Brussels, Tuesday, 30 October 2018

Position Paper on Big data, Artificial Intelligence and their applications

Summary

Sensors of automated and connected vehicles produce huge amounts of data. In addition, data is gathered from road infrastructure sensors, such as cameras. The two categories are here referred to as Big traffic Data.
In combination with Big traffic Data, Artificial Intelligence (AI) techniques play a critical role in the development of automated driving technologies, since Big traffic Data feeds the machine learning algorithms that drive the continuous improvement of AI-based Connected and Automated Driving (CAD) functions towards higher-level automation. Conversely, AI techniques are essential for analysing and annotating collected Big traffic Data and converting it to useful information that can too be used in the development and validation of CAD functions.

Utilising Big traffic Data, AI techniques and their applications for CAD developments introduces several challenges in various domains. This paper summarise the different perspectives and priorities of the CARTRE thematic group members on these challenges. This group involves the CARTRE partners as well as stakeholders who are concerned with or interested in the topic. Throughout the CARTRE project, the inputs and debates of the thematic group members are provided in form of challenges and statements:

- *Challenges*: include existing and future expected challenges
- *Statements*: include diverse opinions about the addressed challenges and how much thematic group members agree with them.

In this paper, the challenges are clustered in four different categories: 1) technical, 2) policy, 3) organisational ecosystem and 4) user acceptance. The related statements and debates are also summarized. As a result of the intensive debates, the main impact of Big traffic Data and AI on CAD developments as well the input to the EU research agenda are provided.

CARTRE and SCOUT are funded by the European Union Horizon 2020 Work Programme

Introduction

Sensors of Connected and Automated Driving (CAD) systems continuously produce massive amount of data. This includes both macro (logistics, traffic flow) and micro real-life traffic data (individual car movements). In other words, it includes the data collected from road infrastructure sensors (e.g. camera, induction loops, laser scanners) and on-board vehicle sensors (e.g. camera, radar, LIDAR, ultrasonic, infrared and GNSS systems (e.g. Global Positioning Systems (GPS)), which are either raw or pre-processed. The macro-level traffic data is partially addressed by the CARTRE thematic area 'Digital and physical infrastructure'. In combination, micro and macro data form what is referred as Big traffic Data in this paper.

The world of Big traffic Data and Artificial Intelligence (AI) applications is not only rapidly growing but also very dynamic. The push towards increasing levels of automation of CAD systems forces the market to accelerate the development of new automated driving services, activities and applications including (but not limited to):

- Development of new Big traffic Data analytic tools which will lead to improved insights in driver modelling, driving and re-charge patterns, and understanding of the impact of CAD for safety, comfort and mobility.
- Validation of CAD systems through development of scenario-based assessment methodologies, including the classification of real-life scenarios, development of CAD functions, safety monitoring and validation, and traffic management.
- Improving situational awareness through development of applications using AI techniques, for example:
    - improved on-board sensor fusion to estimate the relative position and velocity of the surrounding vehicles - improved camera and radar fusion (see also CARTRE thematic area 'In-vehicle enablers'),
    - prediction models for other road user behaviour (especially complex behaviour like pedestrians) and
    - building accurate maps for on-road concurrent mapping and navigation.
- Improving the performance of CAD functions by self-improving mechanisms.
- Open new fields of applications such as taxi services, car sharing or find-a-parking-spot services.

To enable these developments and applications, many challenges in different domains and research areas have to be addressed. These challenges are categorised and discussed in the following section. Then, statements by thematic group members that are used for discussion are presented in two categories: common ground and open for discussion. In the final two sections, the long-term as well as short-term future research needs and their impacts are presented.

Challenges

CAD development processes need to find solutions for several challenges in different domains and research areas. It is expected that the challenges that are encountered until 2020 will most likely still carry key uncertainties in 2040. These challenges can be arranged under four different categories:

1. Technical challenges

This category includes the challenges to be addressed by technology enablers to collect, store, (pre)process and use real-life Big traffic Data. Two sub-categories can be defined:

1.1. (Big traffic) Data collection and storage

This sub-category includes hardware/software tooling for collecting, storing and sharing Big traffic Data.

- What CAD data is most valuable to store and share?
  This is a debatable point: one could claim that the value of CAD data depends on the stakeholder perspective. While others could claim that for researchers, any available CAD data is relevant.

CARTRE and SCOUT are funded by the European Union Horizon 2020 Work Programme

- Where and how to store real-life Big traffic data?
  Data is collected from many different sources and by different parties. The storage and accessibility of this data will become a real challenge at some point. Are data storage solutions, i.e. cloud solutions, huge servers, capable of handling this problem? Are new technology developments needed?
- How to handle the reducing validity of Big traffic Data over time?
  Some data may be still very relevant (incidents), other data is no longer representative.
- How to obtain sufficient high quality CAD data for analysis and development?
  It is of high importance that the collected data is useful for testing and validation of CAD functions. This leads to another challenge: how to determine and measure the quality of the data?
- What value in improved functionality can make sharing data attractive?
  Data sharing among different parties (researchers & OEMs) can be motivated by returning benefits to the data owner.
- How to provide reliable and privacy proof interfaces for Big traffic Data repositories?
  Different data sources communicate through safeguard interfaces. This includes interfacing with Internet of Things (IoT) and cloud repositories. On a 2020 horizon, this appears to be mostly a technical issue. However, on a 2040 horizon, we need to have addressed the full ethical, social and political implications.

1.2. Big traffic Data processing and AI techniques and applications

This sub-category includes hardware infrastructure and software tools for processing the collected data and converting it into valuable information. It also includes AI data-driven applications, e.g. environmental perception, prediction and modelling of other road user behaviour.

- How to compare algorithms in a reproducible way? How to compile a training data set in such a way that no bias is expected?
  Algorithms and AI applications should be tested and compared using generic approach and unbiased data set.
- How to obtain the ground truth information (e.g. from video streams) without extensive manual work?
  For algorithm training, the ground truth ('What really happened?') must be available and reliable. This may include external reference systems, human annotation in real time or in retrospect. The reliability of the annotation will determine the reliability of the algorithms.
- How to assess the completeness of the scenarios used in training an AI function? How to select a data set for validating specific AI CAD function?
  This point mainly addresses the completeness of testing. In other words, is the data/scenarios set that is used for training or testing a certain CAD functionality sufficient to represent real world driving.
- Are existing black box AI techniques able to cope with new challenges introduced by CAD?
  This suggests that a new revolution for AI is needed to satisfy the CAD needs toward higher level of automation including transparency and traceability of these techniques.

2. Policy challenges

This category includes the challenges that are or will be faced by policy makers and legal organizations on the national and EU levels concerning the reusability of collected data, developed AI techniques and sharing them between different parties.

- How to overcome privacy and security barriers for sharing Big traffic Data?
  This requires appropriate anonymisation solutions and security approaches on a 2020 horizon. For 2040, the views and policies on privacy and surveillance may have completely changed (key uncertainty).
- What is an acceptable distinction between company-sensitive data, personally sensitive data and research-relevant data?
  Various stakeholders have legitimate concerns on privacy, liability, confidentiality and security. The data sets need to be handled accordingly.
- What type of data owned by different stakeholders could be shared across them and others and for which purposes?

CARTRE and SCOUT are funded by the
European Union Horizon 2020 Work Programme

Data type and processing state define the value of a given data set. Data sharing practices will depend on the sharing partners and the intended use of the data.

3. Organisation ecosystem challenges

This category includes the challenges that are or will be faced by (competitive) decision makers and researchers in the automotive industry (car manufacturers, suppliers and research labs) to share/reuse each other's driving/traffic data for research and development.

- How to handle different data sharing needs across various stakeholder types?
  The sharing needs do not match automatically. What fosters sharing for particular constellations? Should data sharing remain voluntary or be prescribed?
- How can we share the investments for developing AI functions?
  Currently, massive parallel investments are made in machine learning for environmental perception using sensor systems. Is there a way to cooperate and exploit synergies?
- How to stimulate, convince and manage openness about driving data?
- How can we set up a framework in which OEMs share driving data and road operators share infrastructure data?
  The benefits of data sharing and the degree of sharing that are required to ensure international competitiveness of European research, development and innovation should be investigated further. Which incentives can be created for data sharing?

4. User acceptance

This category includes the challenges that are or will be faced by decision makers and the automotive industry (car manufacturers, suppliers and research labs) to promote CAD technologies to the end-users.

- How should the difference in media coverage for human and AI errors be approached?
- How much safer must an AI driver be compared to human drivers to foster user acceptance?
  Due to the difference in acceptance for accidents caused by humans or AI, which is further exacerbated by the media coverage, automated vehicles must significantly outperform humans. How much better must AI systems perform? How can a balanced reporting of benefits and challenges be achieved?

Statements

In this section the statements by thematic group members are presented. The statements are used for discussion and not considered as definite answers. Statements are presented in two main categories: common ground and open for discussion

1. Common ground statements (agreed on)

- Driving data and sensor data only becomes valuable with good annotation (ground truth).
- Not every bit and byte of data is worth sharing. The data quality determines the usability of the data.
- An AI function will evolve from 'near real time' to 'real time' to 'before real time', including prediction of other road user behaviour.
- We need an agreement on benchmark data sets for validating AI functions.
- We need performance criteria for an AD function which includes AI.
- We need agreement on how to test a certain CAD function built with AI.
- Machine learning needs to be integrated in the automated vehicle to deal with new situations.
- The more automated systems on the street, the easier to develop a vehicle for it.
- For connected cars, the security of the data exchange is crucial for safety and privacy.
- Any data where the driver can be recognized (e.g. video, GPS destination, license plate) is personal data, thus governed by the General Data Protection Regulation (GDPR).
- Any anonymised data showing behaviour of other road users can be shared (without compromising privacy).
- Some Big traffic Data is so sensitive to the OEMs that it will not be shared.
- The EU should find ways to enforce and/or stimulate data sharing like open data pilots.

CARTRE and SCOUT are funded by the European Union Horizon 2020 Work Programme

2. Open for discussion statements (not agreed on)

- The value of CAD data depends on the stakeholder perspective.
- For data sharing, semantic compatibility on the data sets is required and not standardized format. The power and the beauty of big data techniques are to be able to deal with different format which are interlinked at a semantic level.
- Vehicles will become adaptive, learning systems.
- We need AI to identify the best combination of scenarios for training automated driving functions.
- After the hype, machine learning will become a technique just like any other.
- OEMs will share data using a common web interface for services.
- OEMs only want to share driving data for further improvement of the vehicle.

Future research needs

There are various challenges in sharing and using Big traffic Data and AI, both in the short and long term. The CARTRE working group concluded the following focus points for the EU research agenda:

1. Short-term research needs

- **Data sharing:** Policy and regulations for data ownership are strongly needed to stimulate and encourage car manufacturers, other institutions and even individuals to share driving data. Very diverse opinions exist on who owns the data. It is unclear whether data ownership may be regulated or driving data may become open within privacy, security and confidentiality constraints. This has a great impact on the re-usability of data that has been collected or aggregated. The importance of data sharing is underlined by its relevance for data-hungry machine learning algorithms.
- **Privacy and security:** Policy and ethics on the appropriate use of driving data are needed. Big traffic Data has a significant impact on privacy and security. Current policies are fragmented and were not prepared with Big traffic Data in mind. The willingness of users to accept limited privacy may also increase, if a clear returned benefit is perceived.
- **Regulations, ethics and liability:** A framework adapted to the use of AI functionality is required. The shift from human to AI vehicle control poses fundamental questions concerning regulations, ethics and liability, e.g. concerning the programming of AI behaviour for accidents and the subsequent determination of the guilty party.
- **Data storage and accessibility:** A framework for storing and accessing shared driving data must be defined. Data is collected from many different sources, by different parties and in different formats. Technically the storage and accessibility of this data will become a real challenge at some point. This may require developing new data storage solutions as well as new strategies for data reduction and removal (i.e. what to keep and what to remove and when?).
- **Assessment and validation of CAD functions:** There is an urgent need for a harmonised framework for the validation and assessment of CAD functions. This includes the harmonisation of test cases and training data sets for AI functions and the question of completeness of training scenarios.

2. Long-term research needs

- **Investment for developing CAD functions:** Encourage and enable sharing the investments for developing AI functions for automated driving. Currently, massive parallel investments are made in machine learning and AI technologies. Sharing investment will accelerate these developments and naturally leads towards harmonisation.
- **Evolution for new AI techniques:** Encourage new initiatives working on developing new AI techniques that fulfil future CAD functionality needs (including transparency and traceability) especially in the research and academic sectors.

CARTRE and SCOUT are funded by the
European Union Horizon 2020 Work Programme

Expected Impact

Several impacts are expected when the above mentioned proposed agenda is supported, among those

- Promoting/accelerating development of CAD
- Closer cooperation between stakeholders
- Increase users trust and acceptance for CAD
- Ease procedure toward new standard traffic regulation for CAD
- Providing solutions for several issues concerning liability AI based CAD

CARTRE and SCOUT are funded by the
European Union Horizon 2020 Work Programme