An XES Extension for Uncertain Event Data

Marco Pegoraro, Merih Seran Uysal, Wil M.P. van der Aalst, RWTH Aachen University

International Conference on Business Process Management
8 September 2021

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

© Marco Pegoraro (RWTH Aachen University)

Event Log

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 thoroughly	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	08-01-2011:12.05	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	...

Case 1:
<register request, examine thoroughly, check ticket, decide, reject request>

Case 2:
<register request, check ticket, examine casually, decide, pay compensation>

Case 3:
<register request, examine casually, check ticket, decide, reinitiate request, … >

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data

- Events with **uncertain attributes**: might correspond to **two or more activities**, or the order between events is **lost**
- Correspond to multiple **possible scenarios** in real life
- Normally, these anomalies are considered noise and **filtered**
- **How do we extract insights from uncertain traces?**

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Example

- A patient enrolls in a clinical trial for a treatment against blood cancer. The enrollment includes a lab exam and a visit with a specialist.

- The lab exam, performed on the 8th of July, detects thrombocytopenia (TP) in the patient's blood, **which might be primary** (symptom of the cancer) **or secondary** (from other causes).

- At the visit on the 10th, the patient reports **a possible episode** of night sweats occurred on the 5th. Moreover, the physician detects an enlargement of the spleen, but **it is unclear when this developed**.

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Example

- This results in the following uncertain trace:

Case ID	Event ID	Timestamp	Activity	Indeterminacy
ID192	e_1	2011-07-05	$NightSweats$	?
ID192	e_2	2011-07-08	$\{PrimaryTP, SecondaryTP\}$	
ID192	e_3	[2011-07-04, 2011-07-10]	$SpleenEnlargement$	

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Example

- This results in the following uncertain trace:

Case ID	Event ID	Timestamp	Activity	Indeterminacy
ID192	e_1	2011-07-05	$NightSweats$	?
ID192	e_2	2011-07-08	$\{PrimaryTP, SecondaryTP\}$	
ID192	e_3	[2011-07-04, 2011-07-10]	$SpleenEnlargement$	

The NightSweats event might not have occurred: we mark it with this additional attribute

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Example

- This results in the following uncertain trace:

Case ID	Event ID	Timestamp	Activity	Indeterminacy
ID192	e_1	2011-07-05	*NightSweats*	?
ID192	e_2	2011-07-08	$\{PrimaryTP, SecondaryTP\}$	
ID192	e_3	[2011-07-04, 2011-07-10]	*SpleenEnlargement*	

Event e_2 has two possible activity labels, indicating either primary or secondary thrombocytopenia

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Example

- This results in the following uncertain trace:

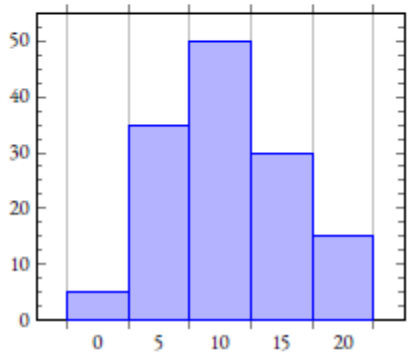
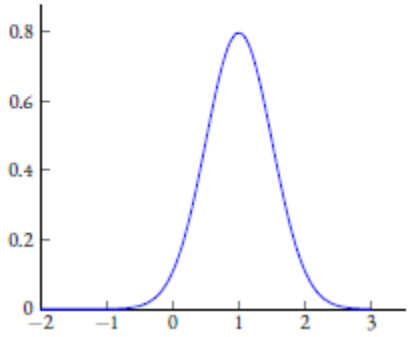
Case ID	Event ID	Timestamp	Activity	Indeterminacy
ID192	e_1	2011-07-05	$NightSweats$	?
ID192	e_2	2011-07-08	$\{PrimaryTP, SecondaryTP\}$	
ID192	e_3	[2011-07-04, 2011-07-10]	$SpleenEnlargement$	

Event e_3 might have occurred in any moment within an interval of time

Notice that the exact order of events is lost or unknown

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data - Types

	Weak uncertainty	Strong uncertainty
Discrete data	Discrete probability distribution 	Set of possible values $\{x, y, z, \dots\}$
Continuous data	Probability density function 	Interval $\{x \in \mathbb{R} \mid a \leq x \leq b\}$

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data Representation

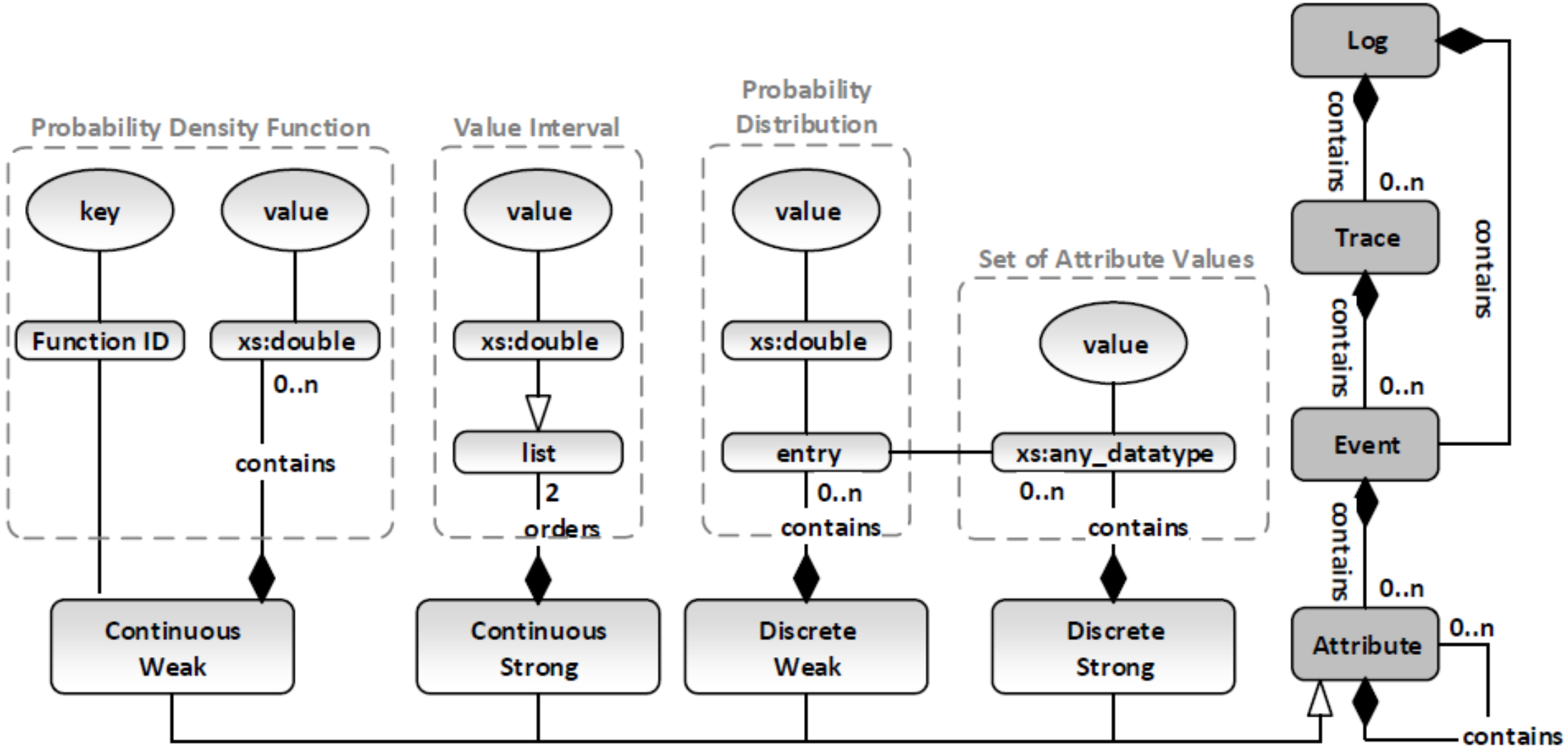
- Process discovery and conformance checking techniques are now available for uncertain event logs [1,2]

- An important practical problem remains: **how to represent uncertain traces in a computer, in such a way that they can be manipulated with ease?**

- We propose an extension of the XES data standard for uncertain event data representation

© Marco Pegoraro (RWTH Aachen University)

Uncertain Data XES Representation



© Marco Pegoraro (RWTH Aachen University)

Uncertain Data XES Representation - Example

Case ID	Event ID	Timestamp	Activity	Indeterminacy
ID192	e_1	2011-07-05	*NightSweats*	?
ID192	e_2	2011-07-08	{*PrimaryTP*, *SecondaryTP*}	
ID192	e_3	[2011-07-04, 2011-07-10]	*SpleenEnlargement*	

```
1  <trace>
2   <string key="concept:name" value="ID192" />
3   <event>
4    <string key="concept:name" value="NightSweats" />
5    <date key="time:timestamp" value="2011-07-05T12:00:00+00:00" />
6    <container key="uncertainty:discrete_strong">
7     <bool key="uncertainty:indeterminacy" value="true" />
8    </container>
9   </event>
10  <event>
11   <string key="concept:name" value="PrTP" />
12   <date key="time:timestamp" value="2011-07-08T12:00:00+00:00" />
13   <container key="uncertainty:discrete_strong">
14    <string key="concept:name" value="PrTP" />
15    <string key="concept:name" value="SecTP" />
16   </container>
17  </event>
18  <event>
19   <string key="concept:name" value="Splenomeg" />
20   <date key="time:timestamp" value="2011-07-07T12:00:00+00:00" />
21   <list key="uncertainty:continuous_strong">
22    <date key="time:timestamp" value="2011-07-04T12:00:00+00:00" />
23    <date key="time:timestamp" value="2011-07-10T12:00:00+00:00" />
24   </list>
25  </event>
26 </trace>
```


© Marco Pegoraro (RWTH Aachen University)

Uncertain Data XES Representation - Example

```
1  <trace>
2   <string key="concept:name" value="ID192" />
3   <event>
4    <string key="concept:name" value="NightSweats" />
5    <date key="time:timestamp" value="2011-07-05T12:00:00+00:00" />
6    <container key="uncertainty:discrete_strong">
7     <bool key="uncertainty:indeterminacy" value="true" />
8    </container>
9   </event>
10  <event>
11   <string key="concept:name" value="PrTP" />
12   <date key="time:timestamp" value="2011-07-08T12:00:00+00:00" />
13   <container key="uncertainty:discrete_strong">
14    <string key="concept:name" value="PrTP" />
15    <string key="concept:name" value="SecTP" />
16   </container>
17  </event>
18  <event>
19   <string key="concept:name" value="Splenomeg" />
20   <date key="time:timestamp" value="2011-07-07T12:00:00+00:00" />
21   <list key="uncertainty:continuous_strong">
22    <date key="time:timestamp" value="2011-07-04T12:00:00+00:00" />
23    <date key="time:timestamp" value="2011-07-10T12:00:00+00:00" />
24   </list>
25  </event>
26 </trace>
```


© Marco Pegoraro (RWTH Aachen University)

Conclusion

- Our XES extension proposal can **effectively represent uncertain event data and its meta-attributes**
- This enables analysis of uncertain log on XES-compliant platforms
- The extension is designed to allow for input, output and editing of uncertain data **on existing XES-compliant software** (**backwards compatible**)
- For more information: https://github.com/proved-py/proved-core/tree/An_XES_Extension_for_Uncertain_Event_Data/data

© Marco Pegoraro (RWTH Aachen University)

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Marco Pegoraro

pegoraro@pads.rwth-aachen.de

www.mpegoraro.net

@pegoraro_marco

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

1. 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.
2. Pegoraro, Marco, Merih Seran Uysal, and Wil MP van der Aalst. "Conformance checking over uncertain event data." *Information Systems* (2021): 101810.
3. 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.