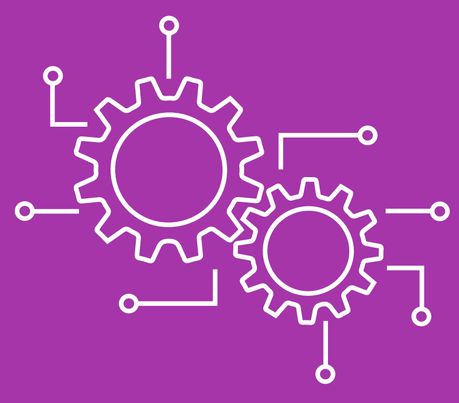
Learned Components in Complex Event Processing Systems

Research Proposal

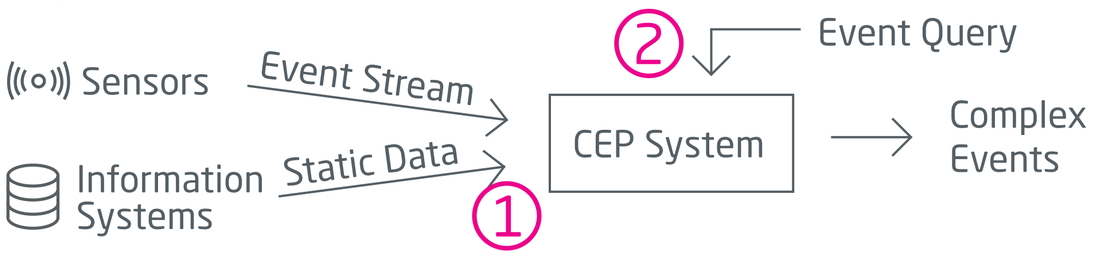


Abstract

Complex Event Processing (CEP) has to process enormous event stream dimensions. In CEP, the ability to react as fast as possible to changing situations provides value. **Learned Components** have been show to improve DBMS when replacing classical, hand-tuned components. The author would like to present a research proposal that enhances *Remote Data Integration* as well as *Event Query Execution* in CEP systems to reduce latency.

Complex Event Processing [1]

... aims to **detect patterns on event streams** with **predefined queries**. The event matching context may depend on **external static data**. [3]



1 Remote Data Integration

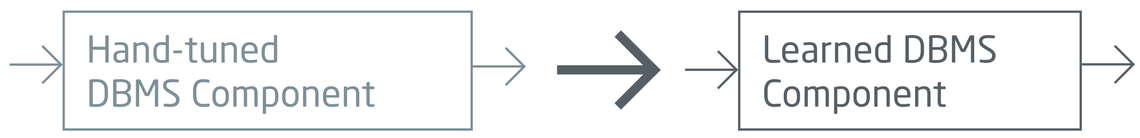
- EIRES framework [3]
  - Uses cost model to estimate the expected utility of remote data elements for query evaluation
  - Performs *Prefetching* or *Lazy Evaluation* depending on utility as well as *Caching*

2 Event Query Execution

- Most Common: Automata-based Execution
  - Transforms query into automaton that stores partial matches
  - Can be parallelized

Learned Components in DBMS [2]

... **improve complex components** and **reduce manual engineering effort** in DBMS by replacing classical DBMS components with ML models.



Workload-driven Learning

Learning 1.0

- Run queries, collect results, train model

— Expensive and repeated execution of training queries

Learning 2.0

Data-driven Learning

- Learn data distribution

\+ Fast updating

— Limited to tasks that do not consider workload

Zero-Shot Learning

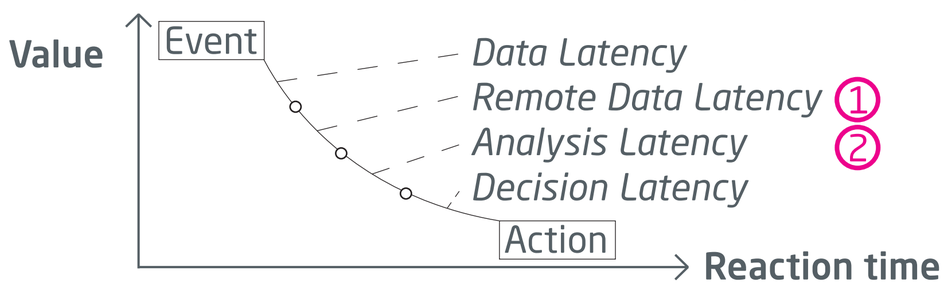
- Learn models that generalize to new DBs and workloads
- Transferable representation and sufficient training data needed

\+ No updating required

Research Proposal

Problem

Efficiency in CEP matters. The system only has limited time to process. The later we react to an event, the less the value of our action (e.g. credit card fraud detection).



Related Work

- [4] showed that rule-based classifiers can be used for detecting rule patterns in CEP systems.
- [5] created *IL-Miner* that discovers patterns from labeled event data by learning abstractions and correlation conditions.
- [6] used ML for pattern creation to detect security attacks in IoT.

→ **Event patterns can be learned**

- [7] summarized opportunities for ML in CEP systems.

Research Goal

Use *Learned Components* to lower technical and human latency by reducing *Remote Data* and/or *Analysis Latency*.

Solutions

1 Reduce Remote Data Latency

- EIRES performs *Utility Estimation* as counts over sliding windows [3]
- Use ML models to estimate utilities

→ **Reduce technical and human latency: improved cost models and no manual cost model design**

2 Reduce Analysis Latency

- Use Learned Component to build partial matches of a query

→ **Reduce technical latency: no automata-based execution**

- Apply *Zero-Shot Learning* to find partial matches of complex queries on event streams while being trained on other streams (instead of expensive *Workload-driven Learning* per stream)

→ **Reduce technical latency: pre-train generalized models**

- Replace hand-tuned queries completely with *Learned Component*

→ **Reduce technical and human latency: no manual engineering**

Author

Vincent Xeno Rahn
Master Student

Lecture Series on Database Research
Winter Term 2021/2022

Digital Engineering | University of Potsdam
February 2022

References

[1] Weidlich, Matthias. "Efficient Complex Event Processing." ***Lecture Series on Database Research 2021/22***, 18.01.2022.

[2] Binnig, Carsten. "Learned DBMS Components 2.0." ***Lecture Series on Database Research 2021/22***, 14.12.2021.

[3] Zhao, Bo; van der Aa, Han; Nguyen, Thanh Tam; Nguyen, Quoc Viet Hung; Weidlich, Matthias. *EIRES: Efficient Integration of Remote Data in Event Stream Processing*. Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21), 2021.

[4] Mehdiyeva, Nijat; Krumeicha, Julian; Enkeb, David; Wertha, Dirk; Loos, Peter: *Determination of Rule Patterns in Complex Event Processing Using Machine Learning Techniques*. Complex Adaptive Systems, Publication 5, 2015.

[5] George, Lars; Cadonna, Bruno; Weidlich, Matthias: *IL-Miner: Instance-Level Discovery of Complex Event Patterns*. Proceedings of the VLDB Endowment, Vol. 10, 2016.

[6] Roldán, José; Boubeta-Puig, Juan; Luis Martínez, José Luis; Ortiz, Guadalupe: *Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks*. Expert Systems with Applications, Volume 149, 2020.

[7] Wanner, Jonas; Wissucheka, Christopher; Janiesch, Christian: *A Review of Real-time Data Analytics for the Industrial Internet of Things*. Enterprise Modelling and Information Systems Architectures, Vol. 15, 2020.

All graphics based on the lectures and modified by the author.

Hasso Plattner Institut
Digital Engineering · Universität Potsdam