Development of Big Data Analytics Platform for Traffic Estimation

Background

Estimating traffic state on a road network is challenging due to high levels of uncertainty. Accurate traffic flow information helps travelers to make better decisions and alleviates the congestion. The advent of Big Data has resulted in revolutionary changes in field of Intelligent Transport Systems (ITS). However, methodologies and regulations in many domains of ITS have not kept pace with the proliferation of Big Data. More specifically, the current traffic estimation approaches significantly rely on data from inductive loop detectors which hinder the benefits of other data sources such as Floating Car Data (CFD). Admittedly, it is urgent to adapt the existing real-time traffic control systems to exploit the proliferation of Big Data.

Goal

The main goal of the project is to develop a data analytics platform based on the Big Data streaming tools (preferably Kafka) as well as raw data warehouse (HDFS or HBase) and NoSQL data storage (e.g. MongoDB). More specifically, a realistic traffic scenario which is already developed in SUMO (an open-source microscopic traffic simulation developed by German aerospace center) is used to generate the related data such as vehicles' speed, traffic lights, bus arrival time, etc. Then, the data should be streamed via Kafka to: (i) be stored in a raw data warehouse and (ii) be delivered to a data analytics engine (i.e. available machine learning algorithms) to estimate the traffic condition on the network.

Figure: proposed system architecture and the potential data sources in SUMO (www.sumo.dlr.de)

Tasks

- Establish data streaming between SUMO and Kafka using Python 2.7
- Plug-in a raw data warehouse and a NoSQL database into the platform
- Respond to the queries with respect to traffic state and visualize the data on a map

Contact

Sasan Amini, M.Sc. | Email: sasan.amini@tum.de

TUM Main Campus | Arcisstr. 21 Munich 80333 | Chair of Traffic Engineering and Control