Real-Time Streaming and Computing of Traffic Congestions for More Efficient and Safer Transportation

Robert Kemi, Sonia Yeh, Vincenzo Massimiliano Gulisano, Marina Papatriantafilou

Chalmers University of Technology - *June 2016*

Objectives

The congestion study aims to create new information that helps to reduce congestion and improves urban sustainability. This study has three main tasks:

1) Establish "typical" congestion patterns for cities at detailed spatial and temporal resolution;
2) Develop real-time notifications of abnormal congestions that deviate from the "typical" traffic patterns;
3) Make predictions of real-time congestion patterns (spread and durations of congestions) when an accident occurs.

Data

Real-time traffic data can be obtained from HERE Maps Traffic Application Programming Interface (API). HERE gathers real-time location and speed data collated from GPS enabled mobile devices in-vehicle. Minute-by-minute speed information can be collected for any point on the map as well as other information including time, location (latitude and longitude), the direction of the traffic (+, -), free-flow speed, speed limit, and incidents. Traffic incidents are classified according to type (accident, congestion, construction, etc), status, criticality, profile, or start and end times.

Method

Real-time traffic information for selected cities was processed and analyzed. Given the large amount of data, choosing a suitable framework and/or platform will be needed for streaming, processing and maintaining user-defined statistics efficiently as specified by the project. This is a game-changing aspect of this project. Existing studies which rely on static datasets collected over limited periods of time (and later analyzed) usually result in knowledge that is also limited to such period, which might be non representative of the behavior later observed in the live urban environment. In contrast to this, this project will leverage data streaming in order for the analysis techniques to be developed to continuously analyze live data, thus providing up-to-date knowledge in a continuous and real-time fashion.

Average and Maximum congestion

Background congestion information is established and continuously updated. Figures 1 and 2 show the maximum congestion indexes for 475 locations for Gothenburg, Sweden, April 21 – May 5 2016.

Fig 1. Box plots of avg and max congestion indexes for 475 locations in Gothenburg, Sweden (April 21 – May 5 2016)

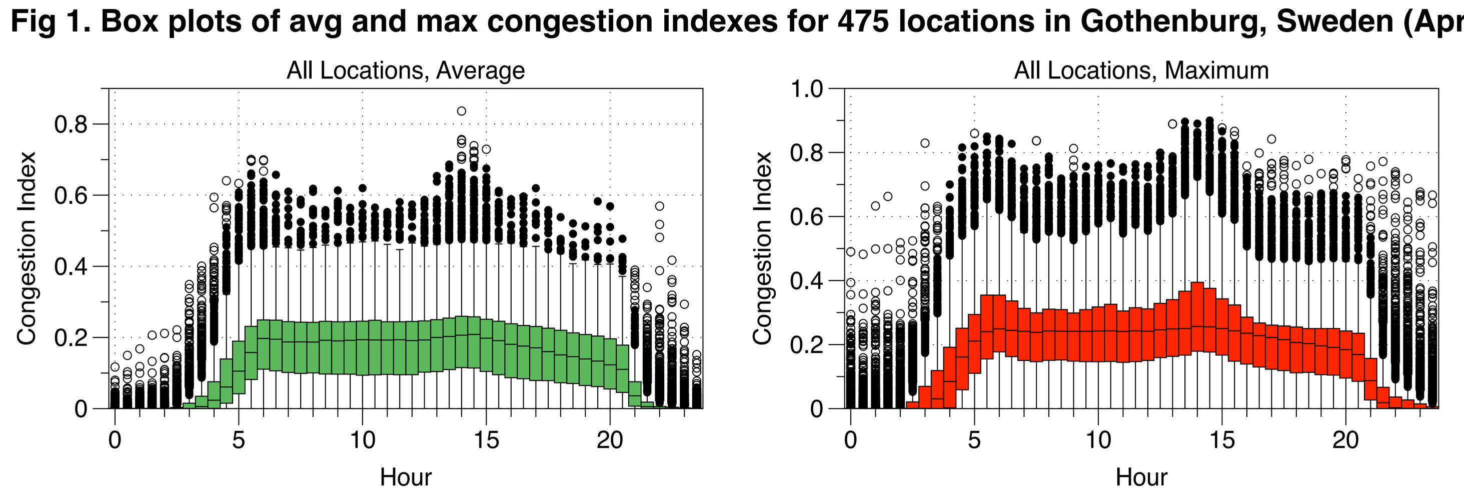
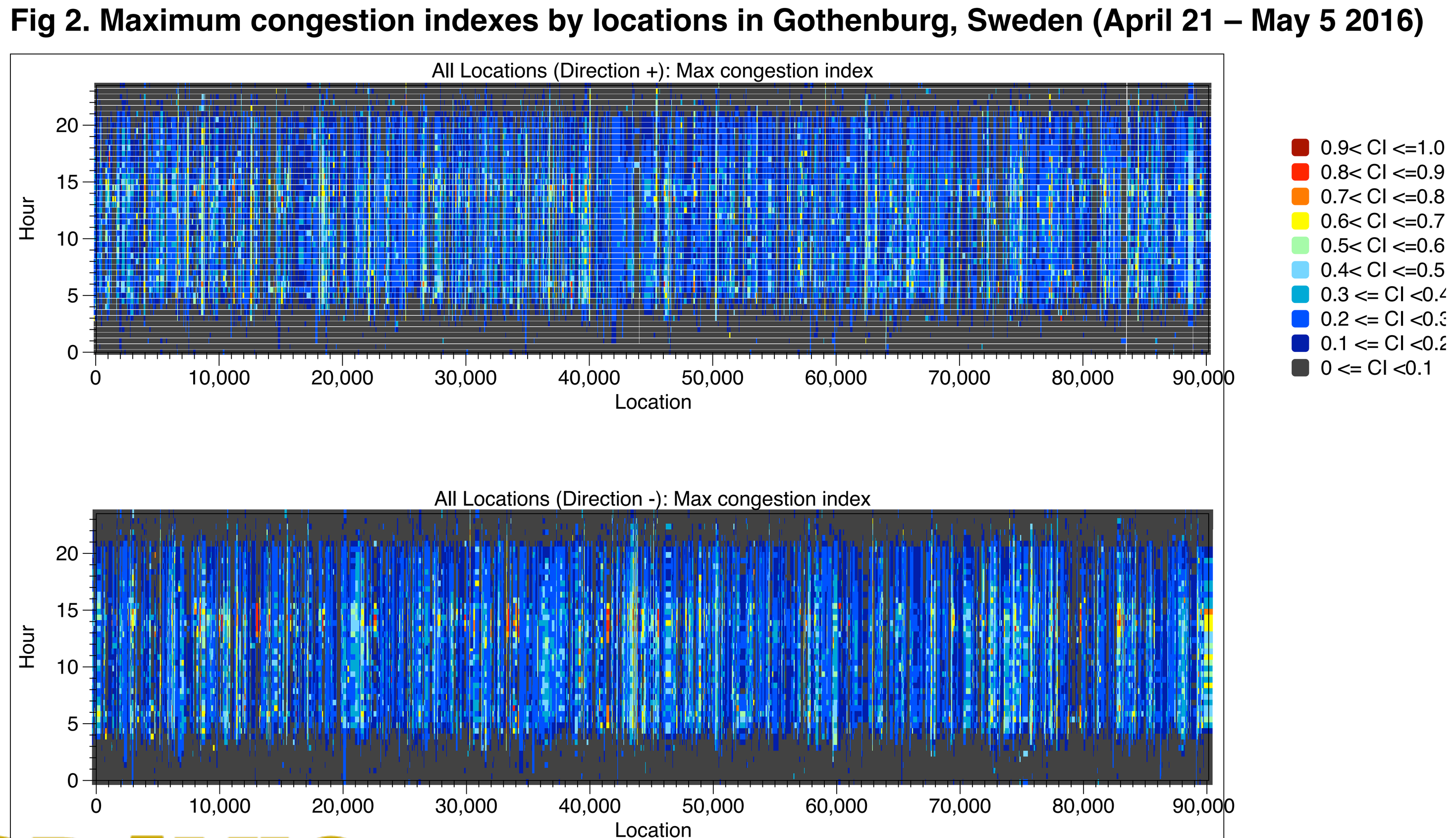


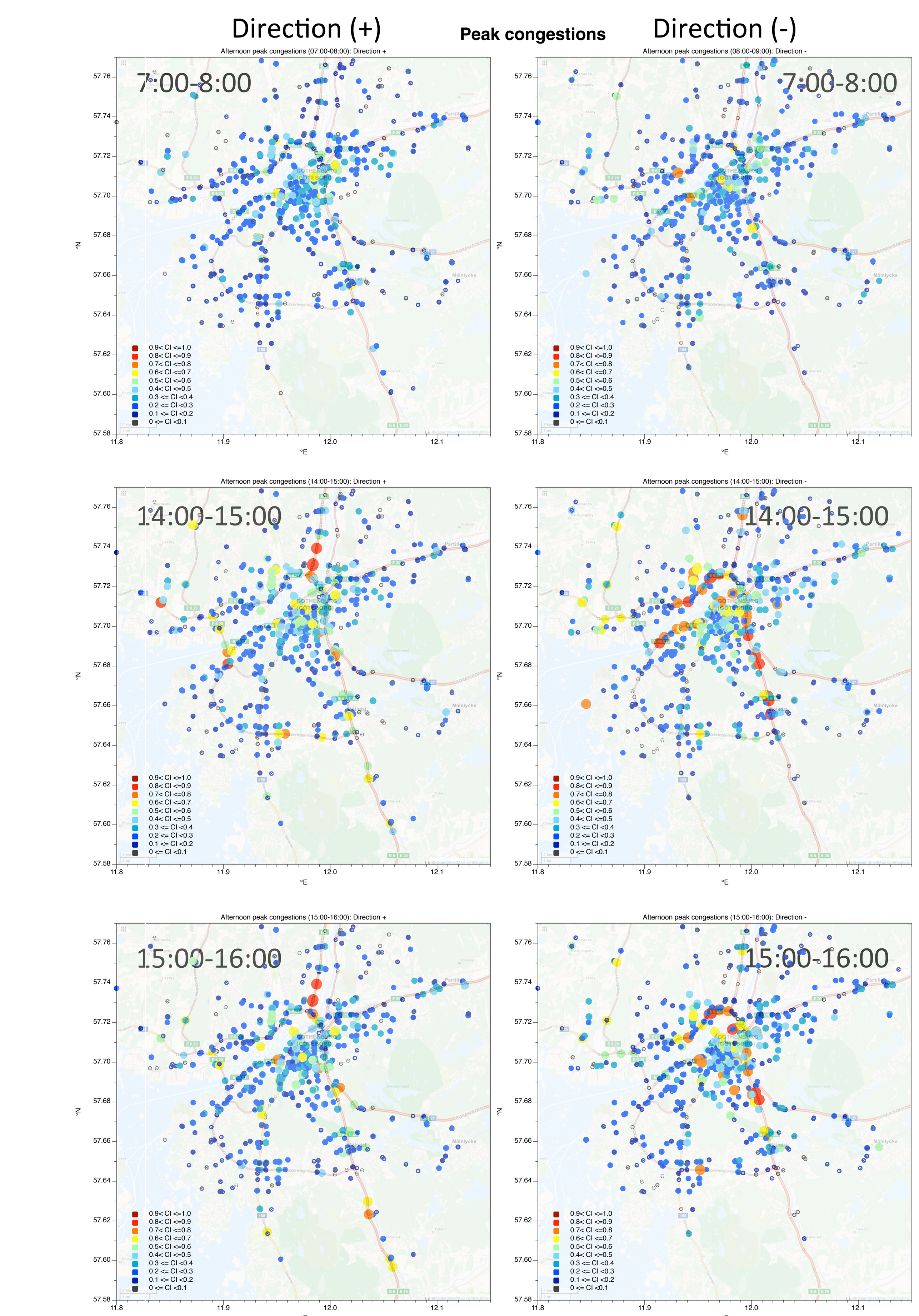
Fig 2. Maximum congestion indexes by locations in Gothenburg, Sweden (April 21 – May 5 2016)



Congestion Index

Congestion Index (CI)= $\frac{free\ flow\ speed - actual\ speed}{free\ flow\ speed}$ and is calculated for each location, time of the day, and direction (+,-). The sampling frequency can range from seconds, minutes, or every hour.

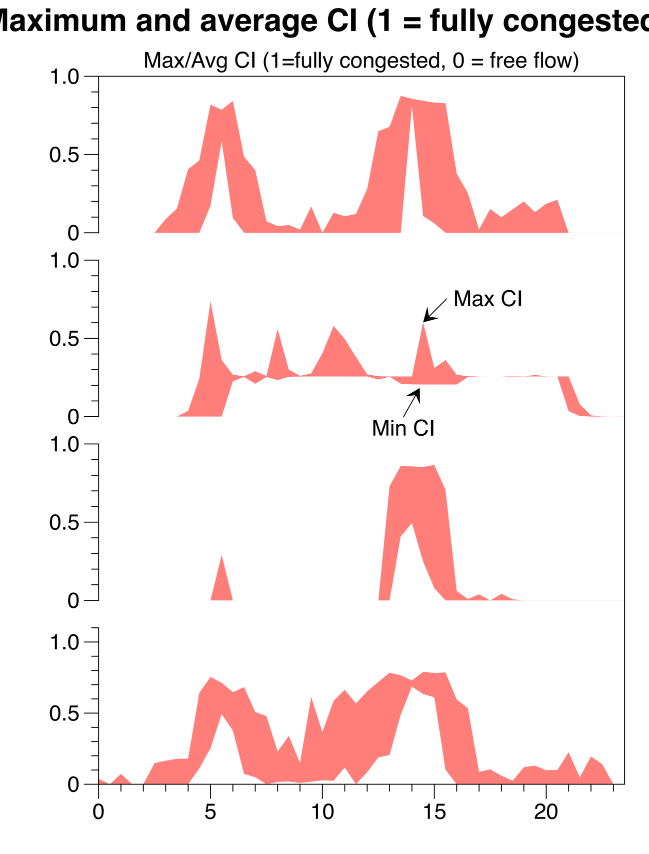
Spatial distribution of congestions



Recurrent/non-recurrent congestions

Figure 3 shows the maximum and minimum congestion index for 4 selected locations that have distinct congestion patterns. The top and the bottom locations show typical typical morning and afternoon rush-hour recurrent congestions; whereas the middle two figures show irregular non-recurrent congestions that are more likely to be associated with traffic incidents.

Figure 3. Maximum and average CI (1 = fully congested, 0 = free flow)



Real-time warning of abnormal congestion

Nonrecurring (or abnormal) congestions is identified in real-time based on the statistics calculated above when the real-time speed exceeds a tolerance threshold (e.g. traffic speed is two-standard deviation below the average speed for location x, in y direction, at time t). An alert will be created in list form or spatially and available to users: travelers, first responders, transit or private transportation service operators.

Our future work is to predict congestion patterns when a traffic incidence has been reported in HERE (which reports incidences by location, time, type, criticality, and start/end time) or identified as a nonrecurring/ abnormal congestion from the task above. A learning algorithm such as deep machine learning, Bayesian neural network or other statistics will be developed to "learn" the spatial and temporal correlations of congestion patterns based on past events and make predictions of the distributions and expected durations of delays in real-time when an incident occurs. An alert will again be created and available to users that allows for advanced planning and respond to avoid the expected congestions.

Table 1 below shows a work-in-progress warnings that match with incidents type; false negatives (incidents) and the false positives (deviations). Work is ongoing to improve the matches, and reduce false negatives and false positives.

Matches	64
Incidents	848
Deviations	5765

Type	Count	Matches
ACCIDENT	220	31
DISABLED VEHICLE	275	19
MISCELLANEOUS	13	0
ROAD HAZARD	242	4
ROAD_CLOSURE	30	0
CONGESTION	60	9
LANE_RESTRICTION	72	1

Table 1. Still work-in-progress: Matching warnings with incidents, matches, false negatives (incidents) and false negatives (deviations)

Theory

The data streaming processing paradigm has gained popularity in the last decade as an alternative to traditional store-then-process paradigms. Differently from the latter, data streaming analysis does not rely on persisted data (i.e. stored database) but rather allows for continuous streams of tuples (e.g., the geo-tagged messages generated by crowd-sourced traffic information applications) to be processed in an online fashion, producing updated results on the fly. As an example, it could allow to continuously aggregate crowdsourcing data referring to a given spatial location in order to maintain up-to-date statistics such as congestion indexes, continuously generating alerts for areas experiencing unexpected congestion levels. By providing efficient *distributed* and *parallel* analysis, data streaming allows for high-throughput analysis (being able to run complex analysis with rates in the realm of hundreds of thousands or millions of tuples per second) with low sub-second latencies.

One of the key hypotheses to be investigated in this project is that data streaming will provide efficient and expressive analysis of the data that is continuously sensed in urban environments. In turn, this will allow for valuable information (e.g., congestion alerts associated with traffic incidents) to be shared with citizens in a real-time fashion.

Contacting the Authors:
Sonia Yeh (sonia.yeh@chalmers.se). *Funding provided by the Adlerbertska Foundation, Sweden for the Adlerbertska Visiting Professorship at Chalmers University of Technology, Gothenburg, Sweden*

UC DAVIS
SUSTAINABLE TRANSPORTATION ENERGY PATHWAYS
An Institute of Transportation Studies Program