Explainable Concept Drift in OCPM

Name - Aum Pathak

Roll no - MML2023006

Guide - Dr. Ranjana Vyas

Overview

Traditional Process Mining

Traditional process mining techniques focus on a single case or process instance at a time, simplifying the analysis but often missing out on the broader and more complex interactions in real-world environments.[2]

Object-Centric Process Mining

Object-Centric Process Mining (OCPM) is an advanced approach to process mining that captures the complexity of real-world processes by considering multiple object types and their interactions within a process.[2]

Benefits of OCPM

By adopting OCPM, organizations can visualize the intricate web of their processes, uncover performance bottlenecks, and improve decision-making based on a more realistic understanding of their operations.[2]

Concept Drift in Process Mining

1

Concept Drift

Concept drift is a significant challenge in process mining. It refers to the phenomenon where the underlying processes change over time, either abruptly or gradually, which can affect the accuracy and reliability of process models derived from event logs.

2

Challenges

Detection and analysing the concept drift in object centric process mining is challenging because the process itself is dynamic and viewing the process from different case notions is difficult to examine.

3

Importance of OCPM and Concept Drift

Incorporating both OCPM and concept drift detection into process mining frameworks enables a more resilient and adaptable approach to understanding complex business processes.[2]

Literature Review:

- There are **types of drifts**, including sudden, gradual, recurring, and incremental, each of which poses unique challenges for detection. Another challenge is **online vs. offline detection**. While many detection techniques are applied in post hoc (offline) analysis, real-time (online) detection is critical for providing timely responses to process changes as they happen.[3]
- Concept drift detection in process mining presents several challenges. One of the primary difficulties is **detection and characterisation**, which involves identifying when and where changes occur within a process model. These changes can affect different perspectives, such as control flow, timing, or resources.[3]
- Additionally, the absence of **standardised evaluation protocols** complicates the benchmarking and comparison of various drift detection methods, metrics are lacking. This is further increases the difficulty of **handling evolving environments**, where real-time process mining techniques need to adapt to rapidly changing conditions in business processes.[1]
- Several approaches to drift detection have been developed. **Statistical hypothesis testing** is commonly used to detect sudden and gradual drifts. **Trace clustering** is another method that groups traces based on behavioural similarities, making it particularly effective in detecting gradual drifts. [1]

Problem Statement

Process mining enables organizations to extract insights from event logs to discover and optimize processes. However, traditional methods assume static processes, while real-world processes are dynamic and involve multiple interrelated objects. Object-Centric Process Mining (OCPM) addresses this complexity by capturing interactions across various object types. A significant challenge in both traditional process mining and OCPM is **concept drift**, where processes evolve over time, affecting model accuracy. In object-centric logs, concept drift can affect different dimensions (e.g., control flow, timing), making it harder to detect and interpret. Developing a method to detect, characterise, and explain drifts is crucial for maintaining actionable insights for an organization.

Framework for Concept Drift Detection

Time series extraction:

- **Object Graph**: Represents object connections through shared events. Connected subgraphs represent process executions.
  - Flattening -
    - Deficiency-Event not associated with any object will be discarded.
    - Convergence-Event associated with multiple object causes duplications
  - Component extraction
    - Large cluster- Slow process execution
  - Leading Object extraction technique-
    - The user selects a "leading" object type (e.g., orders or items) to focus on.
    - For each object of the chosen leading type, a process execution is formed, with the leading object serving as the anchor for the execution.

ID	Activity	Order	Item	Timestamp	Cost
e_1	Place order	o1	i1,i2	03.03.2022 12:15	1
e_2	Pick item		i2	03.03.2022 14:21	3
e_3	Out of stock		i1	03.03.2022 14:57	4
e_4	Place order	o2	i3,i4,i5	04.03.2022 07:12	1
e_5	Reorder		i1	04.03.2022 08:45	2
e_6	Pick item		i3	04.03.2022 14:01	2
e_7	Pick item		i5	04.03.2022 14:01	2
e_8	Pick item		i4	04.03.2022 14:06	3
e_9	Ship order	o2	i3,i4,i5	04.03.2022 14:56	120
e_{10}	Receive payment	o2		06.03.2022 09:00	6
e_{11}	Receive payment	o1		06.03.2022 09:03	4
e_{12}	Reorder arrived		i1	07.03.2022 08:32	10
e_{13}	Ship order	o1	i1,i2	07.03.2022 08:01	80
e_{14}	Place order	o3	i6	10.03.2022 15:38	1
e_{15}	Pick item		i6	10.03.2022 16:59	3
e_{16}	Ship order	o3	i6	10.03.2022 17:07	40
e_{17}	Receive payment	o3		13.03.2022 09:01	4

Fig-1 Event log

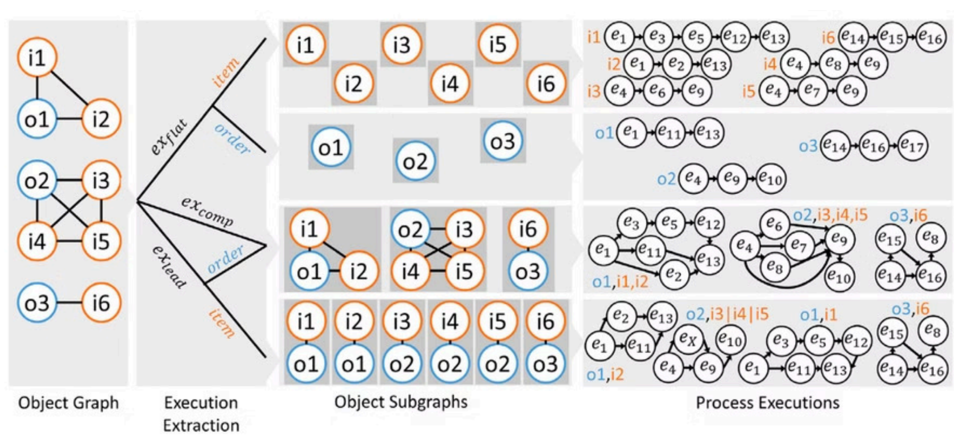

Fig-2 Process Executions

Event Log Segmentation:

- Segment event logs into **timeframes** (e.g., daily, weekly) to monitor process evolution.
- **Inclusion Functions**:
  - Start: Includes executions starting in the window.
  - End: Includes executions ending in the window.
  - Contained: Includes executions fully within the window.
  - Spanning: Includes executions that overlap the window.

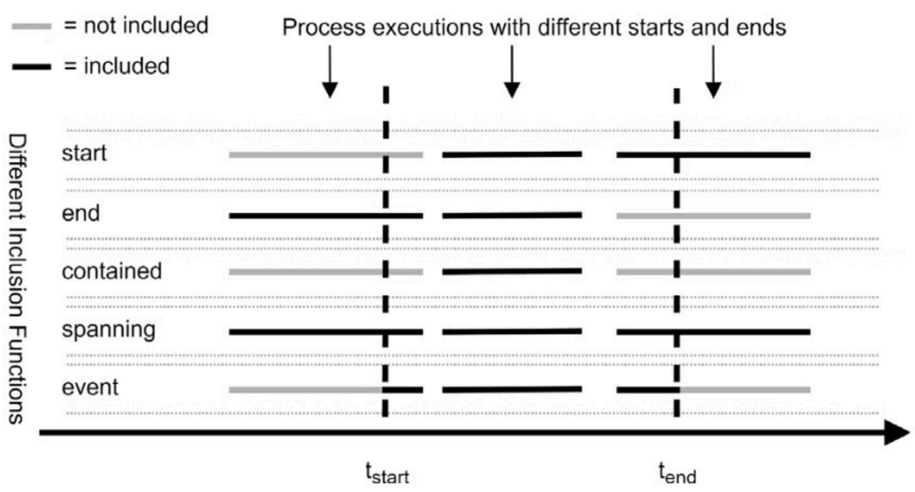

Fig 2-Time Series Segmentation using Different Inclusion Functions

Feature Calculation:

- Compute metrics like **average execution time** and **number of objects per event**.
- Features are concatenated over time to form the **time series** for monitoring process performance and changes.

Expected Outcome

- Identification of significant concept drifts within processes
- Development of a framework for detecting and characterising Drift
- Enhanced capability for organizations to monitor processes efficiently
- Improved responsiveness to process changes for better decision-making and optimization

References

1. Sato, D. M. V., De Freitas, S. C., Barddal, J. P., & Scalabrin, E. E. (2023). A survey on concept drift in process mining. *ACM Computing Surveys*, *54*(9), Article 189. https://doi.org/10.1145/3472752
2. van der Aalst, W. M. P. (2023). Object-centric process mining: Unraveling the fabric of real processes. *Mathematics*, *11*(12), 2691. https://doi.org/10.3390/math11122691
3. Bose, R. P. J. C., van der Aalst, W. M. P., Žliobaitė, I., & Pechenizkiy, M. (2022). Dealing with concept drifts in process mining. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(1), 154-171. https://doi.org/10.1109/TNNLS.2013.2278313