



Process Mining on Uncertain Event Data

Marco Pegoraro, RWTH Aachen University

International Conference on Process Mining, 31 October 2021

Uncertainty in Event Logs

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	05-01-2011:11.22	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request	Pete	50	...
	35654522	30-12-2010:15.06	examine casually	Mike	400	...
	35654524	30-12-2010:16.34	check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	decide	Sara	200	...
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	check ticket	Pete	100	...
	35654531	09-01-2011:09.55	decide	Sara	200	...
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	...

Wil M.P. van der Aalst. "Process Mining: Data science in action." Springer.

Uncertainty in Event Logs

Case id	Event id	Properties				...
		Timestamp	Activity	Resource	Cost	
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	05-01-2011:11.22	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request	Pete	50	...
	35654522	30-12-2010:15.06	examine casually	Mike	400	...
	35654524	30-12-2010:16.34	check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	decide	Sara	200	...
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	check ticket	Pete	100	...
	35654531	09-01-2011:09.55	decide	Sara	200	...
35654533	15-01-2011:10.45	pay compensation	Ellen	200	...	

Is this a casual or thorough examination?

Uncertainty in Event Logs

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	05-01-2011:11.22	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request	Pete	50	...
	35654522	30-12-2010:15.06	examine casually	Mike	400	...
	35654524	30-12-2010:16.34	check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	decide	Sara	200	...
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	check ticket	Pete	100	...
	35654531	09-01-2011:09.55	decide	Sara	200	...
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	...

**Overlapping timestamps!
In which order did the
events occur?**

Uncertain Data

- Uncertain event data: **events with quantified imprecision in their attributes**
- “Quantified” means we can obtain a **description** of the uncertain attribute value(s)
 - For categorical attributes: **a set of possible values**
 - For numerical attributes: **an interval of possible values**
- Often obtained through **pre-processing** and **domain knowledge**

Uncertain Event Logs

Case id	Event id	Properties	
		Timestamp	Activity
1	35654423	30-12-2010:11.02	register request
	35654424	31-12-2010:10.06	examine
	35654425	05-01-2011:15.12	check ticket
	35654426	06-01-2011:11.18	decide
	35654427	07-01-2011:14.24	reject request
2	35654483	30-12-2010:11.32	register request
	35654485	30-12-2010:12.12	check ticket
	35654487	30-12-2010:14.16	examine casually
	35654488	05-01-2011:11.22	decide
	35654489	05-01-2011:11.22	pay compensation

Case id	Event id	Timestamp	Activity
1	35654423	30-12-2010: 11.02	register request
	35654424	31-12-2010: 10.06	{examine casually, examine thoroughly}
	35654425	05-01-2011: 15.12	check ticket
	35654426	06-01-2011: 11.18	decide
	35654427	07-01-2011: 14.24	reject request
2	35654483	30-12-2010: 11.32	register request
	35654485	30-12-2010: 12.12	check ticket
	35654487	30-12-2010: 14.16	examine casually
	35654488	[05-01-2011: 11.22.00, 05-01-2011: 11.22.59]	decide
	35654489	[05-01-2011: 11.22.00, 05-01-2011: 11.22.59]	pay compensation

Uncertain Data

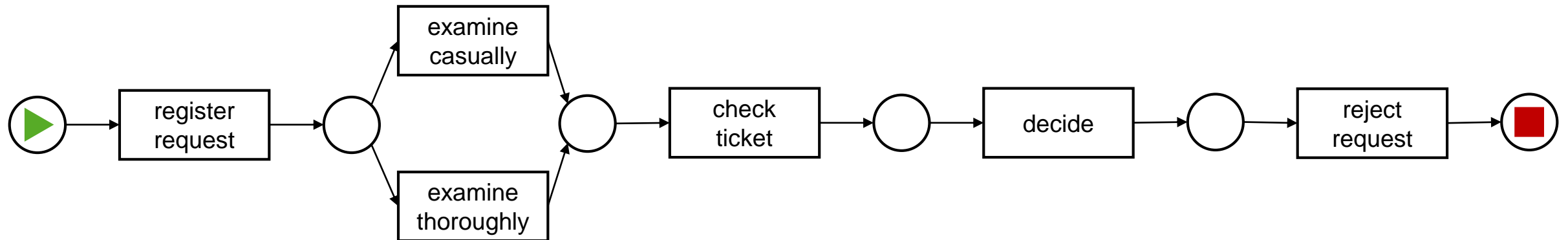
- Correspond to multiple **possible scenarios** in real life
- Normally, these anomalies are considered noise and **filtered or repaired**
- **How do we extract insights from uncertain traces?**

Uncertain Data

- The answer: **we blur the line between data and model**
- Uncertain data contain **behaviour**, which can be represented through a **model**

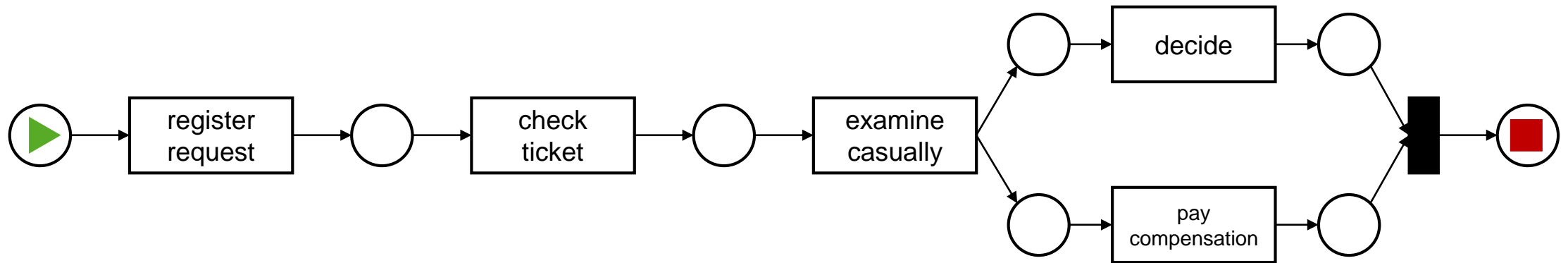
Uncertain Traces

Case id	Event id	Timestamp	Activity
1	35654423	30-12-2010: 11.02	register request
	35654424	31-12-2010: 10.06	{examine casually, examine thoroughly}
	35654425	05-01-2011: 15.12	check ticket
	35654426	06-01-2011: 11.18	decide
	35654427	07-01-2011: 14.24	reject request



Uncertain Traces

Case id	Event id	Timestamp	Activity
2	35654483	30-12-2010: 11.32	register request
	35654485	30-12-2010: 12.12	check ticket
	35654487	30-12-2010: 14.16	examine casually
	35654488	[05-01-2011: 11.22.00, 05-01-2011: 11.22.59]	decide
	35654489	[05-01-2011: 11.22.00, 05-01-2011: 11.22.59]	pay compensation



- Representing uncertain traces through graphs **enables process mining on uncertain event data**
 - We can determine **bounds for conformance scores** of uncertain traces against a reference model
 - We can obtain **models** reflecting the degree of belief about relationships between activities in presence of uncertainty

Conclusion

- Real-life data contain **uncertainty**
- Generally, it is regarded as **noise**: however, it contains **value!**
- Specialized process mining techniques can **gather insights** from uncertain data
- This can be done by representing uncertain data with **models**



Marco Pegoraro

pegoraro@pads.rwth-aachen.de

www.mpegoraro.net

 [@pegoraro_marco](https://twitter.com/pegoraro_marco)

 <https://www.researchgate.net/profile/Marco-Pegoraro-2>

References

Pegoraro, Marco, Merih Seran Uysal, and Wil M.P. van der Aalst. "Discovering process models from uncertain event data." *International Conference on Business Process Management*. Springer, Cham, 2019.

Pegoraro, Marco, Merih Seran Uysal, and Wil M.P. van der Aalst. "Conformance checking over uncertain event data." *Information Systems* (2021): 101810.

Pegoraro, Marco, Merih Seran Uysal, and Wil M.P. van der Aalst. "Efficient time and space representation of uncertain event data." *Algorithms* 13.11 (2020): 285.

Pegoraro, Marco, Merih Seran Uysal, and Wil MP van der Aalst. "PROVED: A Tool for Graph Representation and Analysis of Uncertain Event Data." *International Conference on Applications and Theory of Petri Nets and Concurrency*. Springer, Cham, 2021.