

Pi-CEP: Predictive Complex Event Processing using Range Queries over Historical Pattern Space

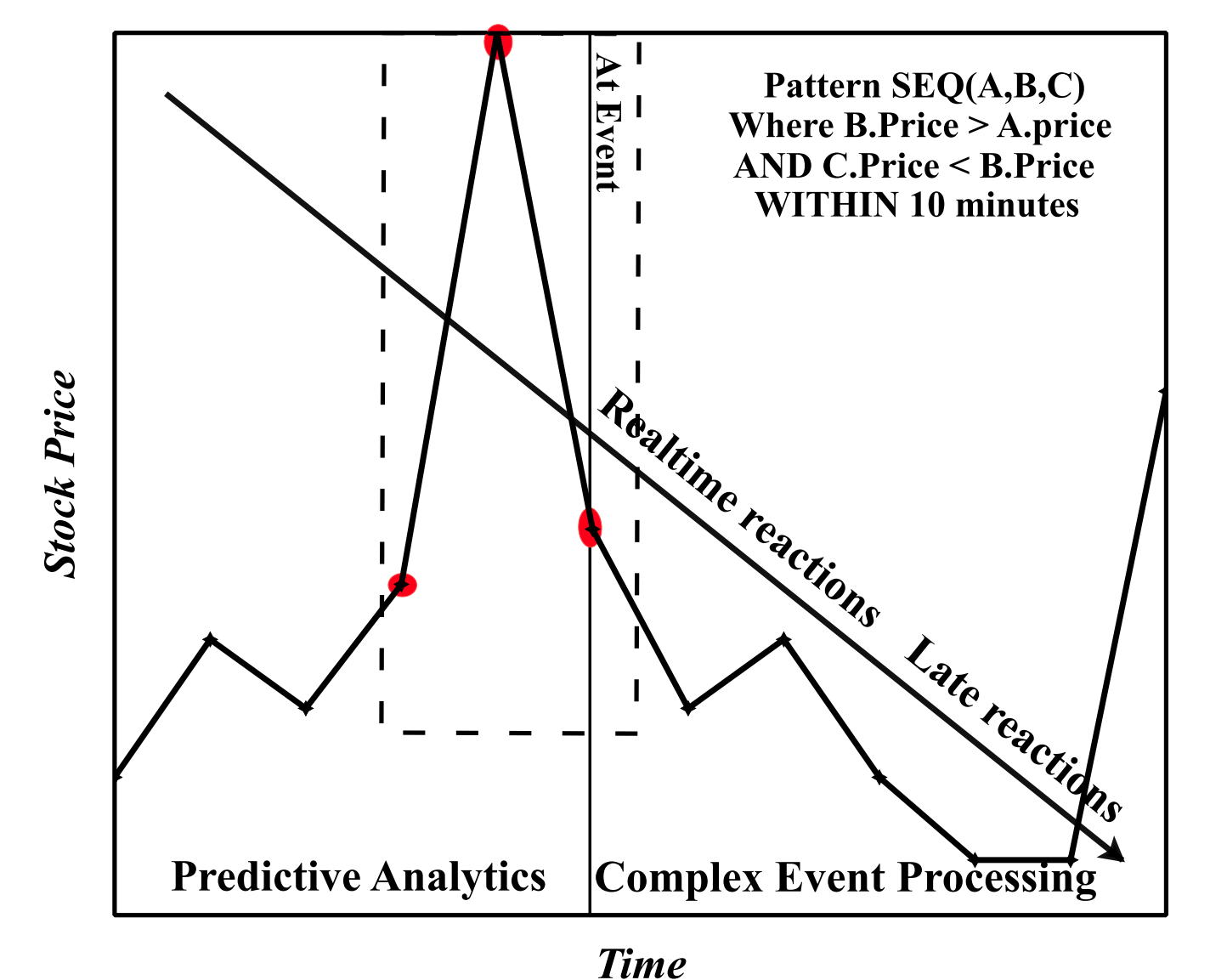
S. Gillani, A.Kammoun, K.Singh, J. Subercaze, C. Gravier, J.Fayolle, F. Laforest
 {firstname.lastname}@univ-st-etienne.fr

Problem Statement

- **Analytics Systems** are moving towards proactive computing and **Predictive Complex Event Processing (CEP)** constitutes the next evolution: it predicts future potential complex patterns.
- **Sequential Pattern Mining Algorithms** are computationally expensive in term of runtime/memory and may not be applied to complex types of data.
- Techniques such as **Markov models** and their variations are impractical in this context, since they suffer from demoting older patterns due to runtime memory constraints.
- With above techniques the historic information is lost without any summarisation technique.

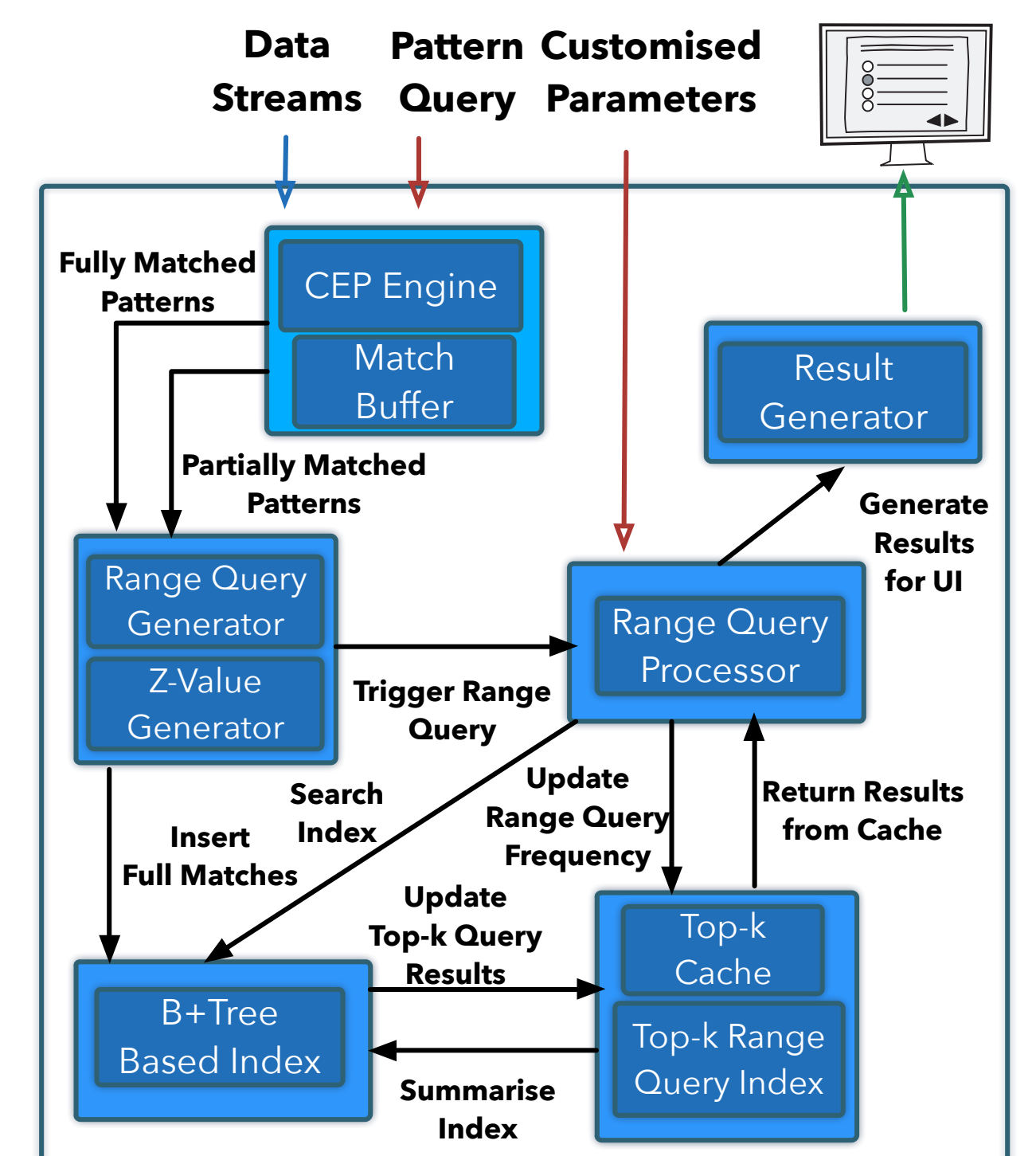
Proposal

- **Our goal** is to design a generic, noise tolerant and efficient predictive approach supporting multidimensional patterns.
1. We map multi-dimensions to one: use z-curve and sort them using B+Tree.
 2. We perform efficient range querying for partially matched (PM) patterns.
 3. We include K Nearest Neighbors patterns.
 4. We summarise full matches enclosed by the most infrequent range queries.



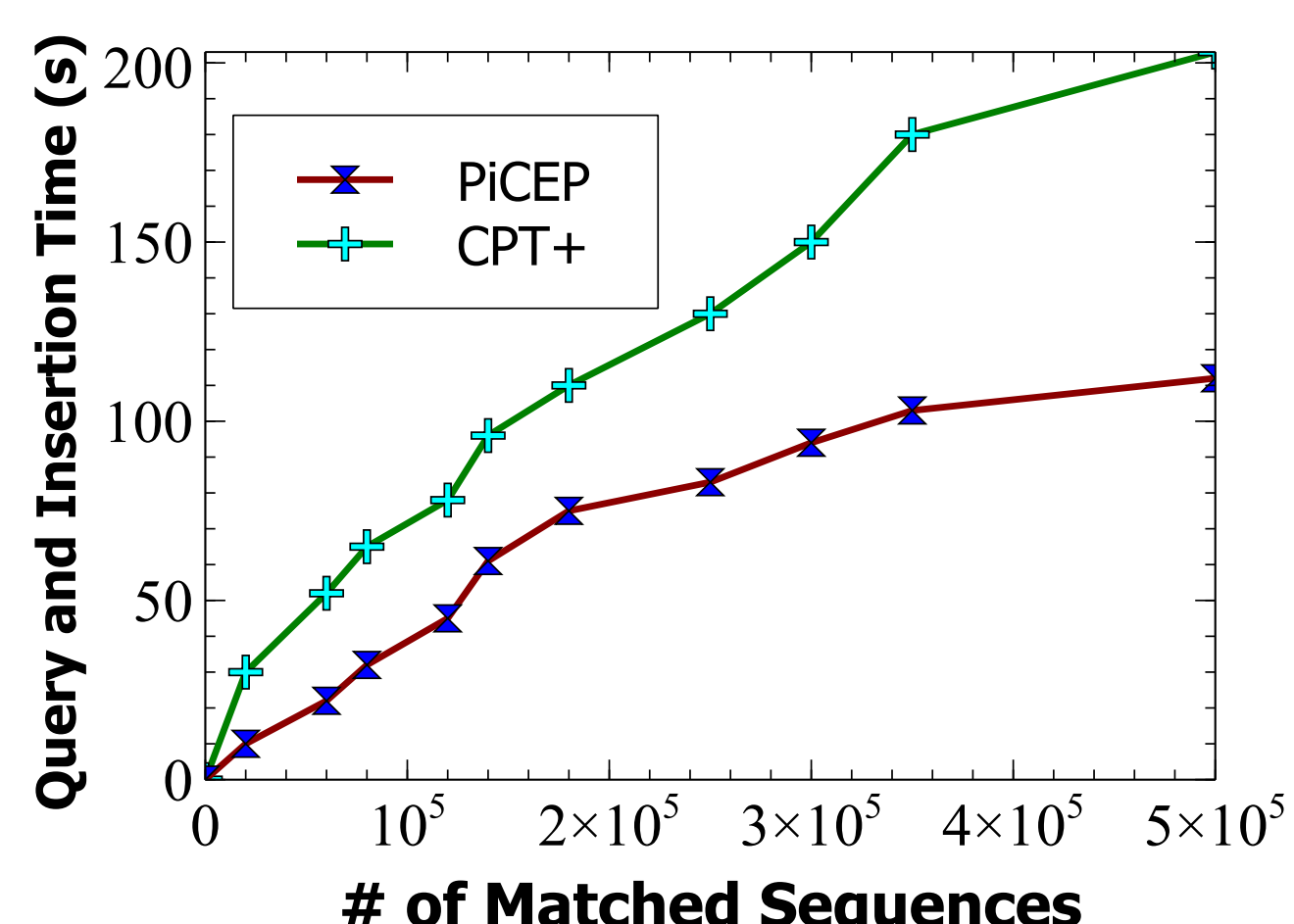
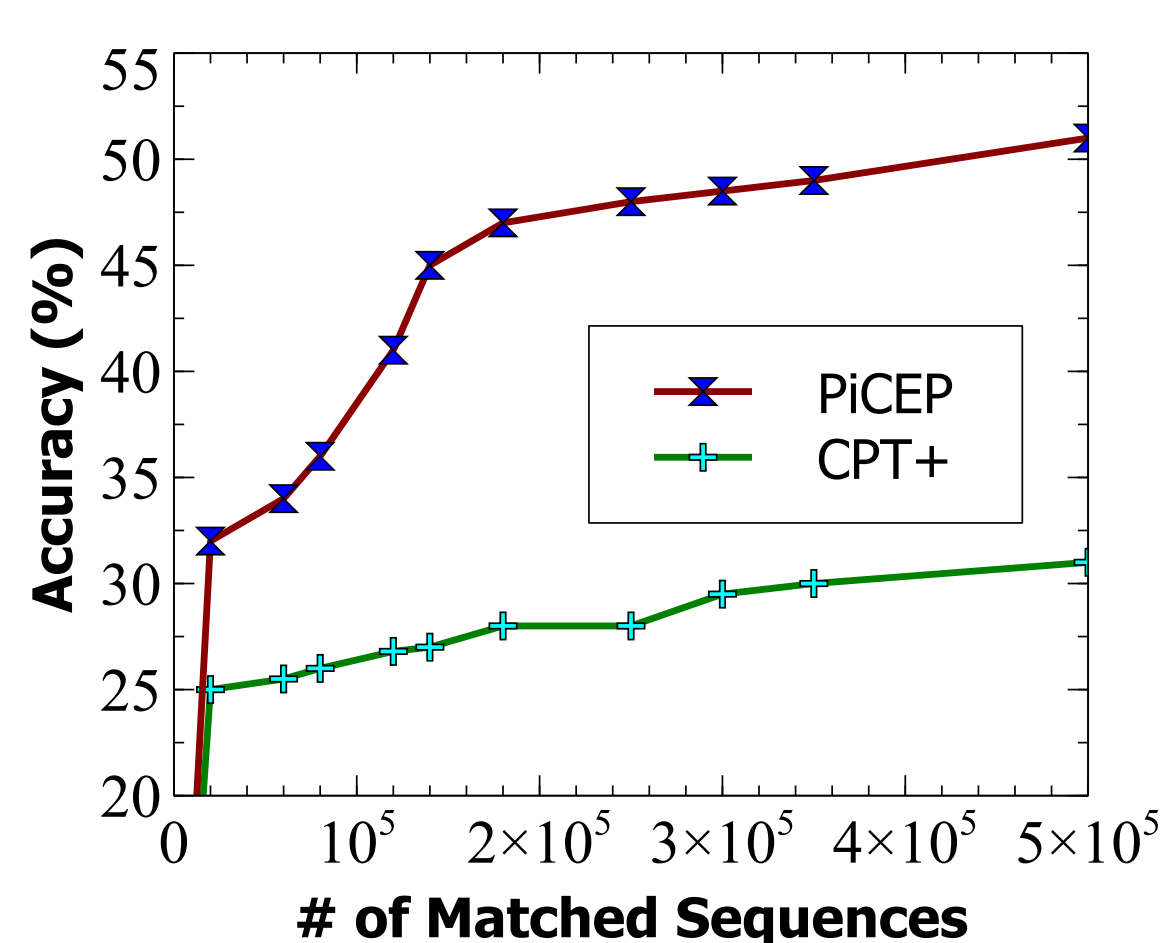
Pi-CEP

- Pi-CEP leverages two key observations:
 - the historical matches can give an expert view of the future matches;
 - summarise older matches according to their relevance (most infrequent pattern).
- We use a N-dimensional database called *historical space* \mathcal{H}_s . We employ range querying on \mathcal{H}_s for the partially matched patterns to determine the predictive events.
- We use the Z-order curve for N-Dimension space encoding, which preserves the proximity of the full matches, while leveraging the effectiveness of linear data structures (such as B+tree) for range queries.
- Our techniques can be used for the general purpose of multidimensional pattern prediction.



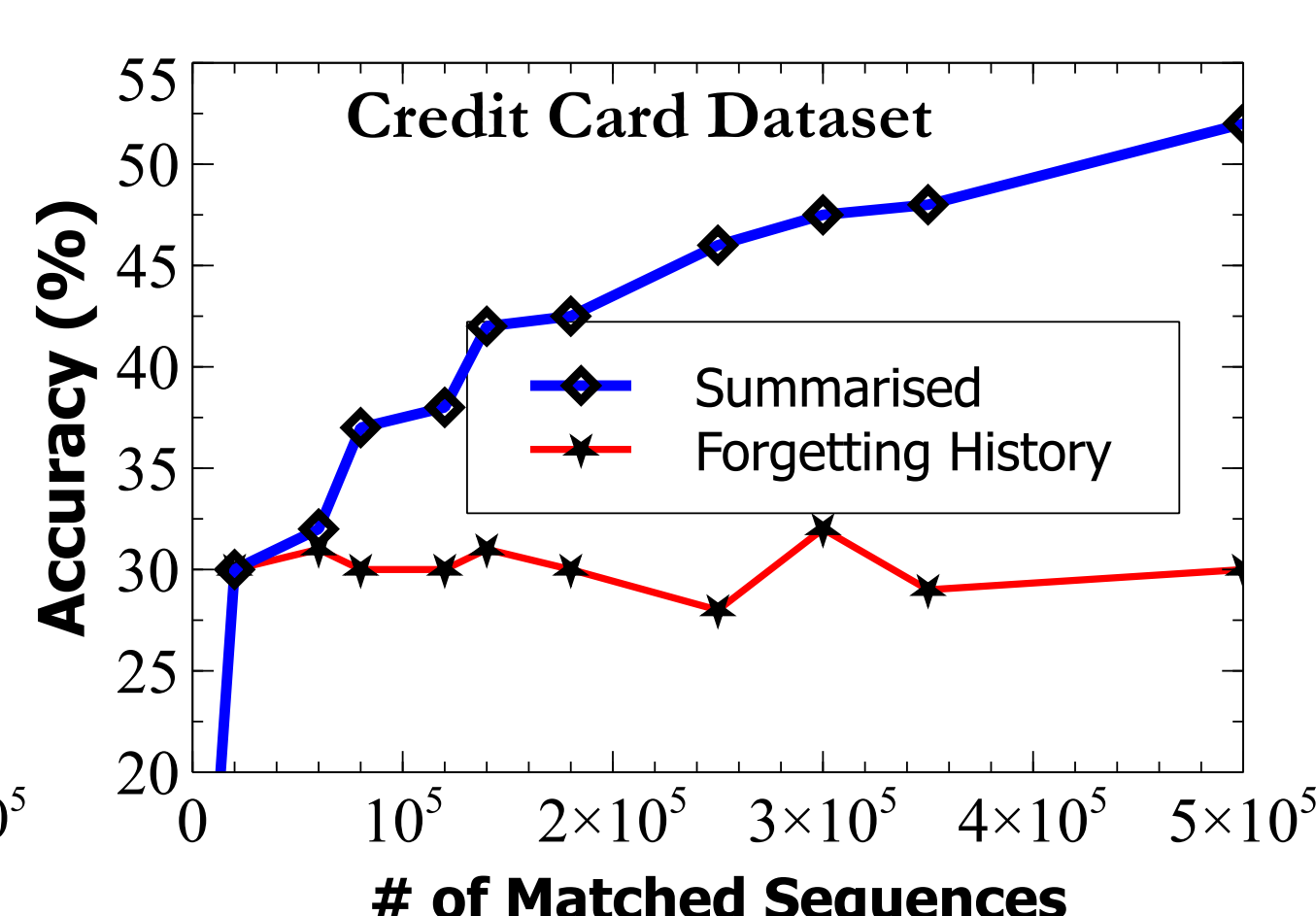
Experimental Evaluation

Comparison with CPT+



Accuracy and query and insertion time performance on credit card dataset

Summarisation vs. forgetting history



Accuracy of prediction algorithms with summarisation versus without

