Getting to the CORE of Complex Event Recognition

Stijn Vansummeren
Data Science Institute, Applied Computer Science Laboratory,
Hasselt University, Diepenbeek, Belgium

Abstract

In this talk, I will give an overview of our recent work on complex event recognition.

2012 ACM Subject Classification Information systems → Query languages for non-relational engines; Information systems → Stream management; Theory of computation → Formal languages and automata theory

Keywords and phrases Complex Event Recognition, automata, enumeration-based query processing

Digital Object Identifier 10.4230/LIPIcs.TIME.2022.3

Category Invited Talk

Funding This research was supported by the Bijzonder Onderzoeksfonds (BOF) of Hasselt University (Belgium) under Grant No. BOF20ZAP02.

Acknowledgements The research surveyed in this talk is joint work with Marco Bucchi, Alejandro Grez, Martín Ugarte, Andrés Quintana, and Cristian Riveros.

1 Extended Abstract

Complex Event Recognition (CER for short) refers to the activity of processing high-velocity streams of primitive events by evaluating queries that detect *complex events*: collections of primitive events that satisfy some pattern. In particular, CER queries match incoming events on the basis of their content; where they occur in the input stream; and how this order relates to other events in the stream. CER has been successfully applied in diverse domains such as maritime monitoring, network intrusion detection, industrial control systems and real-time analytics, among others.

In this talk, I will survey our recent work on developing a formal framework for specifying and evaluating CER queries. This framework consist of a formal, core query language called *Complex Event Logic* (CEL) for specifying CER queries [4]. In contrast to previous proposals, CEL has a compositional and denotational semantics, and encompasses all operators that are considered "common base operators" in the literature. Using CEL, we have been able to get a better understanding of the relative expressiveness of these operators as well as the impact of common evaluation heuristics such as selection policies [3, 4].

The framework also consists of an automaton-based formal computational model for CEL, called *Complex Event Automata* (CEA). Using CEA, we have developed a novel evaluation algorithm for CEL that exhibits strong performance guarantees: under data complexity, the algorithm takes only constant time to process each input event [1]. This is in contrast to existing algorithms that take time proportional to the number of previously processed events, or the size of a time window. As I will explain, our algorithm processes each event in constant time by adopting the framework of enumeration-based query evaluation that is receiving increased attention in the database community [5]. Specifically, it maintains a data structure from which at any point in time all found complex events may be enumerated with so-called

output-linear delay. This means that the time required to output a new recognized complex event C is linear in the size of C, but independent of the number of already processed events or size of time window, and independent of the number of found complex events.

Our framework is implemented in the CORE complex event recognition engine [2]. CORE's firm formal foundation allows it to exhibit stable query performance, even for long sequence queries and large time windows, and outperform existing systems by up to five orders of magnitude on different workloads [1].

I will discuss the essential ideas behind CORE's query language and evaluation algorithm, as well as their limitations, and from these limitations discuss open questions relevant for the TIME community.

References

1 Marco Bucchi, Alejandro Grez, Andrés Quintana, Cristian Riveros, and Stijn Vansummeren. CORE: a complex event recognition engine. *Proc. VLDB Endow.*, 15(9):1951–1964, 2022. URL: https://www.vldb.org/pvldb/vol15/p1951-riveros.pdf.

2 The CORE COmplex event Recognition Engine. https://github.com/CORE-cer/CORE.

3 Alejandro Grez, Cristian Riveros, Martín Ugarte, and Stijn Vansummeren. On the expressiveness of languages for complex event recognition. In *23rd International Conference on Database Theory, ICDT 2020, March 30-April 2, 2020, Copenhagen, Denmark*, volume 155 of *LIPIcs*, pages 15:1–15:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICDT.2020.15.

4 Alejandro Grez, Cristian Riveros, Martín Ugarte, and Stijn Vansummeren. A formal framework for complex event recognition. *ACM Trans. Database Syst.*, 46(4):16:1–16:49, 2021. doi:10.1145/3485463.

5 Luc Segoufin. Enumerating with constant delay the answers to a query. In *Joint 2013 EDBT/ICDT Conferences, ICDT '13 Proceedings, Genoa, Italy, March 18-22, 2013*, pages 10–20. ACM, 2013. doi:10.1145/2448496.2448498.