Track & Know | Issue 6 | Date: May 2020

WELCOME TO THE SIXTH TRACK & KNOW NEWSLETTER!

In this newsletter, you can find:

- NoDa: a recent development
- Update on the upcoming Summer School on Data Science for Mobility, supported by Track & Know
- 5 new software additions to the online observatory
- Find out what the Track & Know Health Care and Fleet Management pilots are about

ABOUT THE NEWSLETTER

This newsletter provides an update on the results and activities of the EU H2020 research project Track & Know. The aim is to ensure that those interested in managing big data are kept up to date with our latest developments, specifically mobility data and the tools/methods we develop to handle, analyse and visualize these datasets. Track & Know aims to answer industry generated questions in 3 test pilots in the transport/mobility, insurance and health care sectors. The business cases explored in these pilots centre on answering questions around minimizing patients travel, carpooling and electric mobility potential and driver behaviour profiling.

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

NoDA: a recent development within Track & Know

Researchers from **University of Piraeus**, Greece (**Track & Know Consortium partners**) have developed a unified operator's framework to access big data that alleviates the burden of learning the query languages of each NoSQL database store. They call this framework **NoDA** (Unified NoSQL Data Access Operators). Currently the operators are developed for mobility related (trajectories-based) big data. Its primary focus is on scalable mobility data management, in particular spatial and spatio-temporal operators. This work is part of the development **of a Big data processing (BDP)** tool box within the Track & Know project.

NoDA resides between application code and data storage as a bridge for data access, and aims at "hiding" the query language of the underlying store from the developer. The key features are as follows:

- It is ***simple*** in terms of using a familiar vocabulary of generic operations (filter, project, sort, etc.), without mixing the data model and the query language of the individual NoSQL store in the application code.
- The framework is ***unified*** because the exact same operations are used for querying different NoSQL stores.
- NoDA does not require the data to be moved; it queries the data on their original NoSQL store which distinguishes it from the existing similar solutions.
- In addition to common operators such as boolean and comparable ones, NoDA supports operators oriented to spatial (2D) and spatio-temporal (3D) data, called Geographical Operators (or Geo-Operators in short).

More details on this development can be read from the article, where the NoDA framework applicability is discussed in detail.

For more updates on Track & Project, please visit the website https://trackandknowproject.eu .

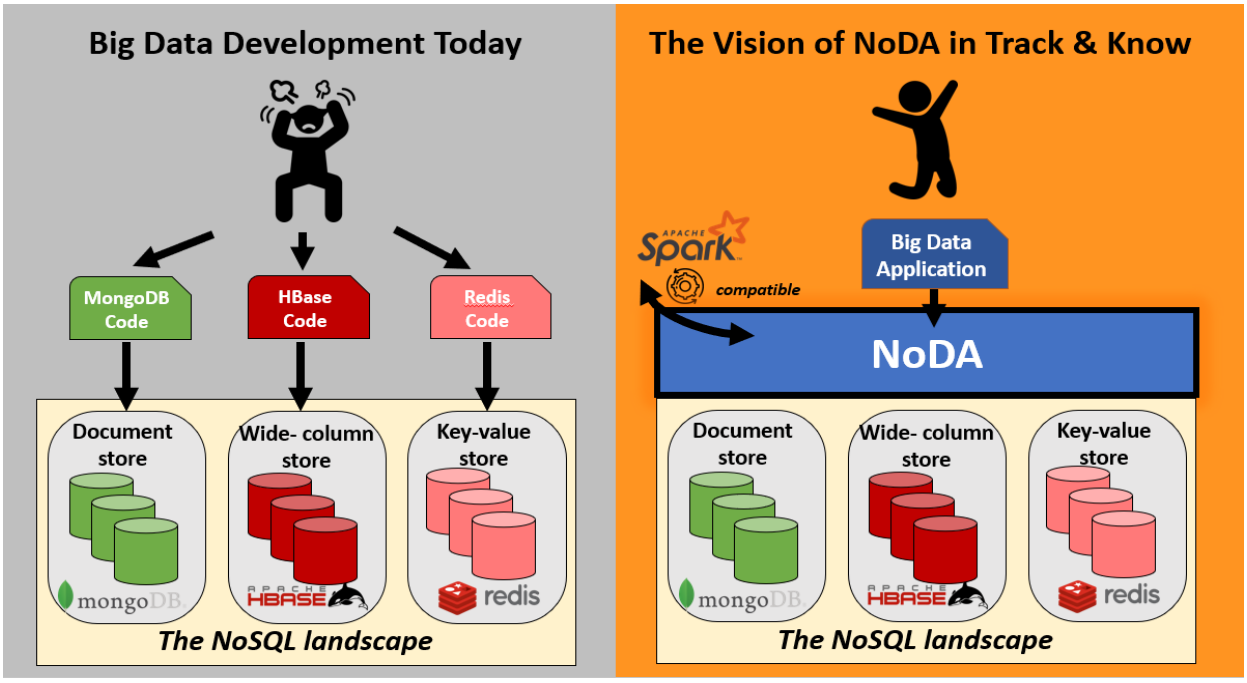


Fig. 1: The vision of NoDa in Track & Know.

Postponed: First International Summer School on Data Science for Mobility

IMPORTANT! Due to the coronavirus outbreak the summer school is postponed to October 12 – 16, 2020. More information and updates on the scientfic and social program will follow shortly.

About

The objective of the **First International Summer School on Data Science for Mobility,** supported by Track & Know, is to offer to participants both visionary keynote speeches and hands-on mini courses. The keynote speeches are given by leading experts on maritime and aviation domains introducing analysis challenges a large quantity of very complex mobility data generated every day.

Aim of the school

Massive amounts of spatio-temporal data, representing trajectories of moving objects are produced by an ever-increasing number of diverse, real-life applications, ranging from mobile to social media apps and surveillance systems, from vehicle tracking systems to IoT mobile sensors. Such mobility-aware traces come in huge numbers and very complex forms, and can be enriched with multiple different semantic dimensions. These semantically enriched trajectories have the potential to unveil novel challenges in several domains, such as urban, maritime and aviation.

The explosion in Data Science is happening now. The Big Data technological infrastructure has reached maturity. Significant interest from the research community is being shown towards the Big Data Value Analytics reference model: data management, data processing, data analytics, data visualization. The time is right for the field of Mobility Data Science to follow the trend!

Course aim

At the end of the course, each attendee will:

- Understand how to analyse mobility data with deep learning techniques
- Understand how machine learning and AI methods can be tailored to mobility data
- Understand how to manage Big Mobility Data
- Gain significant hands-on experience with state-of-the-art technologies and tools
- Have a vision of open research as well as technological challenges customized to key application areas and domains

Click here for full info and updates

5 new software additions to the online observatory

(all part of the BDP tool box)

1. Distributed Subtrajectory Clustering

Trajectory clustering is an **important operation of knowledge discovery from mobility data**. Especially nowadays, the need for performing advanced analytic operations over massively produced data, such as mobility traces, in efficient and scalable ways is imperative. However, discovering clusters of complete trajectories can overlook significant patterns that exist only for a small portion of their lifespan. Here, we address the **problem of Distributed Subtrajectory Clustering (DSC)** in an efficient and highly scalable way.

2. Run-Time Event Calculus

RTEC is an Event Calculus implementation optimised for stream reasoning.

Features

- Interval-based.
- Sliding window reasoning.
- Interval manipulation constructs for non-inertial fluents.
- Caching for hierarchical knowledge bases.
- Support for out-of-order data streams.
- Indexing for handling efficiently irrelevant data.

Applications

RTEC has been used for: Maritime monitoring, Activity recognition, Fleet management, City transport & traffic management.

3. Distributed Subtrajectory Join

Joining trajectory datasets is a significant operation in mobility data analytics and the cornerstone of various methods that aim to extract knowledge out of them. In the era of Big Data, the production of mobility data has become massive and, consequently, performing such an operation in a centralized way is not feasible. Here, we address the problem of Distributed Subtrajectory Join (DSJ) processing by utilizing the MapReduce programming model. The problem that we address is as follows: given two sets of trajectories (or a single set and its mirror in the case of self-join), identify all pairs of maximal ``portions'' of trajectories (or else, subtrajectories) that move close in time and space w.r.t. a spatial threshold ε_{sp} and a temporal tolerance ε_t, for at least some time duration δt.

4. Distributed Subtrajectory Similarity Matrix

An open source implementation of the Distributed Subtrajectory Similarity Matrix solution by employing an updated subtrajectory join procedure and similarity function.

5. Online Relational Learning

Refactored, latest version of OLED & WOLED plus more tools for for online logical & relational learning.

[Download this software via the online Track & Know observatory]

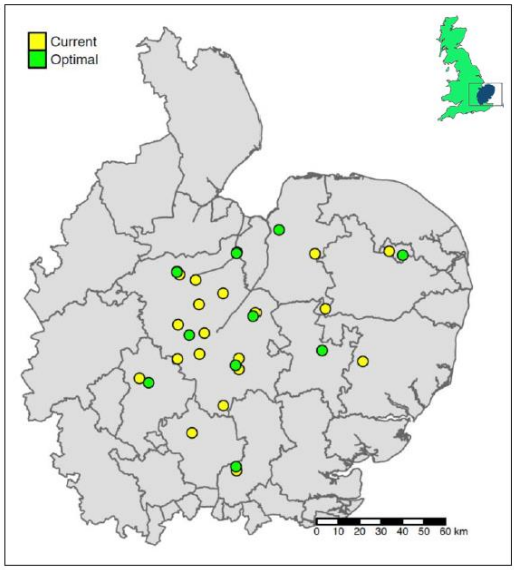
Track & Know Health Care Pilot

BIG DATA IN HEALTHCARE SERVICES

FACT: 1.5 million adults in the UK have Obstructive Sleep Apnoea (OSA) but only 330,000 patients are diagnosed and receiving treatment. OSA patients are around 3 times more likely to be involved in a crash than a regular driver and Sleepiness whilst driving accounts for around 20% of all motor vehicle collisions

PROBLEM 1: Royal Papworth Hospital's patients have to perform at least 4 long journeys to get access to the oximetry test because of improper distribution of oximeters' pick-up points. The results are **increased travel distance, no-show rate and waiting times.**

The role of big data: To use known risk factors of OSA and open source data to predict likely high OSA risk areas and redistribute oximeters to locations where most patients live, improving service quality



Pic. 2: Current vs optimal locations of 10 exchange locations

PROBLEM 2: The absence of measures to qualify whether OSA patients are safe to drive on the road.

The role of big data: To use some of the methods of tracking GPS data and accelerometer to understand driving behaviour of OSA patients

SUCCES STORY

"Sleep well – drive safely – a tale of mobility tracking and Big Data" won "Success Story Awards 2019" at the BDV PPP summit 2019 in Riga (i.e. the primary event for driving European innovation in Big Data and Artificial Intelligence)

Pic. 3: Succes Story Awards 2019

Track & Know health pilot highlighted on ITV Anglia (UK)

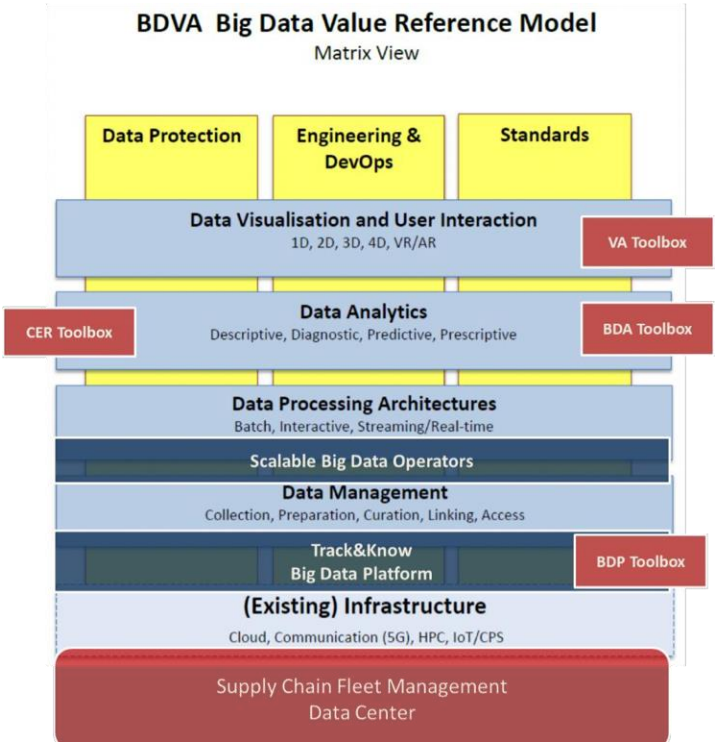
Last March, ITV Anglia News put the spotlight on Track & Know's health pilot at NHS Royal Papworth Hospital in Cambridge, UK. They paid a visit to **Dr. Ian Smith** at the hospital ward where patients with Obstructive Sleep Apnea (OSA) receive treatment and also focussed on the ongoing research within the framework of this H2020 project. **Researcher Kieran Lee** explained how this pilot study measures driving behaviour by recording brain waves and using phone apps. The study ultimately aims to develop a test which will indicate to people if it is safe for them to drive. **Watch it here!**

Track & Know Fleet Management Pilot

BIG DATA INNOVATIONS IN FLEET MANAGEMENT

The enormous volume of mobility data in this new era thanks to on-board devices, sensors and wireless connectivity has posed new challenges in the world of mobility big data management. However, Track&Know's set of toolboxs, including Big Data Processing (BDP), Big Data Analytics (BDA), Complex Event Recognition (CER), Visual Analytics (VA) can enable the applications of big data to become opportunities to innovate the management and operation of fleet management systems.

BDP aims at supporting novel and scalable, solutions of high throughput addressing storage, efficient access, indexing, partitioning and load balancing for Big spatio-temporal data with reliable data collection modes and a set of big data operators. Therefore, increased number of external sources are integrated (e.g. weather, points of interest …); invalid coordinates and invalid speed calculations due to errors are reduced in the fleet monitoring system.



Pic. 4: Positioning of Track&Know Toolboxes against the BDVA reference model architecture

BDA deliver scalable trajectory data mining techniques for voluminous data and real-time techniques to incrementally capture recurring or rapidly evolving phenomena.

- With **support for computing intensive, analytic processing and machine learning techniques**, BDA helps identify driving behavior excess per driver, identify patterns leading to improved fleet maintenance costs and support preventive maintenance recommendations based on tracked parameters (service downtime, etc.)
- With **Future Location Predictation**, BDA helps proactively identify traffic hot spots per day and its alternative routes
- With **Trajectory Prediction**, BDA helps provide recommendations for fuel consumption reduction based on the overall fleet performance optimization, provide accurate estimations of future travel distances and increase the recommendations for alternative routes based on fuel economy and road conditions.

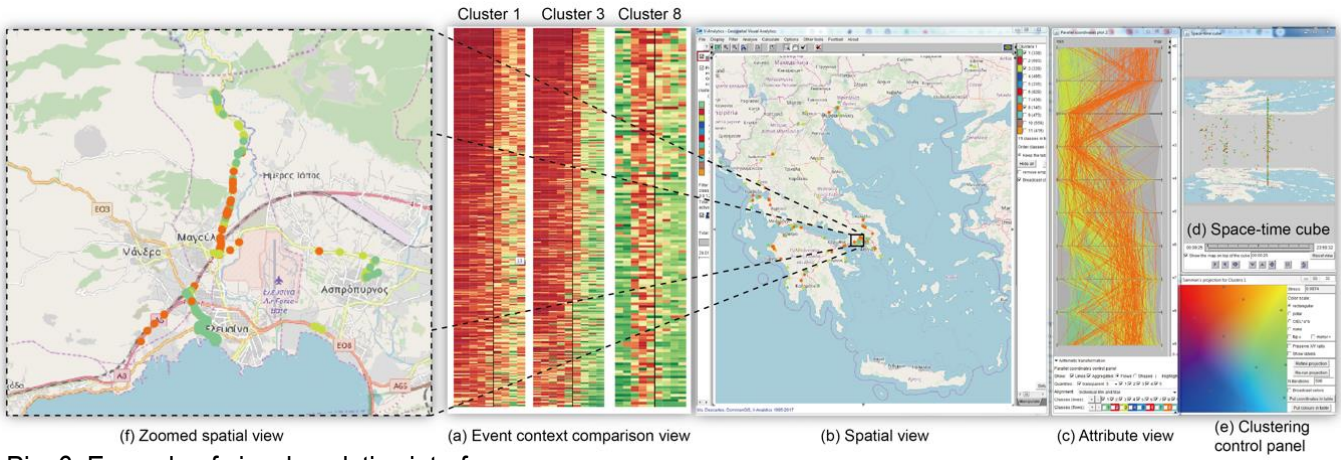
CONTEXTUALISED ANALYSIS OF MOVEMENT EVENTS

To identify and investigate potentially dangerous driving behaviors in commercial fleet vehicles, the approach transforms big data of vehicle trajectories extracted from tracking devices to a visual analytics workflow to analyse dynamic attributes of moving vehicles before and after the event of interest.

1st step: Selection of the events of interest with their times and locations. (e.g. harsh braking, harsh acceleration …)

2nd step: Selection of the relevant attributes (e.g. speed, engine status, fuel amount …), the desired window in relation to the event time and the temporal resolution considering the sampling rate of the available data → Each event is then characterised by a vector of contextual attribute values.

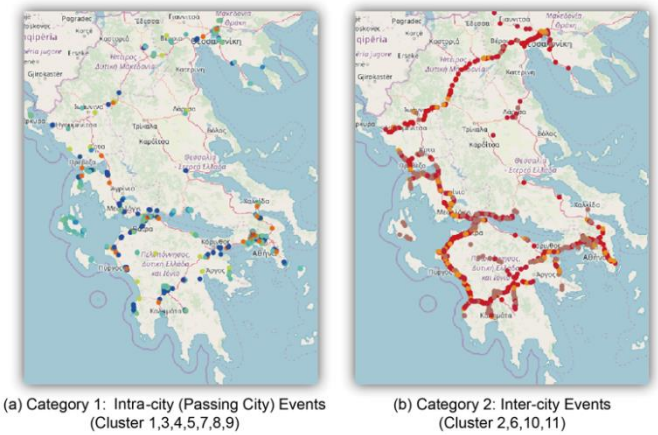
3rd step: Repeated patterns are discovered by applying clustering to the vectors of all events with an appropriate similarity measure and the clustering technique. → Attribute characteristics of the clusters are presented in visual displays for comparison and semantic interpretation.



Pic. 6: Example of visual analytics interface

4th step: Investigation of the spatial and temporal distribution of the clusters → Identification of spatial or spatio-temporal "hot spot".

FACT: Siming Chen and other research collaborators for their research work on "Contextualised analysis of movement events" that received the best paper award in EuroVA 2019 workshop held jointly with EuroVis 2019 Conference in Lisbon (Portugal) on June 3-6, 2019.



Pic. 7: Example of comparison of spatial distribution of events with *distinct context patterns*

RECENT PUBLICATIONS

- Adnan, M, Gazder U., Yasar A. H., Bellemans T., Kureshi, I (2020), Estimation of travel time distributions for urban roads using GPS trajectories of vehicles: a case of Athens, Greece, *Personal and Ubiqutious Computing*, DOI: https://doi.org/10.1007/s00779-020-01369-4, Download **here**
- Theodoridis Y. (2020) Learning from Our Movements – The Mobility Data Analytics Era. In: Tserpes K., Renso C., Matwin S. (eds) Multiple-Aspect Analysis of Semantic Trajectories. MASTER 2019. Lecture Notes in Computer Science, vol 11889. Springer, Cham. Download **here**
- Tampakis, P., Doulkeridis, C., Pelekis, N., Theodoridis, Y. (2020). "Distributed Subtrajectory Join on Massive Datasets." ACM Digital Library. Published version: https://dl.acm.org/doi/10.1145/3373642
- Nikitopoulos, P., Sfyris, G.A., Vlachou, A., Doulkeridis, C., Telelis, O. (2020) "Pruning Techniques for Parallel Processing of Reverse Top-k Queries." Distributed and Parallel Databases (Springer), 2020. Published version: DOI 10.1007/s10619-020-07297-9
- Guidotti, R., Nanni, M. (2020). Crash Prediction and Risk Assessment with Individual Mobility Networks. The 21st IEEE International Conference on Mobile Data Management (MDM 2020). **To appear soon.** Pre-print version **here**.
- Yeghikyan, G., Opolka, F., Lepri, B., Nanni, M., Lio, P. (2020) Learning Mobility Flows from Urban Features with Spatial Interaction Models and Neural Networks. IEEE International Conference on Smart Computing (SMARTCOMP). **To appear soon.** Pre-print version **here**.

Full list of publications 2018-2019-2020 is available HERE

Find & Follow us

Website: https:// trackandknowproject.eu

Twitter: @TrackandKnow

LinkedIn: Track & Know

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