Machine learning for optimizing disposition and planning of vehicles with near real-time IoT events at scale

Dr. Anusch Daemi-Ahwazi[1], Dr. Daniel Rost[2]

Abstract: Cargo vehicles today are equipped with power saving IoT devices measuring various aspects of the vehicle and cargo itself. The real-time stream of IoT events from the vehicles are sending large amounts of data each day, which needs to be correlated with each other and existing data sources to generate business value. The algorithmic challenges for discussion are the handling of noisy data and fast correlation of the sensor data as well as software engineering challenges to ensure the system(s) are highly performant and maintainable over the next decades.

Keywords: Machine Learning; Graph Theory; IoT; Software Engineering

1 Introduction and overview

In this workshop we will focus on an evaluation of complex algorithmic solutions based on graph theoretical approaches as well as machine learning algorithms, enabling use cases related to planning and network optimization of vehicles traveling mostly across fixed relations (e.g. railway corridors across Europe or fixed routes). The challenges for the proposed solutions are related to their expected performance since the results must be available near real-time, considering a high number of input events, e.g. 2.5 million per day for the use case at hand. As the target solution is required to be in place for years - if not decades – further implications on software engineering arise to guarantee robustness and maintainability over the whole application lifetime. The solutions are being evaluated at a major freight railway provider, which digitizes its fleet by equipping vehicles with telematic and sensor units having no external power supply (e.g. freight wagons or containers), instead using batteries with solar panels. Within 2020 the complete fleet of 65.000 vehicles will be equipped and will produce 2.5 million events per day. The event syntax follows an industry standard and provides for example information about the position of the vehicle, its speed and heading, as well as acceleration data across all three spatial dimensions. Further for selected vehicles information about the vehicles loading state is provided. Today's explored algorithmic approaches include the usage of dense, shallow neural networks for prediction, the usage of an advanced caching infrastructure, speeding up the retrieval of events, and first evaluations of complex event processing on streaming data. From a software engineering perspective, agile processes according to SAFe are established, while software quality is validated via automated test frameworks.

[1] accenture GmbH, Augustenstr. 1, 70178 Stuttgart, Germany, anusch.daemi-ahwazi@accenture.com
[2] DB Cargo AG, Edmund-Rumpler-Straße 3, 60549, Frankfurt a. Main, Germany, Daniel.Rost@deutschebahn.com

doi:10.18420/inf2020_67

2 Algorithmic details and challenges

From an algorithmic point of view two major approaches have been evaluated as to generate value from the vehicle events:

- The first approach uses a shallow neural network to forecast the optimal cache size of events for a single vehicle enabling quick retrieval of its route in the past. The neural network uses the event request history for a vehicle as input and predicts the optimal cache length and -strategy for each wagon based on the access patterns of the users. Refer to [Bl20] for further details.
- The second approach described in [We20] aggregates and maps multiple vehicles events driving on a route into a single unit, for example wagons assembled to a train or trucks in a convoy. The approach is based on cluster algorithms as well as a four-dimensional (time + spatial dimensions) fitting and interpolation of spare movement matrices.

Algorithmic challenges for the analysis of the event data arise in the following areas and new ideas are needed for efficient solutions:

- Correct detection of the driving direction and order of vehicles in a train is difficult to achieve using only a simple approach based on turn detection and heading. As the number of available data points is low the error of a false turn detection is multiplied along the trip. Instead, a constant determination of the wagon order based on GPS coordinates should be envisioned.
- Data validation of the vehicle events is challenging as more complex validations require at least two events from the vehicle event stream. Since the vehicle events are sent only in intervals of minutes or hours, an event correlation with at least two events will delay the original event from a business perspective unacceptably long.

3 Software engineering details and challenges

From a software engineering point of view the most prominent challenges lie in the area of validating the algorithmic approach with enough data of high quality. The testing environment is obviously smaller compared to production to save costs and hence the selection of a subset of representative vehicle events from the whole set is critical. Another challenge we face is the correlation of vehicle events with events from existing systems (e.g. transport orders or transport schedules), as these correlations are not unique due to missing, common identifiers. The correlation hence requires intelligent ways for data de-duplication with high performance. From an organizational perspective we need to ensure that all devised solutions can be maintained with minimal effort over a long period of time in an environment where the team which build the solution is not necessarily the same who maintains it. The primary challenge is the evolution of the implemented algorithms which should be understandable and modifiable even after years (from experience just having documentation is not enough), including a stable test infrastructure to easily reproduce the results of the algorithms.

Bibliography

[Bl20] Blatt, Maximilian: Developing a Fast, Scalable and Intelligent Caching System for Historic and Geospatial Data in an Enterprise Environment. Master's thesis, Mannheim University of Applied Sciences, Mannheim, Germany, February 2020.

[We20] Weiser, Andreas: Entwicklung und Implementierung eines Algorithmus für die digitale Abbildung des Güterzugverkehrs auf Grundlage von Sensordaten individueller Güterwagen. Master's thesis, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany, September 2020.