A/B test driven Systems Engineering

Publik rapport

Författare: David Issa Mattos
Datum: 2024-04-04
Projekt inom Electronics, software and communication

FFI Fordonsstrategisk Forskning och Innovation
VINNOVA Energimyndigheten TRAFIKVERKET
FKG VOLVO SCANIA VOLVO

Innehållsförteckning

Kort om FFI

FFI är ett samarbete mellan staten och fordonsindustrin om att gemensamt finansiera forsknings- och innovationsaktviteter med fokus på områdena Klimat & Miljö samt Trafiksäkerhet. Satsningen innebär verksamhet för ca 1 miljard kr per år varav de offentliga medlen utgör drygt 400 Mkr.

Läs mer på www.vinnova.se/ffi.

Instruktioner

- *Instruktioner är skrivna med brun kursiv text – all sådan text tas bort innan redovisning sker.*
- *Ingen begräsning finns för antalet sidor*
- *Infoga inte dokument eller objekt som måste öppnas för att läsas eller ses. Det fungerar inte då slutrapporten omvandlas till en PDF-fil när den laddas upp i portalen.*
- *Denna rapport är publik vilket innebär att rapporten är tillgänglig för alla utan att någon sekretessgranskning kommer att göras. Rapporten läggs ut på FFI:s hemsida.*

1 Sammanfattning

Detta projekt startar i den pågående omvandlingen av fordonsbranschen, från att vara konkurrent/branschdriven och hårdvarufokuserad, till en kundorienterad mjukvaruverksamhet. Det uppkopplade fordonet, som kan både ladda upp och ladda ner data och mjukvara via luften, möjliggör ett skifte i utvecklingsmodell, från kravdriven utveckling till datadriven utveckling, där team använder data från produkter i händerna på kunder för att utvärdera utvecklingsframstegen. Vår huvudhypotes är att denna data kan hjälpa team att fatta rätt beslut – med hjälp av datautvärdering, "värdemodeller", på flera nivåer i företaget för att säkerställa anpassning och ett gemensamt mål. För att validera denna hypotes har vi utfört empiriska studier och pilotförsök i kundfordon i samarbete med de inblandade industriella partners.

2 Executive summary in English

This project starts in the ongoing transformation of the automotive business, from being competitor/branch driven and hardware focused, to a customer-oriented software business. The connected vehicle, able to both upload and download data and software over the air, enables a shift in development model, from requirement driven development to data driven development, where teams utilize data from products in the hands of customers to evaluate the development progress. Our main hypothesis is that this data can help teams to make the right decisions - using data evaluation, "value models", at several levels in the company to ensure alignment and a common goal. To validate this hypothesis, we performed empirical studies, proof of concepts and pilot cases in customer vehicles in collaboration with the involved industrial partners.

3 Bakgrund

The automotive industry is quickly transforming to a software and robotics industry. We see an opportunity to initiate a transition from classic competitor/branch driven development to “data driven” development utilizing Over The Air (OTA) connectivity to vehicles. This technology enables direct collection of data from the customer’s car as well as software updates with new or improved functionality – a direct link between the developing team at R&D and the customer. In this new way of development, known already from IT/WEB business, development teams can work directly towards quantitative targets, using e.g. customer data and product metrics. These targets are also continuously evaluated using data collection to ensure that the team as well as the organization/product is moving towards a common goal. This is a fundamental shift from the traditional requirements driven approach in which functionality is built according to a specification and where teams have little impact on how to reach the desired solution. One also note the inherent shift from top down hierarchical decision making towards more empowered teams with end to end responsibility for their features. This second shift is since long time recognized in software business and in “agile” transformations. However, as a prerequisite for teams to be empowered and adopt autonomous ways-of-working in a large and distributed organization like automotive OEMs, fast and reliable feedback cycles and data analysis at many levels is essential to ensure that what the team optimize for at a feature level align and support the overall and company level strategy and goals. Already now, some approaches exist that

realize parts of the vision we outline here, e.g. A/B testing. In this technique, teams work in exploratory and experimental ways to develop and improve functionality based on data from carefully selected groups of customers receiving different versions, in very short loops. Translation to automotive systems is not straightforward, but OEMs in the software development front as Tesla are already utilizing similar methods using OTA software updates. However, and to the best of our knowledge, there is currently no documented approach for how to maximize team performance by adopting data driven development on larger scale, balancing multiple teams and sub organizations to get true alignment. In our view, the benefits with such an approach would significantly accelerate product development as well as development of new digital and data driven services and solutions in the automotive industry. With VCCs current investments in data collection and user experience (UX), big data analysis, centralized computation and in-house software development, there are tremendous opportunities to explore how to (1) maximize use of customer and product data, (2) start utilizing A/B test methods and data driven evaluation using OTA and company cars (3) support the young agile organization in how benefit from this data and new methods to align team and business goals.

4 Syfte, forskningsfrågor och metod

This project objective is addressing the industrial needs in an expected near future when the electrified premium vehicle is a connected device in an increasingly digitalized transport system. In this near future the car is a true software product, or service, and its development must follow the state of the art of software development at the time – including methods as Continuous Deployment and efficient utilization of User Data collection. Note that user data is not limited to UX or user interaction, but can also be any data collected from the product – as e.g. motor performance. There are to our best knowledge no other alternatives to stay competitive as an OEM.

In this project we have worked with the following general research hypotheses:
H1: By shortening feedback loops and enabling continuous analysis of real time customer data, teams will have an effective mechanism for organizational alignment in addition to the ability to take rapid and correct decisions.
H2: A hierarchical framework for value modeling at a team – system – business level [1], forms an effective basis for alignment in a large and distributed organization developing complex products and systems

In practice these hypotheses have been translated, according to the scope of this project that,

(1) introducing experimentation and causal inference into the automotive development process we are able to shorten the feedback look from customers to development teams.
(2) The use of metrics and data collected from experiments and causal inference analysis on customer data can be used to create a hierarchical view of the value delivered by different development efforts, and this can be used as an alignment tool for communication, prioritization and development of complex products and systems

The following research questions have been used to investigate both hypotheses
RQ1: What are the main challenges seen in practice when running A/B test in a large-scale automotive in commercial vehicles?

RQ2: How can the challenges identified in RQ1 be addressed, which are still open challenges and what are the lessons learned from the tried solutions?

RQ3: How to run large-scale online experiments in the automotive domain in a fast and reliable fashion?

RQ4: How can the inherent limitation of sample sizes in the automotive domain be addressed?

RQ5: Can causality be concluded in the absence of randomization in automotive software evaluation?

RQ6: How can causal inference models be used in communication and as boundary objects between different teams?

To address these research questions, we have used a range of different research methods.
RQ1 and RQ2 have been based on an interview-based case study research method with two of the companies participating in this project. Both manufacture commercial vehicles for private and enterprise consumers.

Research question three utilized a mixed method approach, that combined a literature review, case studies and empirical experiments.

RQ4 has also utilized a mixed method approach combining case studies, empirical experiments, and empirical validation in real cases.

RQ5 has utilized an empirical case further validated by empirical experiments in internal customer vheicles

RQ6 has used concrete cases of causal graphical models to support communication and discussion across teams, as both a visualization tool as well as to communicate physical and theoretical aspects of complex problems. We have utilized a case study and experience reports based on concrete cases conducted in one of the companies

5 Mål

Din text här…

This project proposes the use of new development methods to increase the competitiveness of the Swedish automotive industry in both development speed and quality. The proposed development methods are based on an experimentation methodology (e.g. A/B testing) that aims to understand causal impact of development efforts. This methodology has been successfully applied in a wide range of software development industries, but it is new to automotive software development.
To enable the use of such methodologies, it is required to integrate products (vehicles) with the local IT-infrastructure making use of V2X communication, agile development of connected functions and data utilization to support faster and more accurate decision making.
This type of data driven development shortens the loop between developers and customers through vehicle communication and development iterations, compared to the traditional survey-based approach used by most automotive manufacturers and suppliers.
The complexity of today's vehicles the higher user demand of car-technology and integration requires the used of more efficient development process. One of the largest use cases is to improve energy efficiency of vehicles. By understanding customer usage, being able to deploy new features and create interventions fast, companies can verify that in every deployment the vehicle is fulfilling energy requirements and leading optimization in this area.

6 Resultat och måluppfyllelse

This project has delivered several academic publications, company internal tools, training material and software artifacts. Below we list the contributions that are available publicly.
The project resulted in the following licentiate thesis presented at Chalmers:

1. *Yuchu Liu. "Online experimentation in automotive software engineering", Chalmers University of Technology, Department of Computer Science and Engineering, Licentiate thesis, 2022*

The project resulted in the following academic publication presented in international conferences:

1. *David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Aita Maryan Korshani, Jonn Lantz "Automotive A/B testing: Challenges and lessons learned from practice" Published in the 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 2020*

2. *Yuchu Liu, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, "An architecture for enabling A/B experiments in automotive embedded software". Published in 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). 992-997.*

3. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, "Size matters? Or not: A/B testing with limited samples in automotive embedded software". Published in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 300-307.*

4. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, "Bayesian propensity score matching in automotive embedded software engineering". Published in 2021 28th Asia-Pacific Software Engineering Conference (APSEC 2021). 233-242.*

5. *David Issa Mattos and Yuchu Liu, "On the Use of Causal Graphical Models for Designing Experiments in the Automotive Domain", Published at the 2022 25th Evaluation and Assessment in Software Engineering (EASE 2022),*

The following publication was submitted to an international academic journal and the preprint is available in the provided link

1. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, "Bayesian causal inference in automotive software engineering and online evaluation". In submission to a Software Engineering journal. Preprint available at: https://arxiv.org/abs/2207.00222*

The following public software repositories contain artifacts that are connected to this project and support the publications with the tools and developed method:

- *https://github.com/yuchueliu/BOAT*
- *https://github.com/yuchueliu/BPSM*
- *https://github.com/davidissamattos/ease-2022-causal*

Of the general FFI goal of "Strengthening the international competitiveness" we achieved this goal by creating a large awareness in the need for online experiments and how it leverages data to develop and evolve products that are valued by the customers in the highly competitive automotive industry. This included the development of internal processes to handle that online experiments in internal customer vehicles, development of product usage kpis, competence building on how to conduct and analyze experiments and several use cases.

For the general goal EMK goal of "Electronics, Software and communication" and the subarea "Architecture outside the vehicle, the connected system" we developed data analysis methods and the required infrastructure to analyze with speed experiment data coming from the vehicles. This infrastructure is aligned with the software and electronic architecture evolution of the new vehicles and takes advantage of V2X communication where two of the partners (Volvo Cars and Zenseact) are in the forefront in the automotive industry.

For the subarea "Human-machine interaction", the awareness and methods we developed were used to improve the interaction of the user with the system to deliver better experience in different areas.

For the subarea "Green, safe, autonomous and connected functions" we utilize the developed infrastructure to improve energy consumption in different areas of our electric and plugin hybrid vehicle. Use cases for autonomous vehicles were evaluated but no concrete application was yet made.

7 Spridning och publicering

7.1 Kunskaps- och resultatspridning

Hur har/planeras projektresultatet att användas och spridas?	Markera med X	Kommentar
Öka kunskapen inom området	X	We have developed tools and presented novel ideas in academic conferences exploring and extending the knowledge in the area of automotive online experimentation. Additionally, we have gained knowledge on how to introduce such techniques in the context of large scale agile organization
Föras vidare till andra avancerade tekniska utvecklingsprojekt	X	The pilot projects that we run have been used as basis to run other experiments in the technical development of projects
Föras vidare till produktutvecklingsprojekt		This technique is not intended to lead to the development of a new product in the automotive sector, but rather to evaluate and improve solutions in existing products
Introduceras på marknaden		Partially due to technical limitation and part because of the approval process for software to be introduced in a vehicle we are currently limited to internal customer vehicles. This technique has not been used in external customer vehicles
Användas i utredningar/regelverk/ tillståndsärenden/ politiska beslut		This has not been in the scope of this projects

7.2 Publikationer

The project resulted in the following licentiate thesis presented at Chalmers:

1. *Yuchu Liu. "Online experimentation in automotive software engineering", Chalmers University of Technology, Department of Computer Science and Engineering, Licentiate thesis, 2022*

The project resulted in the following academic publication presented in international conferences:

1. *David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Aita Maryan Korshani, Jonn Lantz "Automotive A/B testing: Challenges and lessons learned from practice" Published in the 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 2020*

2. *Yuchu Liu, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, "An architecture for enabling A/B experiments in automotive embedded software". Published in 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). 992-997.*

3. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, “Size matters? Or not: A/B testing with limited samples in automotive embedded software”. Published in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 300-307.*

4. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, “Bayesian propensity score matching in automotive embedded software engineering”. Published in 2021 28th Asia-Pacific Software Engineering Conference (APSEC 2021). 233-242.*

5. *David Issa Mattos and Yuchu Liu, "On the Use of Causal Graphical Models for Designing Experiments in the Automotive Domain", Published at the 2022 25th Evaluation and Assessment in Software Engineering (EASE 2022),*

The following publication was submitted to an international academic journal and the preprint is available in the provided link

6. *Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, Jonn Lantz, “Bayesian causal inference in automotive software engineering and online evaluation”. In submission to a Software Engineering journal. Preprint available at: https://arxiv.org/abs/2207.00222*

The following public software repositories contain artifacts that are connected to this project and support the publications with the tools and developed method:

1. *https://github.com/yuchueliu/BOAT*
2. *https://github.com/yuchueliu/BPSM*
3. *https://github.com/davidissamattos/ease-2022-causal*

8 Slutsatser och fortsatt forskning

Experimentation and causal inference in automotive software development allows companies to focus resources and understand customer usage and its interaction with the environment in a systematic way that leads to data-driven decision making and increase development efficiency across the product. Nevertheless, this topic is new in the automotive domain from both research and implementation aspects. Most of our efforts in this project was to understand the current limitations in practice and propose tools on how to overcome them for future scaling in a continuous data-driven software development.

To enable software experimentation in a broader setting beyond user experience tweaks one must leverage from causal modelling. Similarly, to most research efforts in causal inference and experimentation, we assume the input and output variables are known as well as the structural process. However, in practice, the causal structure might be vastly unknown prior to the experiments. This means that we cannot confidently design and conduct and experiment and be certain of its benefits for the development organizations. Therefore, a causal discovery process might be required prior to design of an experiment especially in complex systems such as autonomous drive or active safety related functions. Future research can explore the area data- and knowledge-driven causal discovery methods, and run empirical studies to analyze their applicability in automotive software engineering.

On the aspect of scaling, future work will likely see many challenges when experimentation is applied in a broad and diverse customer population. The presence of heterogenous groups, i.e. subgroups of the population that present different effects, and vehicle localization variants (vehicles tailored for a specific subgroup) can lead to challenges in planning and designing these experiments. For instance, understanding how much of the effect is due to the modification introduced or due to the localization features when developing common software. We started to explore that in the publication 5 ("On the Use of Causal

Graphical Models for Designing Experiments in the Automotive Domain") but this is an area largely underexplored.

Finally, experimentation might see some new challenges for future research in face data protection legislations, where experiments will need to be localized and results aggregated for a global market assessment. Causal inference tools, such as the ones we explored, and meta-analysis techniques can provide useful tools to address such potential challenges.

For more insights and detailed information about the results and the scientific method utilized in this project we refer to the attached academic publications.

9 Deltagande parter och kontaktpersoner

David Issa Mattos – Volvo Cars
Jan Bosch – Chalmers University
Helena Holmström Olsson – Malmö University
Sofie Tressing – Zenseact
Peter Thorngren – Volvo GTT