

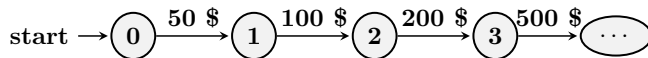
Event Forecasting with Pattern Markov Chains

Elias Alevizos, Alexander Artikis, George Paliouras

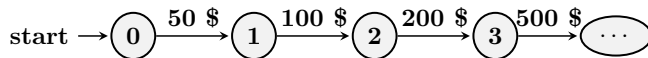
Complex Event Recognition lab,
Institute of Informatics & Telecommunications
National Centre for Scientific Research "Demokritos"

<http://cer.iit.demokritos.gr/>

Motivation

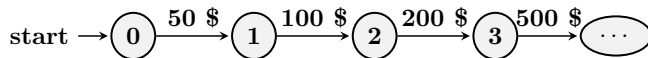


Motivation



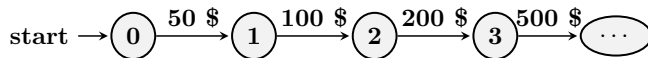
- ▶ Is this a fraud?

Motivation



- ▶ Is this a fraud?
- ▶ How long will it last?

Motivation



- ▶ Is this a fraud?
- ▶ How long will it last?
- ▶ With what probability?

Online Probabilistic Complex Event Forecasting

- ▶ Patterns defined as regular expressions.
- ▶ Consume streams of events and forecast when a pattern is expected to be fully matched.
- ▶ Revise forecasts to reflect changes in the state of the pattern.
- ▶ Remember “arbitrarily” long sequences.

Assumptions

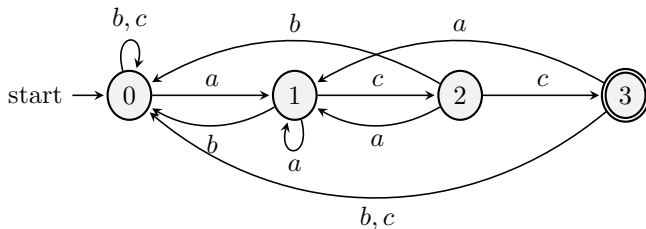
- ▶ Selection strategy: *(partition)-contiguity*.
- ▶ Stream generated by a m -order Markov process.
- ▶ Stream stationary.
- ▶ A forecast reports for how many *transitions* we will have to wait until a full match.

Regular Expression \rightarrow Pattern Markov Chain

$$R = a \cdot c \cdot c.$$

$$\Sigma = \{a, b, c\}.$$

No memory.

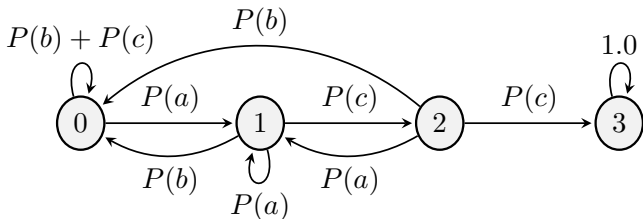
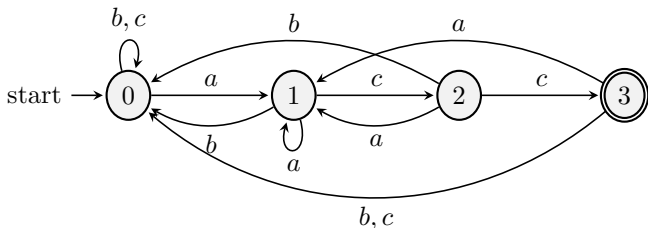


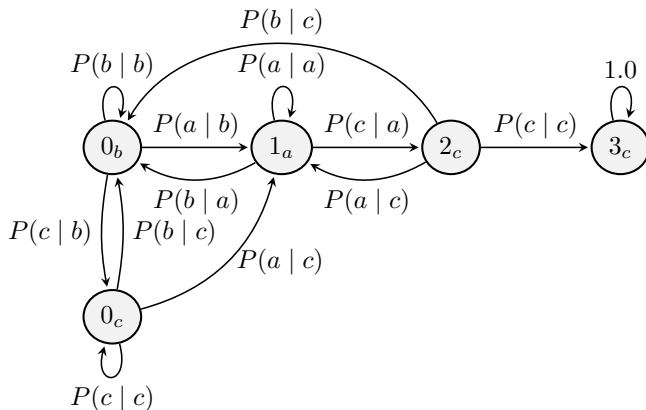
Regular Expression \rightarrow Pattern Markov Chain

$$R = a \cdot c \cdot c.$$

$$\Sigma = \{a, b, c\}.$$

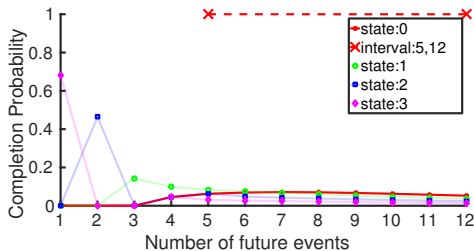
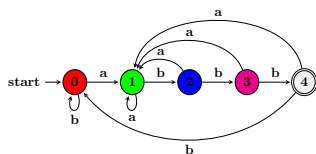
No memory.



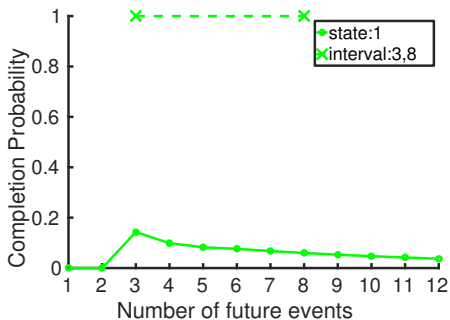
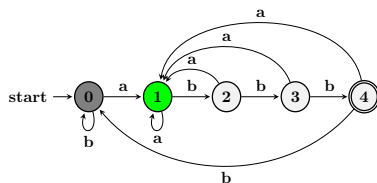
Regular Expression \rightarrow Pattern Markov Chain

Waiting-Time and Forecasts

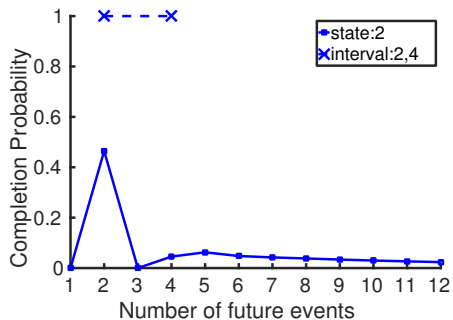
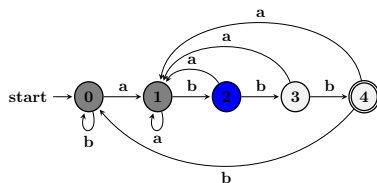
- ▶ Warm-up period to learn distributions.
- ▶ Set a threshold, e.g., $P_{forecast} = 50\%$.



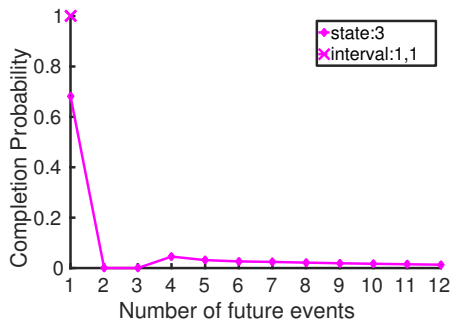
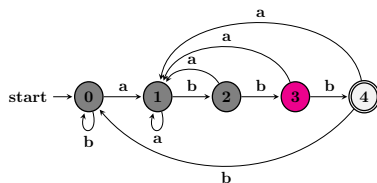
Example: $R = a \cdot b \cdot b \cdot b$.

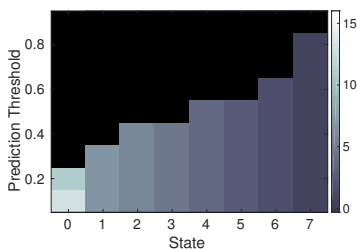
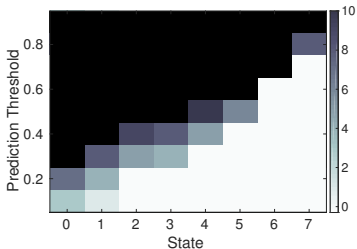
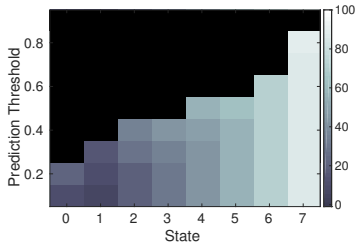


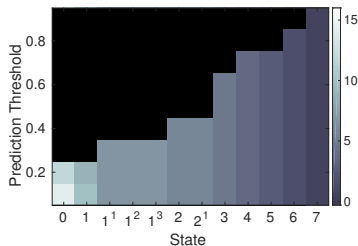
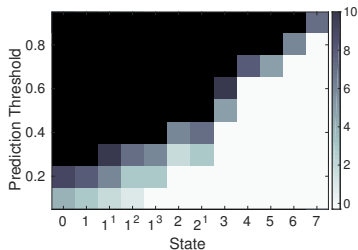
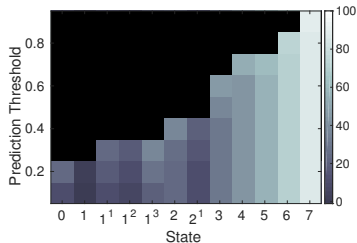
Example: $R = a \cdot b \cdot b \cdot b$.



Example: $R = a \cdot b \cdot b \cdot b$.



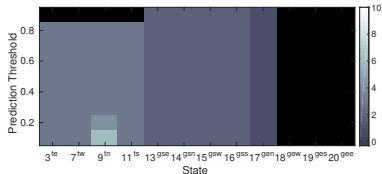
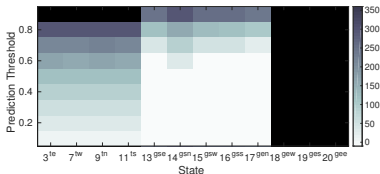
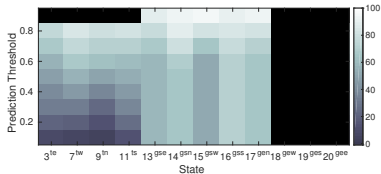
Credit Card Fraud Management: Real Dataset ($m = 1$).

Credit Card Fraud Management: Real Dataset ($m = 3$).

Maritime Monitoring: Real Dataset.

$$R = \text{Turn} \cdot \text{GapStart} \cdot \text{GapEnd} \cdot \text{Turn},$$

where $\text{Turn} = (\text{TurnNorth} + \text{TurnEast} + \text{TurnSouth} + \text{TurnWest})$



Summary & Future Work

- ▶ Contributions:
 - ▶ Regular expressions as opposed to sequential patterns.
 - ▶ Forecasts with guaranteed precision, if Markov process.
 - ▶ Useful forecasts even in applications where we do not know beforehand the stream properties.

Summary & Future Work

- ▶ Contributions:
 - ▶ Regular expressions as opposed to sequential patterns.
 - ▶ Forecasts with guaranteed precision, if Markov process.
 - ▶ Useful forecasts even in applications where we do not know beforehand the stream properties.

- ▶ Future work:
 - ▶ Constraints on event properties.
 - ▶ More selection strategies.
 - ▶ Support drift.
 - ▶ Forecasts that correspond to real time (not transitions).

Appendix

Validation tests

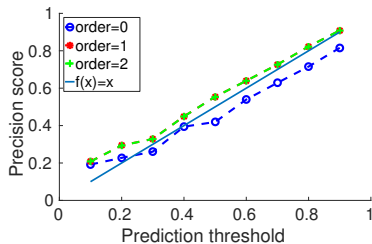
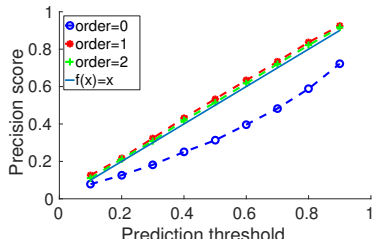
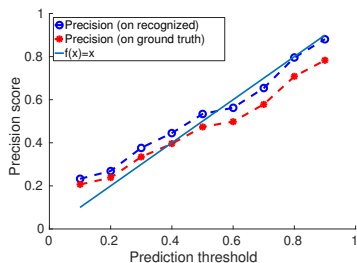
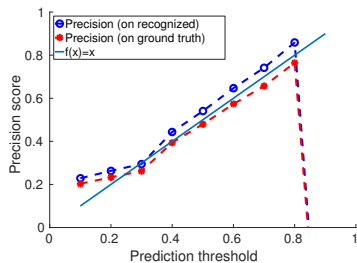
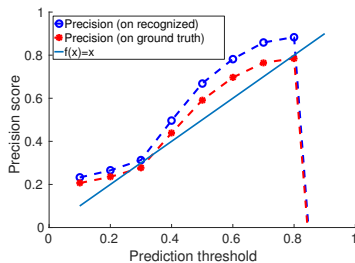


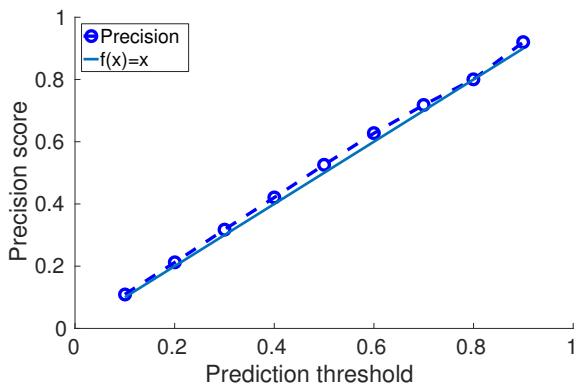
Figure: $R = a \cdot (a + b)^* \cdot c$



Credit cards (precision for $m = 1, 2, 3$)

Maritime (precision)

$$R = Turn \cdot GapStart \cdot GapEnd \cdot Turn$$



Maritime (precision)

$$R = \text{TurnNorth} \cdot (\text{TurnNorth} + \text{TurnEast})^* \cdot \text{TurnSouth}$$

