NeuroFlinkCEP

Neurosymbolic Complex Event Recognition Optimized across IoT Platforms

Authors: Ourania Ntouni, Dimitrios Banelas, Nikos Giatrakos
Technical University of Crete

ATHENA Research & Innovation Information Technologies

VLDB 2025 WELCOME TO LONDON

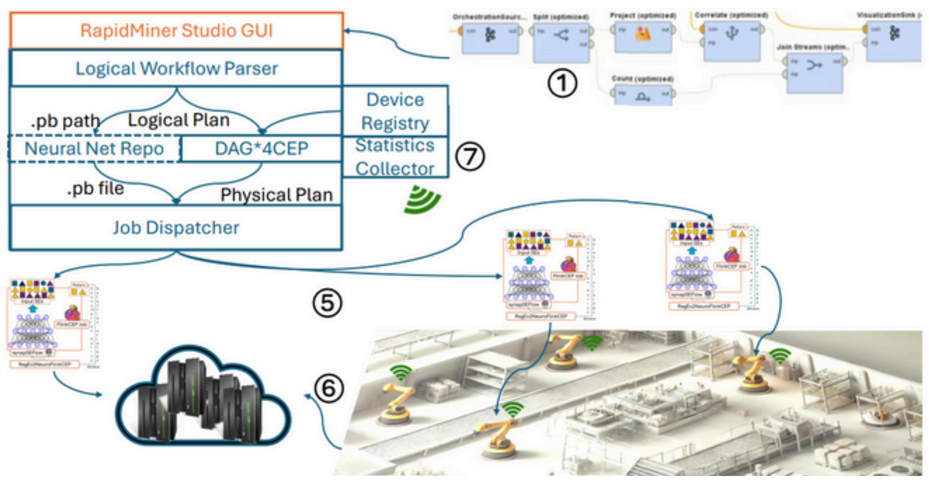
Problem & Gap

- **Neural Net** (aka sub-symbolic representation): From raw streams to Simple Events (SE) → symbols with learned instead of crisp definitions
- **Symbolic Representations**: Rules/Patterns expressing Complex Business Events (CEs) of interest
- **Neurosymbolic CER:** Neural Inference for SEs+ Patterns for CEs

***Challenge**: Neurosymbolic engine that scales (a) with the volume and velocity of rapid streams (b) across IoT platform*

From extended RegEx → auto-generated FlinkCEP jobs per device, with embedded neural inference and an optimizer that places operators across cloud/edge.

Architecture



Optimizer: What's new in DAG*4CER

- Pattern Decomposition
- Early Filtering
- Reordering
- Pushing Predicates Upstream

Demo Scenarios

Robotic Scenario

Robots Movements → Classes as SEs → FlinkCEP → Collision Recovery, Station Skipping, Prolonged Stop, Successful Stop

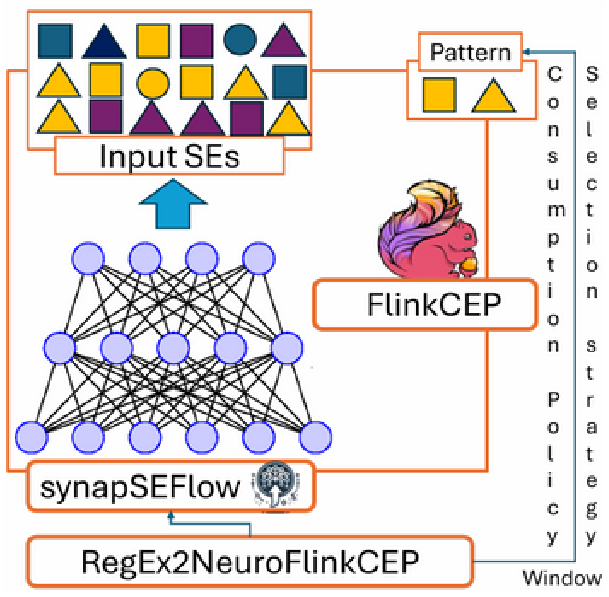
Telecom Scenario

Call Records → Classes as SEs → FlinkCEP → Long Calls at Night, Frequent Long Calls, Frequent Long Calls at Night, Frequent Each Long Call

CEs for detection

Contribution

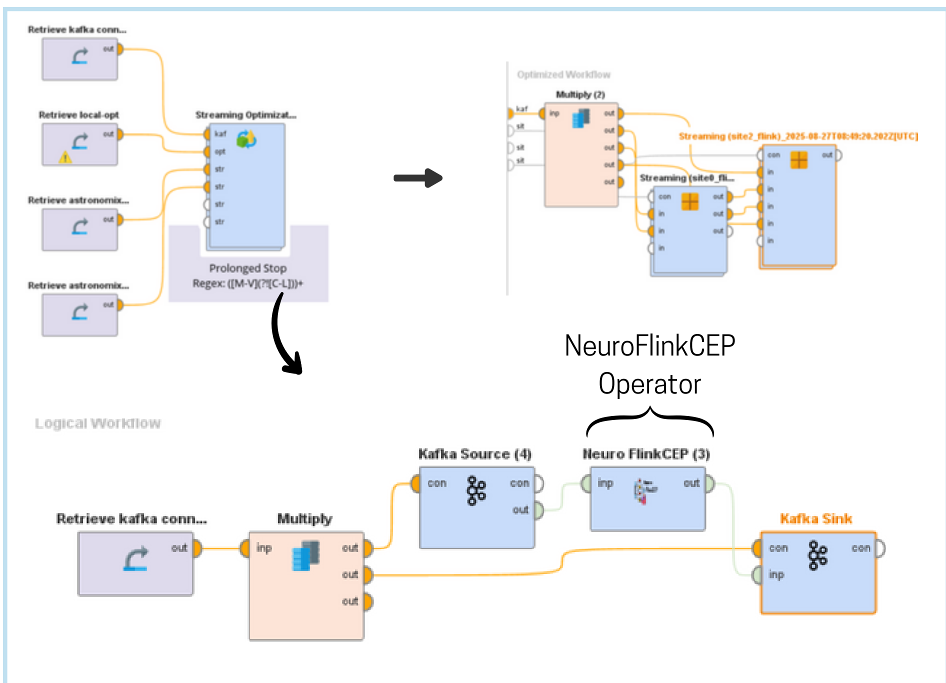
- First **neurosymbolic CER framework** on a state-of-the-art Big Data platform (Flink) with IoT-aware optimization
- **synapSEflow**: embed TensorFlow (.pb) models inside Flink jobs
- **RegEx2NeuroFlinkCEP**: compile extended RegEx patterns into FlinkCEP code (with window, selection, consumption policies).
- **DAG*4CER Optimizer**: extends DAG* with CER-specific rewrites for optimal edge/cloud placement
- Integration in RapidMiner Studio, no-code authoring.



User Experience

1. **Design** Logical workflows with NeuroFlinkCEP operators
2. **Optimize** physical workflow
3. **Human in the Loop**, Modify workflows before submitting
4. **Deploy** across the Cloud to Edge Continuum
5. Visual Analytics Dashboards

Care for manual physical plan creation?



DataSets

Robotic Stream

Record format: robotID, time, px/py/pz, vx/vy, idle, linear, rotational, Deadlock_Bool, RobotBodyContact

Neural net: predicts goal_status → mapped to simple events A–V (e.g., moving/stopped @ station). These feed CEP patterns

Simple events: A–V = collision (A), stopped-unknown (B), moving to 0–9 (C–L), stopped at 0–9 (M–V)

Telecom Stream

Record format: date/time, caller/callee, direction, charge, duration

Neural net: multi-label for A/B/C (premium, night, expensive) — can co-occur. CEP then derives E via aggregates.

Simple events: A=Call made to a premium location, B=Call made during night hours, C=Long call, E=Sum of call durations exceeds 60 minutes

Intrested for more?

NeuroFlinkCEP Website Videos and more info

Github Repository

CREXDATA Critical Action Planning over Extreme-Scale Data

EVENFLOW