Big data for mobility tracking and knowledge extraction in urban areas (Track & Know)

Muhammad Adnan
Transportation Research Institute, Hasselt University
(Track & Know Consortium member)

Introduction

Track & Know is a recently concluded Horizon 2020 project. The project primarily focused on mobility-based big data and developed innovative software stacks and toolboxes that addressed key questions in emerging cross-sector markets such as commercial Internet of Things (IoT) services, car insurance and healthcare management.

Existing and constantly generated new mobility data (big data) largely remains decoupled and isolated within private infrastructures. Efficient handling and inferring knowledge by effectively blending such big data can open new opportunities for business in particular and societies in general.

In today's society, the ability to retrieve knowledge from mobility and contextual data is becoming more and more critical for the competitiveness of all the economic, political and cultural entities. Multidisciplinary research teams from mobility data management, complex event recognition, geospatial modelling, complex network analysis, transportation engineering and visual analytics—as part of a **Track & Know** project consortium—developed a new big data management platform and various toolboxes for processing and analysing mobility big data.

To demonstrate the applicability of a developed platform and toolboxes, the project also focused on resolving key business cases for three test pilots, namely transport/mobility, insurance and healthcare. Business cases explored in these pilots are as follows but not limited to: minimising patient travel, accident risk estimation, carpooling and electric mobility potential, driver behaviour profiling etc.

Key achievements of the project

Big Mobility Data Integrator (BMDI)

The BMDI is a Kafka based, horizontally scalable, distributed streaming platform. It facilitates the efficient, high performance and horizontally scalable ingestion of data from a variety of data sources while ensuring its persistence and availability through a multi-client publish/subscribe model. It enables the decoupling and straightforward integration of various processing and toolbox components while ensuring secure communications with load balancing characteristics. It advances parallel processing capabilities by introducing data partitioning and subsequent processing by multiple instances of toolbox code deployed on the required hosts, enabling the setup of big data processing pipelines. It introduces Kafka Connect functionality where a variety of custom and community available connectors can be utilised. It provides a set of GUI components for platform administration and necessary dashboards for monitoring various performance-related metrics. The BMDI platform is able to efficiently interoperate with modern data storage technologies of a big data ecosystem such as RDBMS, NoSQL, HDFS Hadoop, Apache HBase, etc. as well as other persistence approaches such as Mongo, MySQL, JDBC, etc. Big data applications (toolboxes) can be implemented in all-important big data languages, including Python, Java, R and Scala. Traditional programming approaches (C/C++, Ruby, Perl, PHP) are also supported. Big data toolboxes developed within the project such as big data processing (BDP) toolbox, big data analytics (BDA) toolbox, complex event recognition (CER) toolbox and visual analytics (VA) toolbox are integrated into the platform to demonstrate the business cases of three pilots.

Adobe Stock © WrightStudio

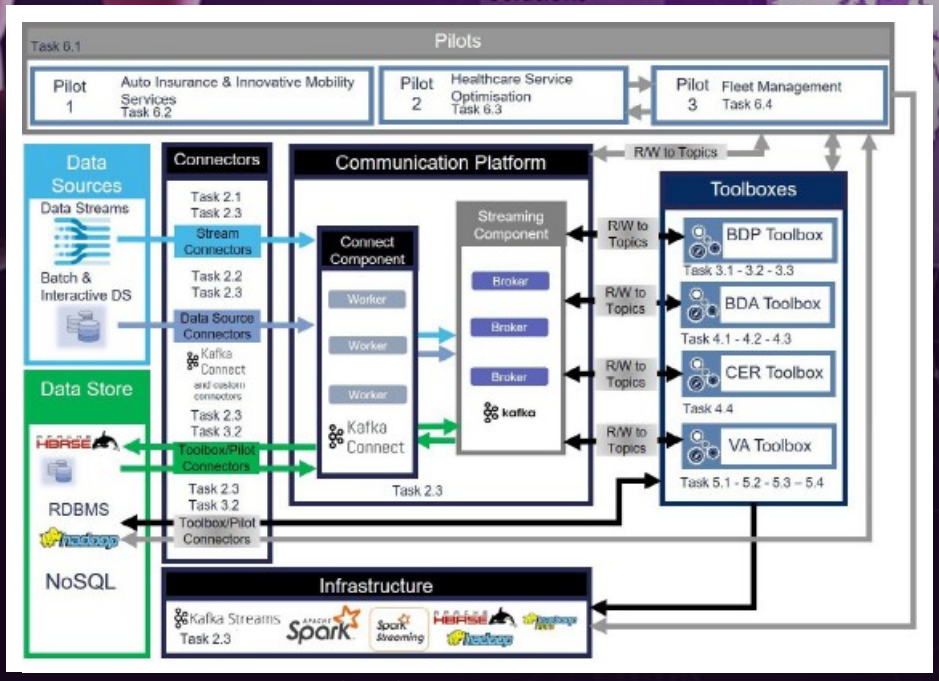

Figure 1: Big mobility data integrator (BMDI).

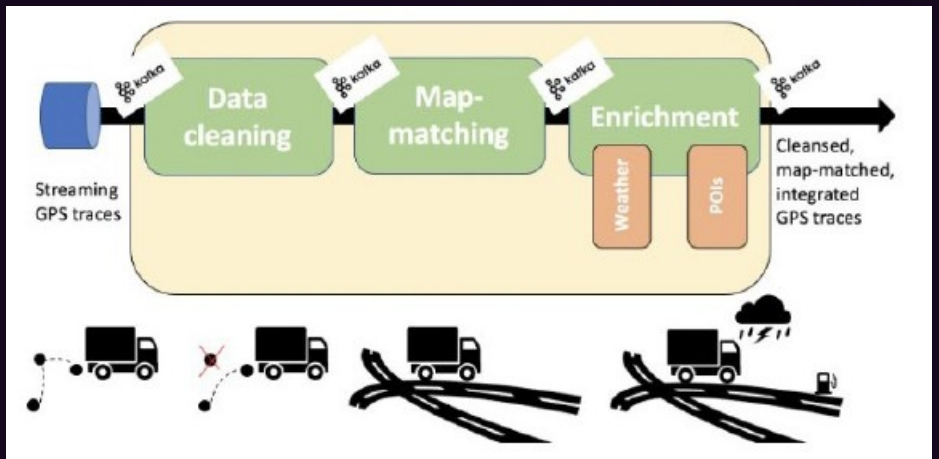

Figure 2: Data pipeline.

Data processing and enrichment pipeline

A data pipeline is developed as part of the BDP toolbox, as shown in Figure 2. It processes the raw GPS data and provides in its output cleansed trajectories of moving objects. The pipeline uses processes such as data cleansing, annotation, interpolation, map-matching and enrichment of data with the weather (e.g. wind, rain, ice etc.) and point-of-interest (e.g. fuel station etc.) information. This is developed with a viewpoint to support processes in BDA toolbox, to help answer questions on accident risk estimation, hot spot analysis, electric mobility analysis. The salient features of the enrichment architecture include operating in an online manner and being reusable across diverse mobility data (urban, maritime, air-traffic). Furthermore, the developed enrichment approach can handle more complex geometries than simple positions (e.g. associating weather with a 3D sector), and produce output in resource description framework (RDF).

Unified NoSQL Data Access Operators—the NoDA API

Track & Know researchers developed an approach within a BDP toolbox which is an abstraction layer in the form of an API over the NoSQL stores. The work carried out contains: (a) the specification of an API for data access/retrieval operators over NoSQL stores; and (b) its implementation for a specific NoSQL store, namely MongoDB, HBase and Redis. The designed

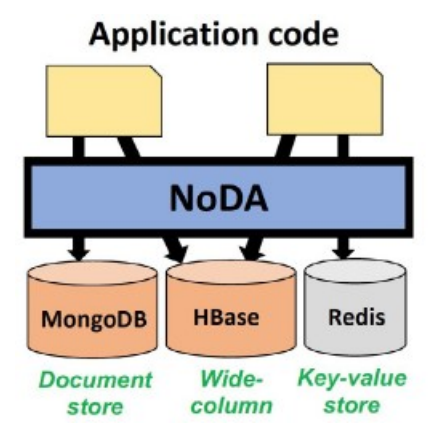

Figure 3: NoDA (Unified NoSQL Data Access Operators) framework.

API for big data access operators can be used by big data developers for easy access to NoSQL stores while hiding implementation details relevant to the specific NoSQL store. This allows developers to write programs against this API, which is much easier, more comprehensive, and removes the 'entry barrier' of having to learn the internal operation of a specific NoSQL store. NoDA is JVM-compliant and can return results as Spark DataFrame.

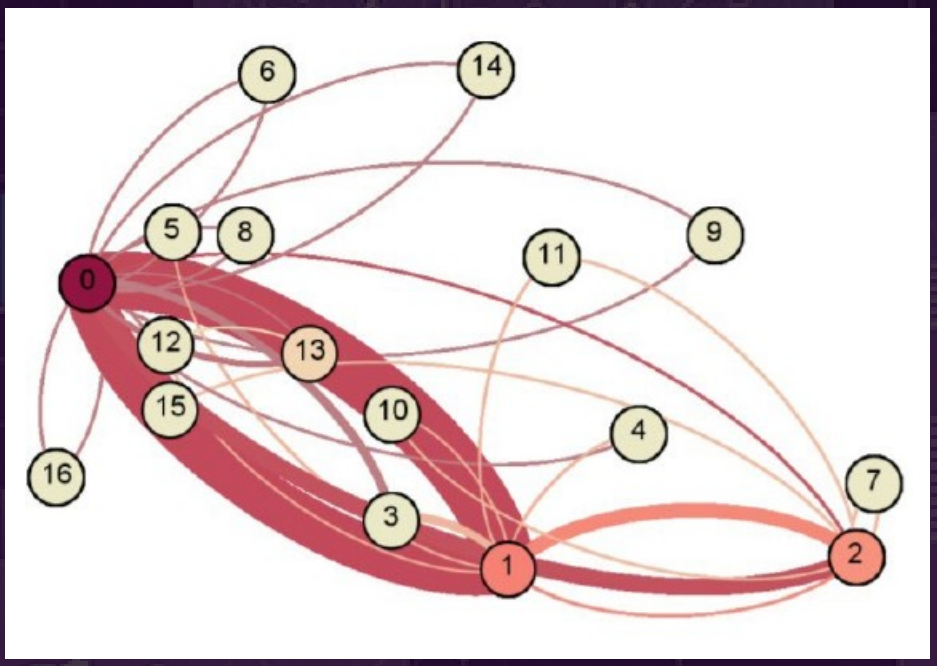

Figure 4: Individual mobility network (IMN).

Extraction and semantic annotation of individual mobility networks (IMN) for the purpose of crash risk prediction

IMNs are a concise graph representation of the mobility history of individuals. From raw GPS traces, the trajectories of a single mobility user are reconstructed and processed to infer the relevant locations that the user visited (the nodes of IMNs) and aggregate the trips between two

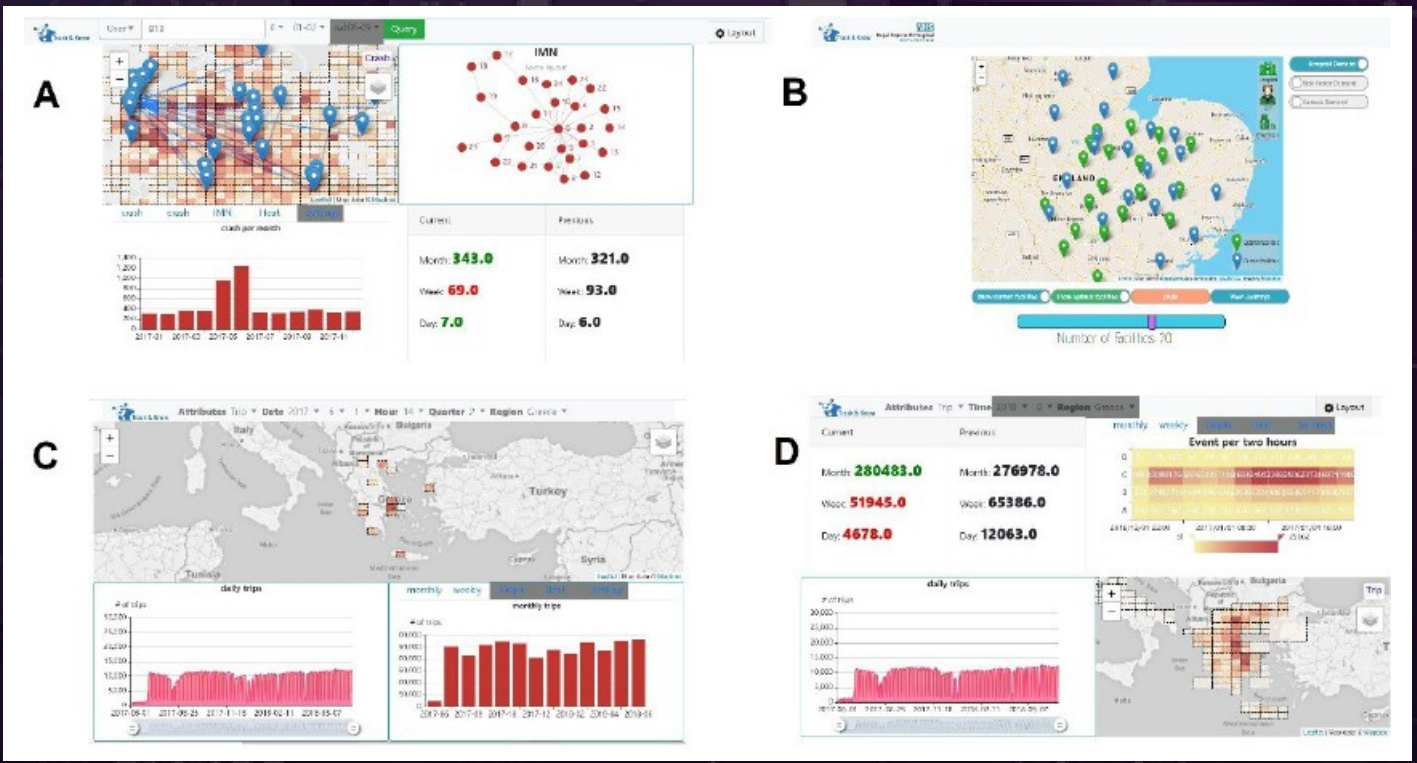

Figure 5: Project dashboards.

locations (the edges of IMNs). Several features are extracted that are used as key variables in the model. These are as follows: the number of nodes and edges, node and edge frequency, clustering co-efficient of IMN, network modularity index, and measure of IMN change with respect to time that reflects changes in habits or mobility needs.

As part of BDA toolbox, a predictive model for crash risk is developed using a machine learning technique. The model uses IMNs along with other features of an individual mobility which are correlated to events that happen in the mid/long-term future, such as crashes in the next month. The component is implemented in Python, is completely automatised and takes as input the mobility data of the user in the monitoring period, as well as (only for the construction of the model, and not for its usage in prediction) a label specifying whether the event is going to happen.

Track & Know dashboards

As part of the VA toolbox, the T&K dashboard provides an overview of some aggregated statistics, as well as the analytical results generated from the other toolboxes. In short, the dashboard does not serve as an interface of the analytical tools but rather visualises the analytical results, which are generated from off-line big data analytics (thus, it acts as a 'strategic dashboard'). It provides simple interactions (e.g. selection, filtering, and drilling down), allowing users to explore the analytical results on the browser side, partially as an analytical dashboard. It is implemented using web technologies and can be accessed via web browsers to maximise accessibility by end-users. In the background, it connects to the database to receive real-time updates of aggregated statistics and analytical results and visualise them on the dashboard interface.

Concluding remarks

The methods/tools/functionalities discussed are just a few examples of developments/achievements made in the project. There are demonstrating videos available on project website and its Youtube channel, where these methods/tools are utilised in addressing a particular research question for the three pilots' cases. Interested readers are requested to visit the Track & Know website to explore more developments and demo videos. In addition to this, the website also contains an online observatory, where users can request and explore the sample data sets, download the developed software packages and other resources generated within the project.

Track & Know researchers and consortium members are acknowledged for providing the necessary materials that have been compiled and presented in this article.

Track & Know

PROJECT SUMMARY

Big Data for Mobility Tracking Knowledge Extraction in Urban Areas (Track & Know). The overall aim is to research, develop and exploit a new software framework that aims to increase the efficiency of big data applications in the transport/mobility, motor insurance and health sectors. A variety of toolboxes (that contain specific methods for various types of data aggregation, manipulation and further analysis) are developed within the project and integrated into a software platform.

PROJECT LEAD PROFILE

Inlecom Group is coordinating the Track & Know project. A European SME founded in 1996, Inlecom is now an established leader in digital ecosystem platforms and technologies. The core competences include: project management, governance and OPEX , solution design, prototyping and POCs, security and compliance, public/private/hybrid cloud, IoT, analytics and big data, basic and applied research etc.

PROJECT PARTNERS

The Track & Know consortium is composed by complementary partners, coming from addressed research, technological and commercial domains that have a proven track record of high-quality research capacity in the field of big data handling, analytics and visualisation tools. The consortium includes organisations from nine European countries (four industry corporations, seven scientific research institutes, and three research capable SMEs).

CONTACT DETAILS

Dr Ibad Kureshi, Senior Research Scientist
Inlecom Systems, Square de Meeus 38/40, 1000 Brussels, Belgium

Ibad.kureshi@inlecomsystems.com
https://trackandknowproject.eu
@TrackandKnow

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No.780754.