Track & Know | Issue 1 | Date: April 2019

WELCOME TO THE FIRST TRACK & KNOW NEWSLETTER!

In this first newsletter, you can find news on:

- The overall scope of the project
- The first results based on our first data analysis
- Publications of 2018

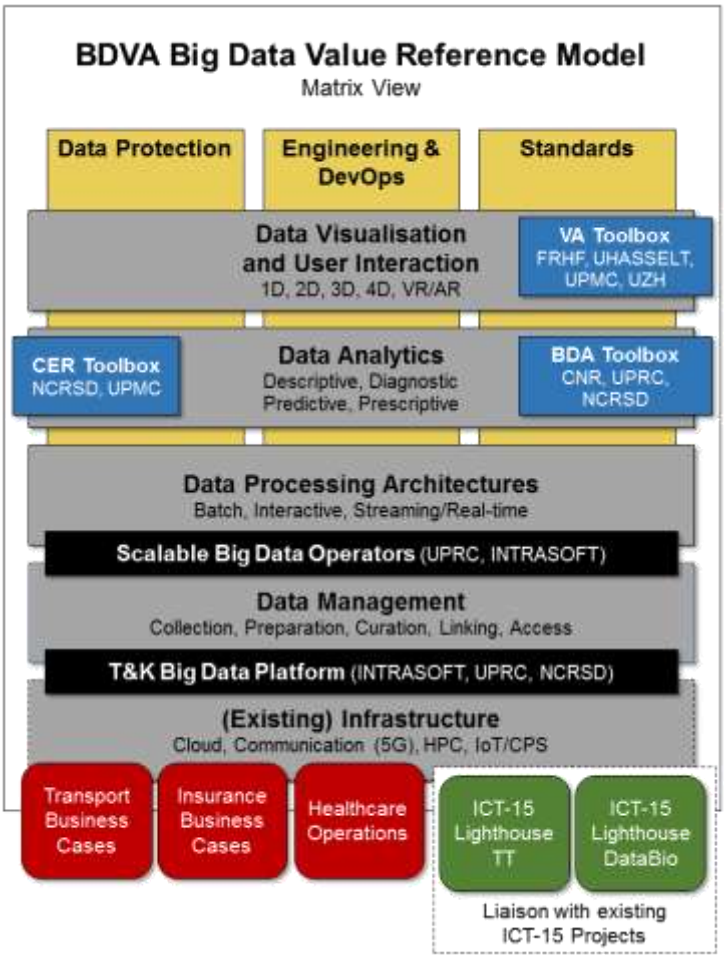
ABOUT THE NEWSLETTER

This newsletter informs you about the results and activities of the EU H2020 research project Track& Know. The aim is to keep all relevant actors interested in managing big data, more specifically on the type of big data we focus on in the project and the tools/methods we develop to handle, analyse and visualize these datasets. T&K focuses on resolving key business cases for 3 test pilots, namely transport/mobility, insurance and health care. Business cases which will be explored in these pilots are as follows but not limited to: minimizing patients travel, carpooling and electric mobility potential, driver behaviour profiling etc.

Project coordinator
Dr. Ibad Kureshi
Senior Research Scientist – Inlecom Systems
Square de Meeus 38/40
1000 Brussels
Belgium
Ibad.kureshi@inlecomsystems.com

THE OVERALL SCOPE OF THE PROJECT

Track & Know researches, develops and exploits **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 / functions / algorithms for various types of data aggregation, manipulation and further analysis) are developed within the project, and integrated in a software platform.



A **Big Data Processing (BDP) toolbox** is developed to implement data acquisition technology that captures data from heterogeneous data sources. Current available solutions can handle spatial and temporal data separately and therefore fail to exploit spatiotemporal correlations present in mobility data. The BDP toolbox extends the current solutions and delivers a tool for efficient access, indexing, partitioning and load balancing for Big spatiotemporal data.

A **Complex Event Recognition (CER) toolbox** detects complex event occurrences by analysing patterns in simple events. To do that, it uses contextual information and results from the BDA toolbox. For example: the toolbox may infer a complex event (such as dangerous driving or non-economical driving) by analysing patterns based on vehicle speed, direction, driver events, fuel consumption and other contextual information such as weather etc. The CER toolbox will advance the state-of-the-art by developing online learning techniques for complex events in big mobility data.

The **Big Data Analytics (BDA) toolbox** is developed to analyse heterogeneous data and to draw conclusions about the spatiotemporal distribution of mobility patterns. Current available analytical models fail to scale for big datasets and/or online streaming data. The BDA toolbox delivers scalable data mining techniques (such as clustering, sequence mining, hot-spot analysis) for voluminous offline and online trajectory data.

A **Visual Analytics (VA) toolbox** develops interactive and scalable methodologies to visualize data at all steps of analysis. Current approaches supports visualization for only a few analysis steps and therefore, does not provide explicit understanding. The VA toolbox can efficiently handle both historical and streaming spatiotemporal data originating from different sources, with varying levels of resolution and quality.

To put theory into practice, **we integrate the toolboxes in a platform and test them in pilot cases** that serve as a test-bench for the stakeholders to evaluate developments using realistic data. **We organise pilot cases in three domains**: transport/mobility, insurance and health care. **A variety of business questions will be answered in each pilot** using the developed toolboxes. For example: detecting errors in the trajectory data and their corrections, travel route features and predictions, driving behaviour categorisation and risk scoring, location of mobile health units so as to minimize travel distance for patients, potential for carpooling and electric mobility etc. For more information, please visit the project website: www.trackandknowproject.eu.

SOME PRELIMINARY RESULTS

Geographical Distribution of patients who may be suffering from Obstructive Sleep Apnoea (OSA)[1]

Patients from Royal Papworth Hospital (RPH), Cambridge, England and Outreach Clinics (outreach exchange facility) in the surrounding region.

The figure (Source: Inlecom Systems BVBA, Belgium) **shows the results of the first retrospective analysis of the geographical distribution of patients requiring OSA testing**, monitored from 2013-2018. Patients, monitored by RPH are visualised with a green triangle. Patients, monitored by an outreach exchange facility are visualized with black and blue stars.

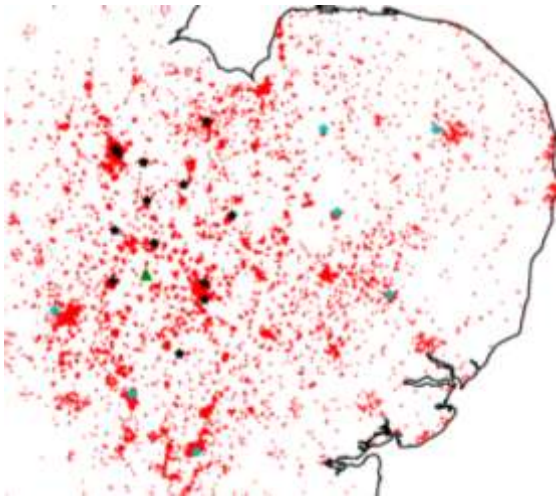


Figure 1: All patients in catchment 2013-2018

This image has allowed us to **visualise the distribution of OSA patient clusters and how these clusters relate to exchange facility locations**. You can see that many outreach facility centres are ideally placed, hitting the areas with lots of patients. But additionally you can also see the service is far from optimal and that many patients have to travel vast distances to get access.

The next analysis will look at what is the best way of re-distributing the exchange facilities to reduce overall travel distance, and adding another layer to the map showing the spatial variation of demographic risk factors that would make someone likely to need to undergo a diagnostic test for OSA. In other words to **find out if there are geographical areas that are in need of tests, but not currently getting access to them**. Although these results are very early and there is a lot more we still hope to achieve in the Track and Know project, the **approach being used is generating a significant degree of interest**. The type of research is fresh and novel within the area of OSA and possibly underutilised in the field of medicine as a whole.

At present these preliminary results are being used by RPH consultants to justify the importance of maintaining outreach services. In particular the hospital is having to defend the need for an outreach in Harlow (Lowest blue star), and to get General Practitioners (GP's) in Peterborough (2 black starts top left) to improve their level service. **The fact that research results are being used, already at such an early stage highlights how well this type of analysis is being received**, and offers a good deal of promise that our next steps will lead to something highly beneficial.

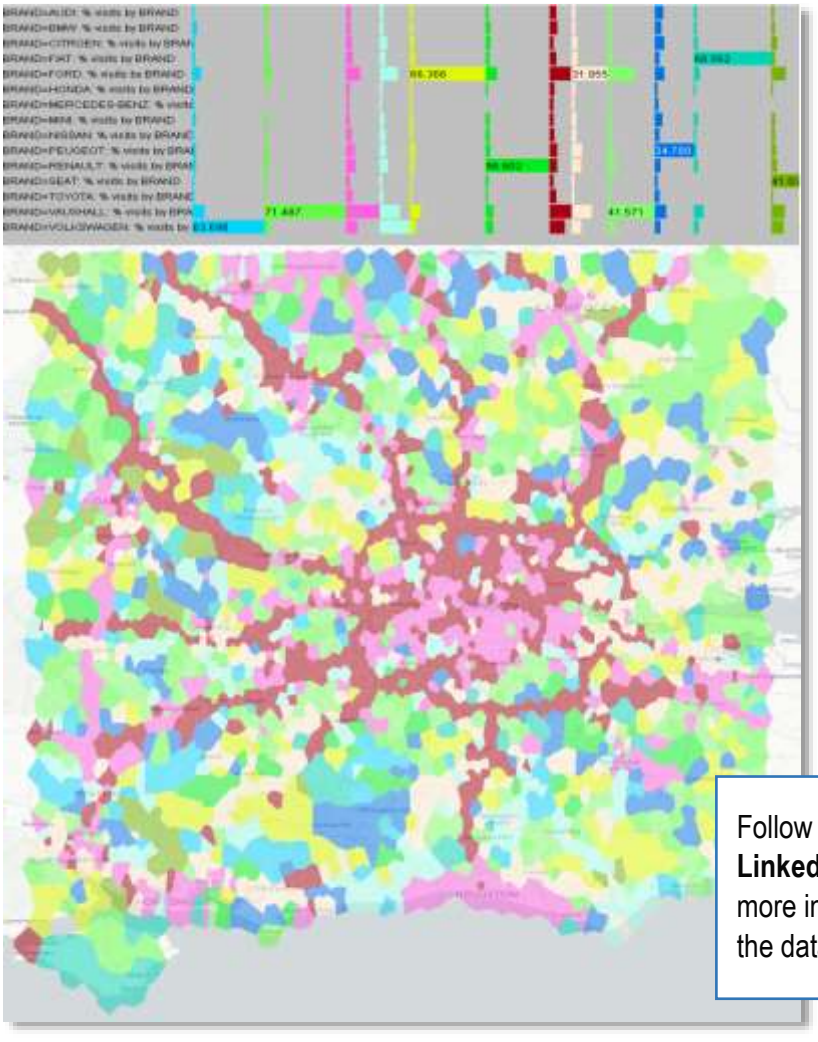
[1] Obstructive sleep apnoea (OSA) is the most common type of sleep apnoea and is caused by complete or partial obstructions of the upper airway. The patients involved in the study are not diagnosed with OSA yet, but there is a strong suspicion that these patients may suffer from OSA. For that reason these patients have to use an expensive device

Distribution of car type usage

Initial Analysis of dataset obtained from Sistematica, ICT innovators (SIS), Italy for London region

The analysis provides an idea about which type of car appears on which type of road. This makes it possible **to depict their geographical markets and to see where a particular brand has a higher accident risk** based on the geographical distribution. Coupled with other contextual information, the analysis can help to **answer questions relevant to the insurance pilot**.

About **two weeks of data** were gathered in the UK (greater London area) regarding the distribution of car brands. **15 top brands** that have at least 80 cars during the considered period were selected. We applied our trajectory summarization procedure for **tessellating the area into about 3,500 polygons of similar sizes** (from our visual analytics toolbox). Next, we **aggregated the trajectories** by these polygons separately for the 15 brands. As a processing result, we have **clustered the polygons** according to the similarity of these "car population structures" and got a nice map. Polygon colours reflect the similarity of the car population profiles. Colour meaning is explained in the bar chart at the top.



The map gives a nice overview of the **structure of the road network and land use in the area**. Main motorways are dominated by VAUXHALL, FORD and VOLKSWAGEN. Central London and Brighton are characterized by a mix of everything, with some prevalence of VAUXHALLs and FORDs. You can find nice FIAT, FORD, SEAT, PEUGEOT and VOLKSWAGEN "villages".

Follow the Track & Know project on **Twitter**, **LinkedIn** and periodically visit our **website** for more interesting analyses and visualizations of the dataset used in the project.

PUBLICATIONS 2018

- Nikitopoulos P., Paraskevopoulos A.-I., Doulkeridis C., Pelekis N., Theodoridis Y. (2018) Hot Spot Analysis over Big Trajectory Data, *In Proceedings of the 2018 IEEE International Conference on Big Data (IEEE BigData 2018)*.
  Download here

- Markovic N., Sekula P., Vander Laan Z., Andrienko G., Andrienko N. (2018). Applications of Trajectory Data From the Perspective of a Road Transportation Agency: Literature Review and Maryland Case Study, *IEEE Transactions on Intelligent Transportation Systems.*
  Download here

- Collins C., Andrienko N., Schreck T., Yang J., Choo J., Engelke U., Jena A., Dwyer T. (2018) Guidance in the human-machine analytics process, *Visual Informatics*, vol. 2(3), pp.166-180.
  Download here

- Li J., Chen S., Zhang K., Andrienko G., Andrienko N., (2018) COPE: Interactive Exploration of Co-occurrence Patterns in Spatial Time Series*, IEEE Transactions on Visualization and Computer Graphics*.
  Download here

- Katzouris N., Michelioudakis E., Artikis A., Paliouras, G. (2018). Online learning of weighted relational rules for complex event recognition. *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 396-413, Springer, Cham.
  Download here

Find & Follow us
Website: https:// trackandknowproject.eu
Twitter: @TrackandKnow
LinkedIn: Track & Know
Researchgate

The project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 780754.