv preprint arXiv:1602.07566 (2016)


LaFM
Loop aware Footprint Matrix

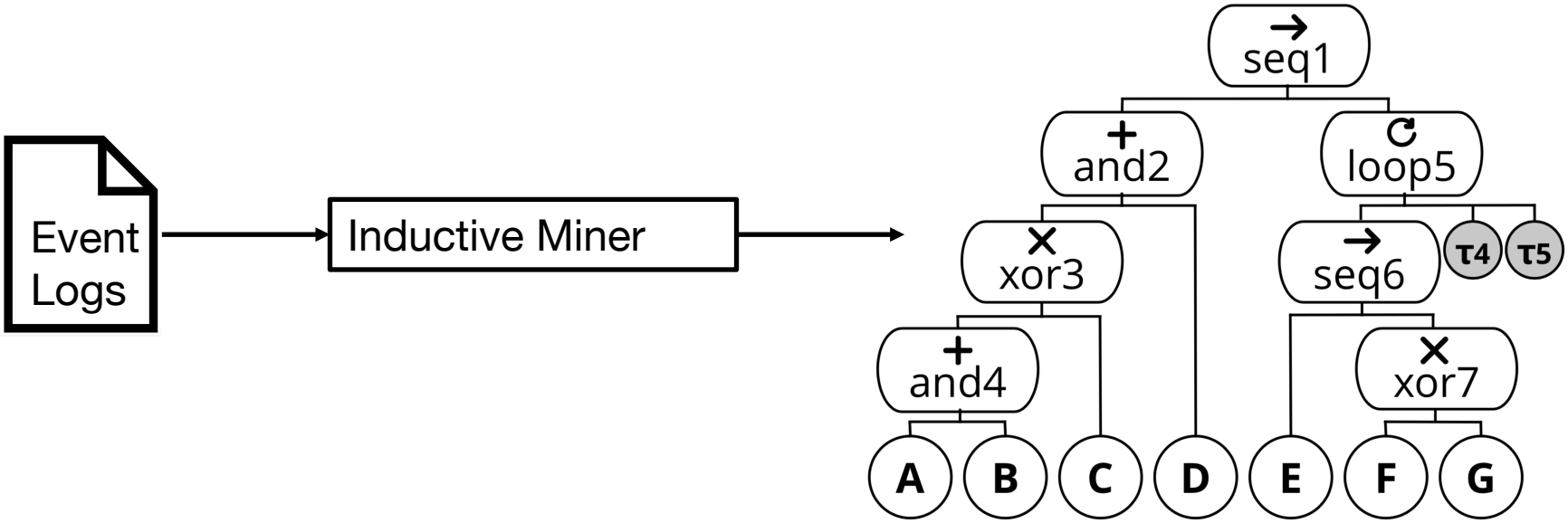
LaFM Loop aware Footprint Matrix

Step 1: Discover a process model

Step 2: Build a footprint

Step 3: Make prediction using the footprint

Discover a process model



Capturing the Behaviors

Parallel +

Order of execution

Exclusive choice ×

Branch executed

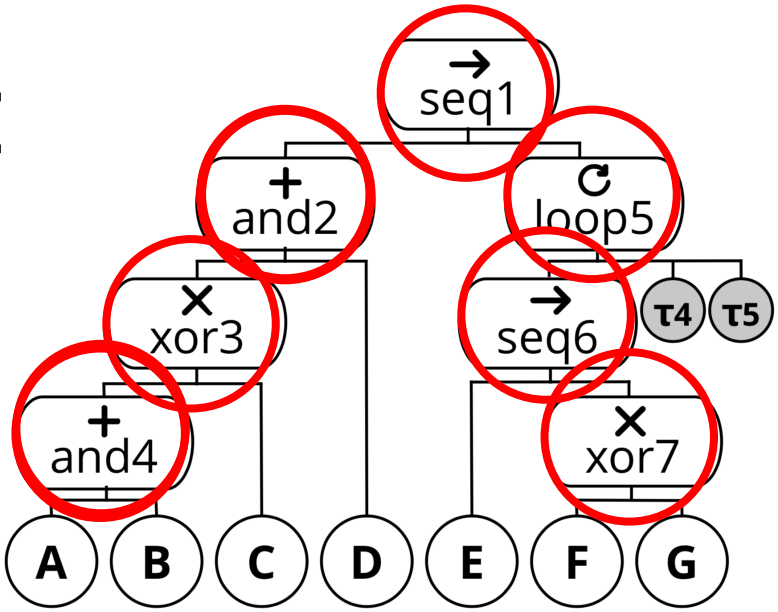
Loop ↻

Number of times loops are executed

Build the footprint

To Record:

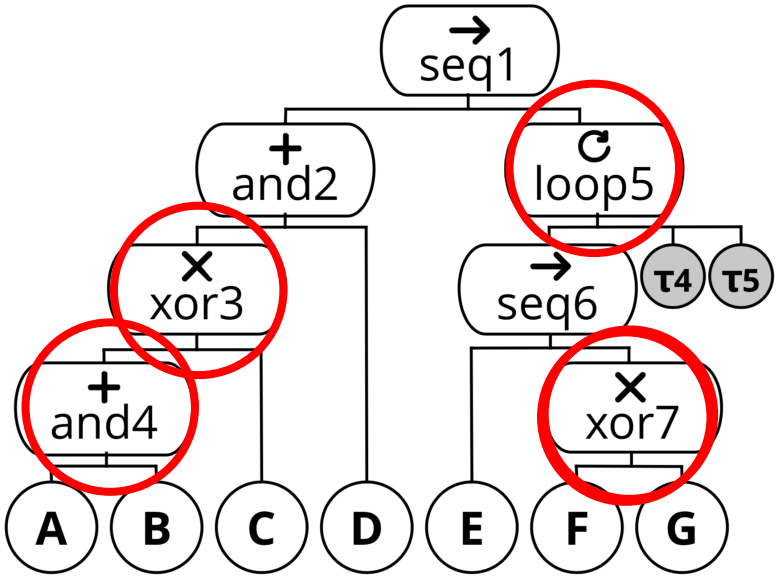
- **Parallel** +
- **Exclusive choice** ×
- **Loop** ↻



Traces	and2(1)|	and2(2)|	and2(3)|	and4(1)|	and4(2)|	loop5|	xor7|loop5{1}	xor7|loop5{2}	xor3|
ABDEF	1	1	2	1	2	1	1	∅	1
BDAEGEF	1	2	1	2	1	2	2	1	1
DCEFEG	2	1	∅	∅	∅	2	1	2	2
…									

Predict

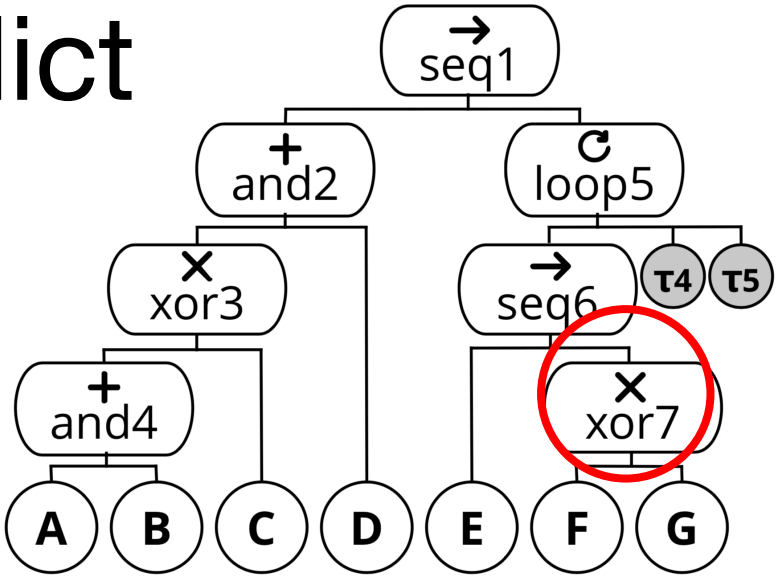
- Prefix: DABEFEF



Traces	and2(1)|	and2(2)|	and2(3)|	and4(1)|	and4(2)|	loop5|	xor7|loop5{1}	xor7|loop5{2}	xor3|
ABDEF	1	1	2	1	2	1	1	∅	1
BDAEGEF	1	2	1	2	1	2	2	1	1
DCEFEG	2	1	∅	∅	∅	2	1	2	2
…									

Abstract and predict

- Prefix: DCEFEFEFEFE..
- Xor7|Loop5{6} => ?



Traces	and2(1)|	and2(2)|	and2(3)|	and4(1)|	and4(2)|	loop5|	xor7|loop5{1}	xor7|loop5{2}	xor3|
ABDEF	1	1	2	1	2	1	1	∅	1
BDAEGEF	1	2	1	2	1	2	2	1	1
DCEFEG	2	1	∅	∅	∅	2	1	2	2
…									

Evaluation Procedure

- 30 synthetic datasets, publicly available [1]
- 2/3 => training, 1/3 => test [2]
- Algorithms tested:
  - LaFM, Markov Chain, LSTM
- Metric used for accuracy:
  - Damerau–Levenshtein similarity [3]

[1] https://data.4tu.nl/repository/uuid:7455 4e7-8cc0-45b8-8a89-93e9c9dfab05
[2] Tax, N., Verenich, I., La Rosa, M., Dumas, M.: Predictive business process mon- itoring with lstm neural networks. In: International Conference on Advanced In- formation Systems Engineering. pp. 477–492. Springer (2017)
[3] Damerau, F.J.: A technique for computer detection and correction of spelling er- rors. Communications of the ACM 7(3), 171–176 (1964)

Results

	treeSeed	1	2	3	4	5*	6	7	8	9	10
Round 3	LaFM	**1.00**	**1.00**	**0.58**	0.31	n/a*	**0.70**	**0.57**	**0.46**	**0.66**	0.91
	lstm	**1.00**	**1.00**	0.50	0.29	n/a*	0.60	0.42	0.44	0.50	**0.92**
	markov	0.60	**1.00**	0.20	**0.37**	n/a*	0.60	0.15	0.33	0.46	**0.92**

	treeSeed	1	2	3	4	5	6	7	8	9	10
Round 4	LaFM	**0.84**	**0.90**	0.39	0.30	0.66	**0.54**	**0.39**	0.34	**0.51**	**0.52**
	lstm	0.81	0.85	**0.43**	**0.35**	**0.83**	0.51	0.35	**0.36**	0.50	0.24
	markov	0.55	0.89	0.26	0.31	0.72	0.43	0.17	0.32	0.45	0.50

	treeSeed	1	2	3	4	5	6	7	8	9	10
Round 5	LaFM	0.51	**0.48**	**0.24**	**0.50**	**0.86**	**0.63**	**0.36**	**0.21**	0.48	**0.54**
	lstm	**0.56**	0.36	0.21	0.42	0.85	0.62	0.30	0.14	0.22	0.29
	markov	0.28	0.42	0.13	0.41	0.45	0.56	0.17	0.17	**0.50**	0.47


C-LaFM
Clustered LaFM


PLEASE
CLOSE
GATE

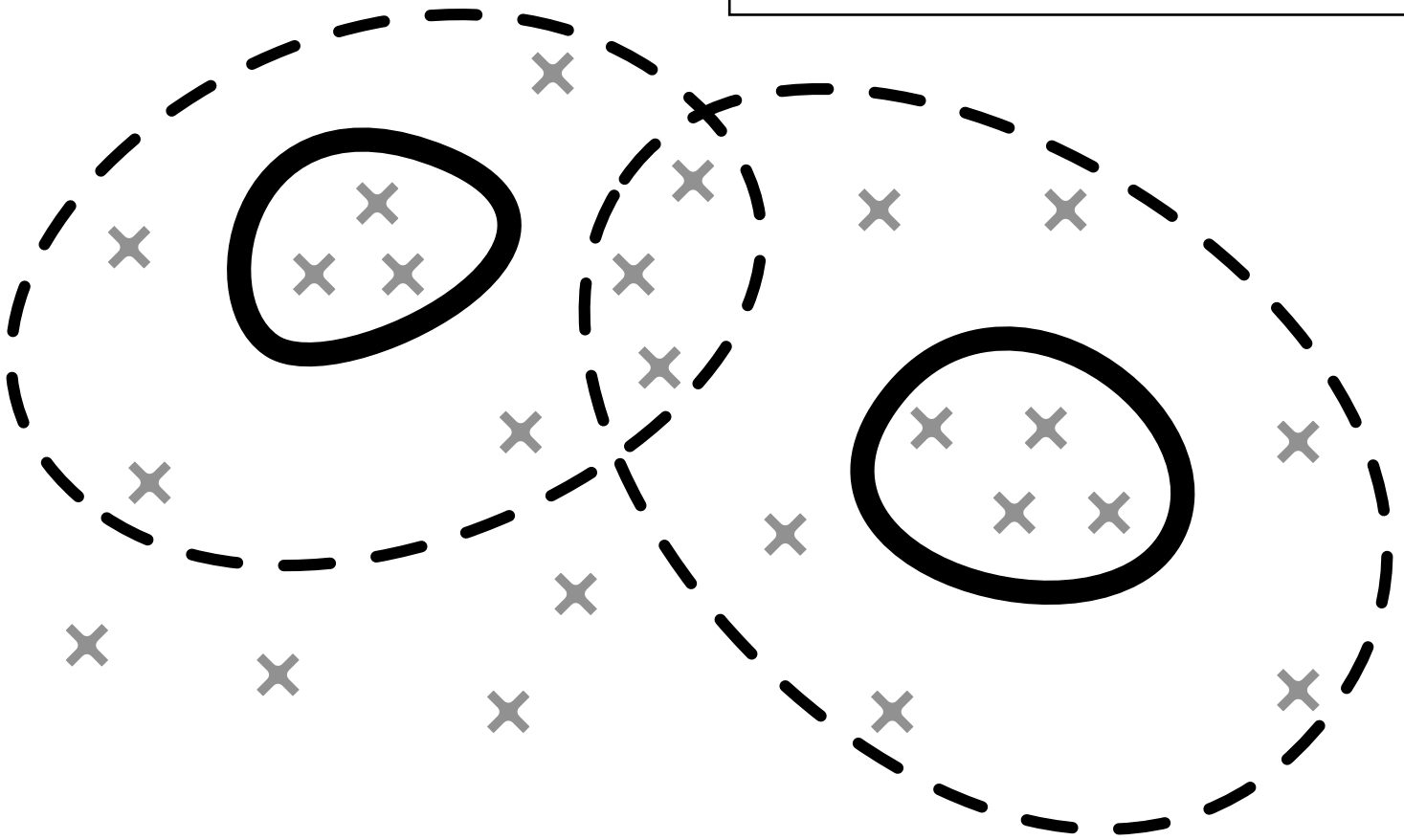
C-LaFM

- Intuition: Complex datasets can be well describe using several process models.
- C-LaFM: Clustered LaFM
  - Based on Ngrams
  - Clustering using HDBSCAN [1]

[1] L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering In: Journal of Open Source Software, The Open Journal, volume 2, number 11. 2017

C-LaFM

Strong representatives	Discover BPM
Weak representatives	Replay



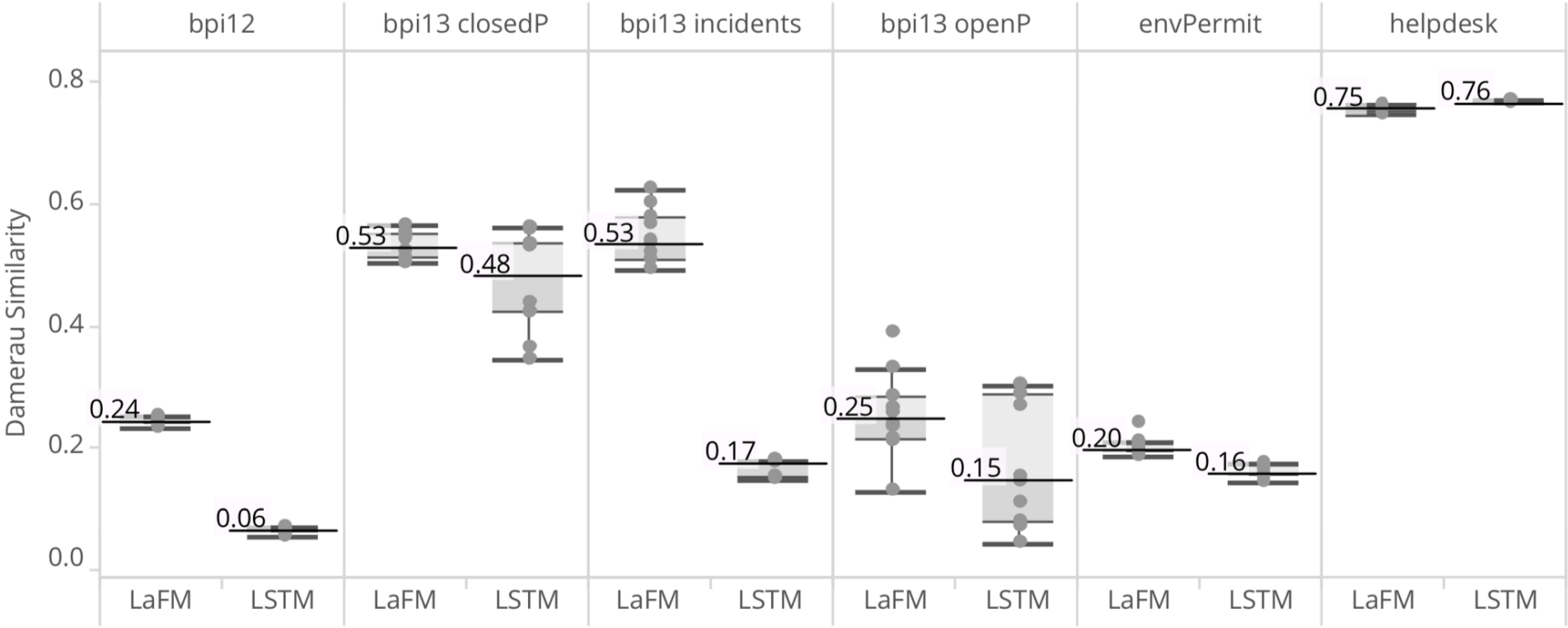
C-LaFM: Classifier

- SGD: Stochastic Gradient Descent classifier
- Training phase:
  - Train the strong representative with all prefix lengths
- Prediction phase:
  - Apply the SGD to assign the prefix to a cluster

Evaluation

Description	#traces	#events
Events from a ticketing system	3’804	13’710
Loan process for a financial industry. Note: keeping only manual task and lifecycle: complete as described in [3]	9’658	72’413
Closed problem - management system from Volvo IT Belgium	6’660	1’487
Incidents - management system from Volvo IT Belgium	7’554	65’533
Open problems - management system from Volvo IT Belgium	819	2’351
Execution of a building permit application process. Note: we pick the Municipality 1	38’944	937

Evaluation


bpi12
bpi13 closedP
bpi13 incidents
bpi13 openP
envPermit
helpdesk
Damerau Similarity
0.0
0.2
0.4
0.6
0.8
0.24
0.06
0.53
0.48
0.53
0.17
0.25
0.15
0.20
0.16
0.75
0.76
LaFM
LSTM

Evaluation

	Dataset	bpi12	bpi13 closedP	bpi13 incidents	bpi13 openP	envPermit	helpdesk
Execution time	LaFM	~45 min	~2 min	~47 min	~1 min	~3.1 hours	~2 min
	LSTM	~15.4 hours	~35 min	~20.6 hours	~18 min	~5.6 hours	~41 min

Conclusion

Conclusion

- Black-Box vs White-Box
- Limitations:
  - Pieces missing for Explainable AI:
    - Intelligible way to propose the prediction
  - Alternatives to Inductive Miner, HDBSCAN, and SGD not tested