2018-00287 - Big Data Analysis and Modeling of Road Traffic

Level of qualifications required: Master's or equivalent
Function: Internship Research

About the research centre or Inria department

The Inria Sophia Antipolis - Méditerranée center counts 37 research teams and 9 support departments. The center's staff (about 600 people including 400 Inria employees) is composed of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrators. 1/3 of the staff are civil servants, the others are contractual. The majority of the research teams at the center are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Six teams are based in Montpellier and a team is hosted by the computer science department of the University of Bologna in Italy. The Center is a member of the University and Institution Community (ComUE) “Université Côte d’Azur (UCA)”.

Context

There exists nowadays a variety of road traffic data sources (magnetic loop detectors, video cameras, oating car data, bluetooth, etc.), which can be used for traffic management either for calibration of traditional models [1], or to directly infer traffic forecasts by statistical techniques [2], or even by combining modeling and data analysis techniques. In particular, the heterogeneity of trac conditions in congested regimes makes it hard to obtain a good matching between simulations and reality, thus preventing from obtaining reliable trac state predictions beyond short time horizons (30 min).

 Indeed, trac models are usually set on a unique choice of the fundamental diagram (ﬂux-density relation), independently of the time of the day and the day of the week. Yet, model parameters may be time dependent, as some trac conditions and driver behavior may be correlated to speciﬁc time ranges. Besides, data analysis techniques may help in opening new perspectives in the interpretation of data and the modeling of road traffic [3].

ACUMES Project-Team has an established experience in (macroscopic) trac models, and disposes of a large set of data coming from 135 loop detectors placed on the freeways in the North of Marseille (A7, A51, A50 and A55). The data-set was provided by the Direction Interdepartementale des Routes Mediterranee (DIRMED) and covers 3 months, from September 1st to November 30th, 2015.

On the other side, ZENITH Project-Team masters the analysis of large amount of data, by means of parallel and distributed approaches [4, 5]. Here, analysis covers different goals such as feature correlation detection, clustering or time series similarity queries [6].

ACUMES Project-Team has an established experience in (macroscopic) traffic models, and disposes of large set of data coming from 135 loop detectors placed on the freeways in the North of Marseille (A7, A51, A50 and A55). The data-set was provided by the Direction Interdepartementale des Routes Mediterranee (DIRMED) and covers 3 months, from September 1st to November 30th, 2015.

On the other side, ZENITH Project-Team masters the analysis of large amount of data, by means of parallel and distributed approaches [4, 5]. Here, analysis covers different goals such as feature correlation detection, clustering or time series similarity queries [6].


Assignment

Within the framework of a partnership (you can choose between)
- collaboration between 2 Inria teams: Acumes and Zenith
- others (please specify)
We aim to study the trac data-set mentioned above by time series similarity techniques to extract new knowledge in synergy with modeling and simulation to improve trac forecast and control.

**Main activities**

The study will be based on the following steps:

1. **Data preparation.** Before analysis, data need to be sorted by sensor and direction, and aggregated over time periods (usually 1 or 6 minutes) to obtain averaged quantities measuring density, mean velocity and flow.

2. **Data analytics.** Analysis will rely on time series data mining using the Matrix Profile. The goal is to extract correlations between time series.

3. **Match with macroscopic models.** The information derived by the previous analysis will be gathered and possible impact on modeling techniques will be investigated. If possible, the combination of the two aspects (modeling and data analysis) will be tested on simple configurations.

**Skills**

Technical skills and level required: background in statistical methods, knowledge of R, Python, Matlab.

Languages: French, English

**Benefits package**

- Subsidised catering service
- Partially-reimbursed public transport
- Flexible working hours
- Sports facilities