Data Science Seminar 2021: Time Series Operators

Mourad Khayati, Alberto Lerner

https://exascale.info/courses/ds-xi/index

September 28, 2021

eXascale Infolab

Contents

Lecturers

- **Mourad Khayati**:
  - Senior researcher @ UNIFR since 2015
  - Ph.D. from University of Zürich
  - Research interests: Time Series repair, Large-scale data analytics, and missing values recovery.

- **Alberto Lerner**:
  - Senior researcher @ UNIFR since 2018
  - PhD from Paris Tech (France)
  - Previous Experience: IBM Research (T.J. Watson and Almaden), Google and MongoDB.
  - Research interests: hardware acceleration, distributed database systems and (time)-ordered data models.

Goals of the Seminar

▶ Gather in-depth knowledge of an advanced topic/paper in data science: **Time Series Operators**.

▶ Focus on one research paper/topic.

▶ Learn how to (critically) study a research paper.

▶ Describe a paper in a report, empirically evaluate the proposed technique(s) and present your findings in front of an audience.

Contents

Requirements and Organization

▶ Good understanding of algorithms, data structures and complexity.

▶ Write a report (first version and final version) of min 6 pages each describing the selected paper.

▶ Make a presentation of 20 min.

▶ Presentations will be split into two seminar sessions by the lecturer. Participants **must** attend **both sessions** (physically or remote).

▶ The seminar participation will be limited to **10 people**.

Contents

Final Grade

- ▶ Quality of the first and the last version of the report.

- ▶ Quality of the presentation (including Q&A).

- ▶ Replicability/Reproducibility experiments.

- ▶ Active participation in the discussions.

Contents

How to write the report

▶ The report[1] should reflect your understanding of the paper and not a rephrasing/summary of the paper.

▶ The structure of report might not follow the same structure of the paper.

▶ The report might contain your own small running example and counter example (illustrating how the proposed solution works).

▶ Plot figures that describe the proposed technique, redefine complex formulas, add algorithmic description, etc.

---

[1]https:
//cs.stanford.edu/~rishig/courses/ref/paper-reading-technical.pdf

How to prepare your presentation

▶ The presentation[2] should explain the proposed contribution(s) and should not summarize the report/paper: what do I need to describe in order to explain the proposed solution(s)?

▶ The presentation should follow a coherent structure and not necessarily the structure of the paper.

▶ A (toy) example helps to better explain the steps of the proposed solution(s).

---

[2]http://matt.might.net/articles/academic-presentation-tips

How to replicate/reproduce the experiments

▶ Rerun the experiments introduced in the paper. You can run extra experiments you deem appropriate.

▶ Use the datasets introduced in the paper (preprocessing might be needed).

▶ Apply different datasets from other sources with similar properties and size.

▶ Discuss the results (efficiency, precision, divergence, etc.).

Contents

How to proceed/1

▶ Select a paper from the set of proposed papers **by September, 30** via: `https://doodle.com/poll/hqftqu4cypcumhms?utm_source=poll&utm_medium=link`.

▶ The papers will be distributed in FCFS basis.

▶ Prepare the first draft of the report and send it two weeks before your talk (Have a look at *Report_example.pdf* on ILIAS).

▶ Time preferences for the two sessions?

How to proceed/2

▶ Meet with the lecturer 1 week before your presentation.

▶ Send your final report before the final deadline (hard deadline).

▶ All the deadlines will be posted on the seminar website:
  https://exascale.info/courses/ds-xi/index. Stay
  tuned!

List of Papers

Title	Conf.	Year	Supervisor
(1) Scalable, Variable-Length Similarity Search in Data Series: The ULISSE Approach	VLDB	2018	Khayati M.
(2) The Lernaean Hydra of Data Series Similarity Search	VLDB	2019	Khayati M.
(3) Data Series Progressive Similarity Search with Probabilistic Quality Guarantees	SIGMOD	2020	Khayati M.
(4) Massively-Parallel Change Detection for Satellite Time Series Data with Missing Values	ICDE	2020	Khayati M.
(5) Efficient Learning Interpretable Shapelets for Accurate Time Series Classification	ICDE	2018	Khayati M.
(6) Temporal Data Management – An Overview	Tutorial	2017	Lerner A.
(7) Optimization of Analytic Window Functions	VLDB	2012	Lerner A.
(8) Efficient processing of window functions in analytical SQL queries	VLDB	2015	Lerner A.
(9) Incremental computation of common windowed holistic aggregates	VLDB	2016	Lerner A.
(10) Building Advanced SQL Analytics From Low-Level Plan Operators	SIGMOD	2021	Lerner A.