Vezető Azonosítás CAN Logok Alapján

Remeli Mina

CrySyS Lab, BME

www.crysys.hu

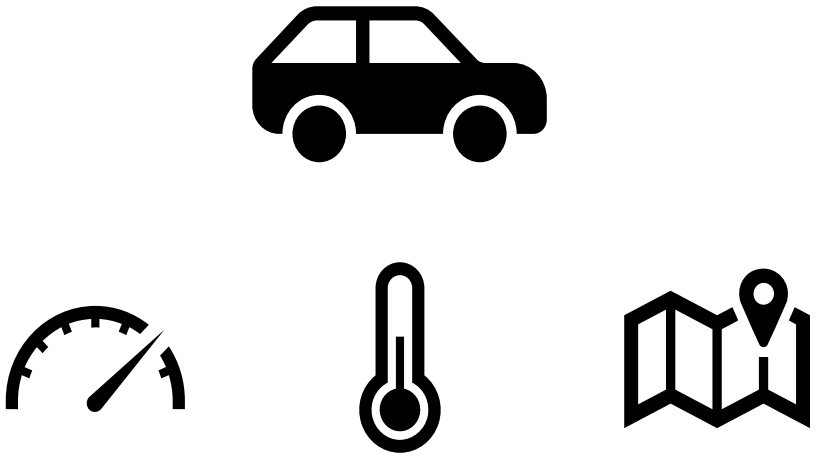
remeli@crysys.hu

Áttekintés

- Bevezető: jármű szenzorok
- Vezető azonosítás
- Privacy megfontolások

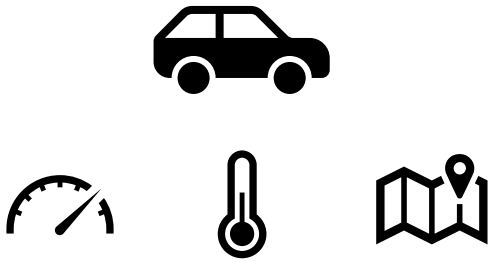
Az előadásom célja

- Általános képet mutassak arról, hogy...
  - Mire használják jelenleg a vezetés közben gyűjtött adatainkat
  - Ezen adatok alapján való azonosíthatóság következményei
    - »GDPR

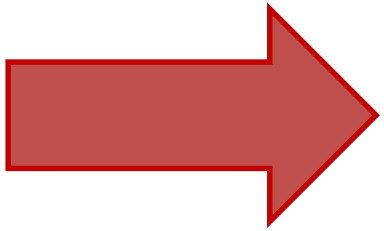


JÁRMŰ SZENZOROK

Mi a szerepük?



Szenzorokból kinyert
vezetési adatok



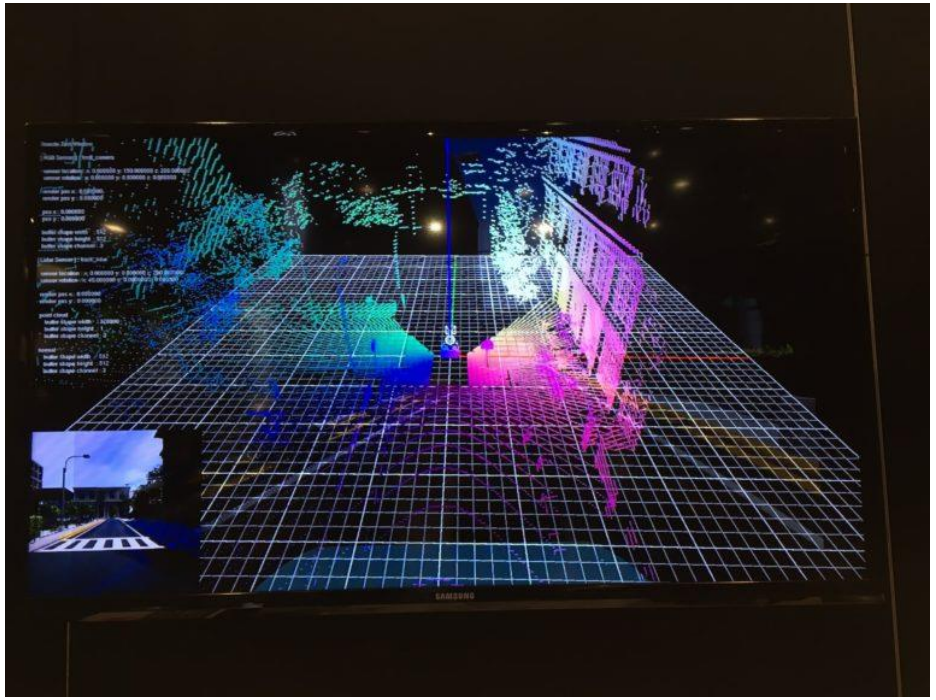
Vezető

Jármű szenzorok

- Amiket a műszerfalon is látunk
  - Fogyasztás
  - Üzemanyagszint
  - Sebesség
  - Fordulatszám
  - Lokáció (gps)

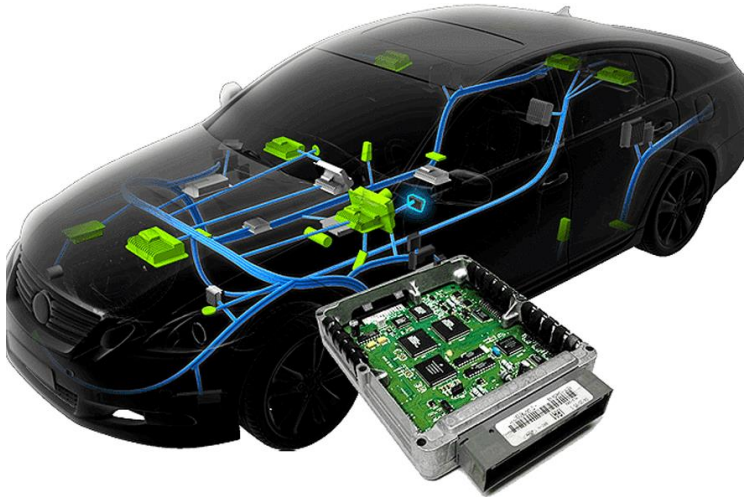
Egyéb szenzorok

- Kormány elfordulás mértéke
- Kuplung / gáz / fék pedál pozíciója
- Oxigén szenzor (kipufogóban)
- Üzemanyag hőmérséklet szenzor
- Kamera
- Radar és LiDAR technológiák
- stb...



Adatok feldolgozása

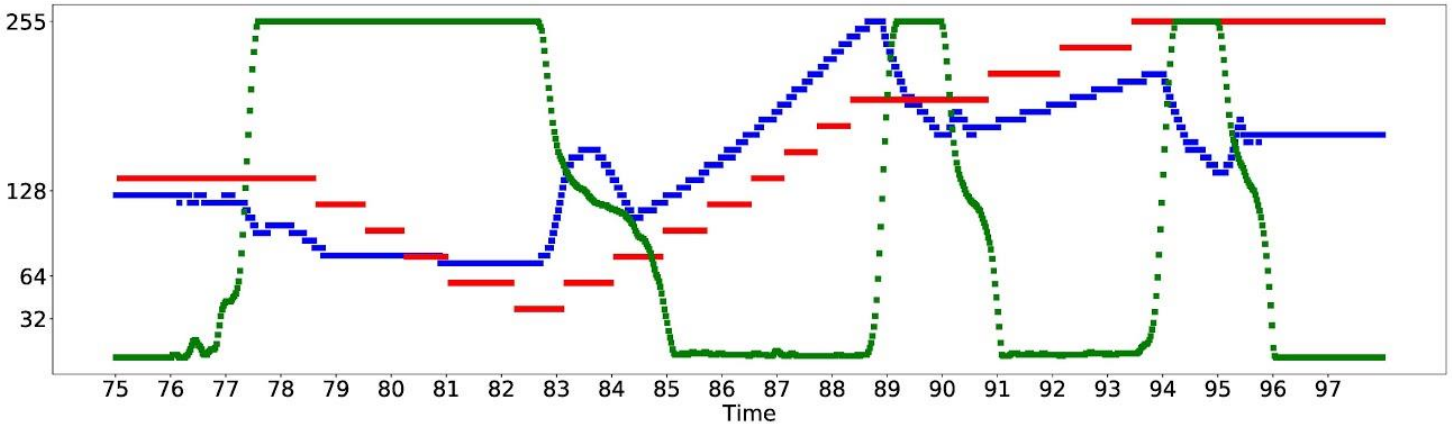
- ECU (Electronic Control Unit)
  - Kommunikáció
    - » Egymás között
    - » Szenzorokkal
  - Feladatok
    - » Adatok feldolgozása
    - » Feldolgozott adatok továbbküldése
  - Kommunikációs csatorna:
    - » CAN busz



Controller Area Network

- CAN busz üzenetek:
  - Üzenetformátum: fejrész + adat
    - » Fejrészben: üzenet ID

Timestamp	CAN-ID	Req	Len	Data
1481492683.285052	0x0208	000	0x8	0x00 0x00 0x32 0x00 0x0e 0x32 0xfe 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc8 0x00 0x0f 0x03 0x00 0x92 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc9 0x00 0x0f 0x00 0x00 0x92 0x4c
1481492674.736055	0x02c4	000	0x8	0x82 0xcc 0x00 0x0f 0x08 0x00 0x92 0x5a
1497323915.123844	0x018e	000	0x8	0x03 0x03 0x00 0x00 0x00 0x00 0x07 0x3f
1497323915.112910	0x00f1	000	0x6	0x28 0x00 0x00 0x40 0x00 0x00
1481492674.736055	0x02c4	000	0x8	0x82 0xd2 0x00 0x0f 0x0c 0x00 0x92 0x5d
1481492674.736055	0x02c4	000	0x8	0x82 0xa1 0x00 0x0f 0xa1 0x00 0x92 0x4d



1., 4., 7. bájt a 0x02c4 ID-jű üzenetben

Hozzáférés…? Könnyű!

- CAN adatok nincsenek rejtjelezve
- OBD dongle
  - Olcsó
  - Könnyen beszerelhető
    - » OBD-II port-ra csatlakozik

- Viszont az adat nem könnyen értelmezhető...
  - » Nem tudni, hogy egy adott ID-jű üzenet melyik bájtján milyen adat utazik.
  - » Nincs szabvány, gyártófüggő.

Adat = pénz

’’McKinzey&Co, egy menedzsment-tanácsadó cég becslése szerint az autókból nyert adatok értéke 2030-ig akár 750 milliárd dollárt fog érni!’’ [1]

- Reklám
- Fizetős szolgáltatások
- Harmadik félnek való értékesítés

Példák

Ford Telematics™

Fleet management software from Ford

Optimizing your fleet to help your operation run more efficiently.

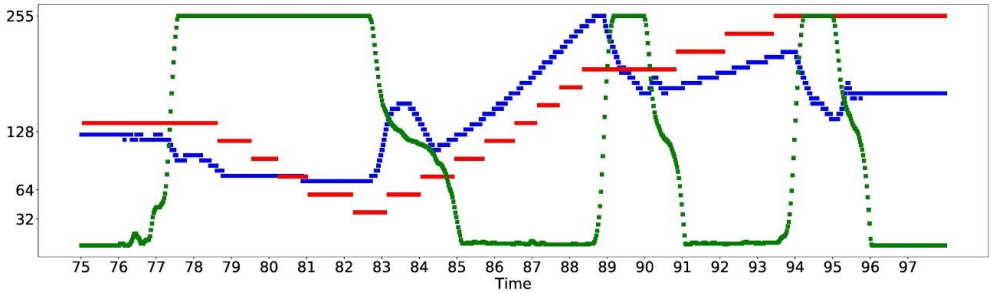
Példák

Példák

VEZETŐ AZONOSÍTÁS

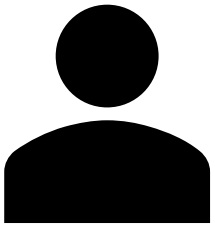
Mit nevezünk vezető azonosításnak?

Timestamp	CAN-ID	Req	Len	Data
1481492683.285052	0x0208	000	0x8	0x00 0x00 0x32 0x00 0x0e 0x32 0xfe 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc8 0x00 0x0f 0x03 0x00 0x92 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc9 0x00 0x0f 0x00 0x00 0x92 0x4c
1481492674.736055	0x02c4	000	0x8	0x82 0xcc 0x00 0x0f 0x08 0x00 0x92 0x5a
1497323915.123844	0x018e	000	0x8	0x03 0x03 0x00 0x00 0x00 0x00 0x07 0x3f
1497323915.112910	0x00f1	000	0x6	0x28 0x00 0x00 0x40 0x00 0x00
1481492674.736055	0x02c4	000	0x8	0x82 0xd2 0x00 0x0f 0x0c 0x00 0x92 0x5d
1481492674.736055	0x02c4	000	0x8	0x82 0xa1 0x00 0x0f 0xa1 0x00 0x92 0x4d



Kuplung
Sebesség
Fordulatszám

Machine Learning



Vezető azonosítás egy kanyarból

Driver Identification Using Automobile Sensor Data from a Single Turn

David Hallac*, Abhijit Sharang*, Rainer Stahlmann‡, Andreas Lamprecht‡, Markus Huber†, Martin Roehder†, Rok Sosič*, Jure Leskovec*

*Stanford University {hallac, abhisg, rok, jure}@stanford.edu
†Volkswagen Electronics Research Laboratory {markus.huber, martin.roehder}@vw.com
‡AUDI AG {rainer.stahlmann, andreas.lamprecht}@audi.de

***Abstract*—As automotive electronics continue to advance, cars are becoming more and more reliant on sensors to perform everyday driving operations. These sensors are omnipresent and help the car navigate, reduce accidents, and provide comfortable rides. However, they can also be used to learn about the drivers themselves. In this paper, we propose a method to predict, from sensor data collected at a single turn, the identity of a driver out of a given set of individuals. We cast the problem in terms of time series classification, where our dataset contains sensor readings at one turn, repeated several times by multiple drivers. We build a classifier to find unique patterns in each individual's driving style, which are visible in the data even on such a short road segment. To test our approach, we analyze a new dataset collected by AUDI AG and Audi Electronics Venture, where a fleet of test vehicles was equipped with automotive data loggers storing all sensor readings on real roads. We show that turns are particularly well-suited for detecting variations across drivers, especially when compared to straightaways. We then focus on the 12 most frequently made turns in the dataset, which include rural, urban, highway on-ramps, and more, obtaining accurate identification results and learning useful insights about driver behavior in a variety of settings.**

soon as the car turns out of the driveway, which member of a household is currently driving it. The vehicle could then automatically adjust the settings to fit the driver's preferences (temperature, radio station, mirror placement, etc.). Furthermore, correctly identifying drivers would allow cars to build driver profiles. Vehicles would be able to determine if certain drivers are more aggressive than others, or if some prefer back-road routes to the main roads when navigating. Analyzing behavior at such small granularity would also allow for profiles of each segment of road [8], for example warning the driver to be careful if the car had previously needed to use the emergency brake at an upcoming intersection. Additionally, this type of analysis could detect changes in driver behavior throughout a drive, such as when an individual uses a handheld cellphone and becomes distracted, since this would manifest itself as a sudden shift in the driving patterns. Note that all of these applications can be implemented locally, without the need for global coordination between different cars. This is imperative because it keeps driver information private, and no

”Vezető DNS”

Characterizing the “Driver DNA” Through CAN Bus Data Analysis

Umberto Fugiglando
MIT Senseable City Laboratory
Cambridge, Massachusetts
umbertof@mit.edu

Paolo Santi
Istituto di Informatica e Telematica del CNR
Pisa, Italy
MIT Senseable City Laboratory
Cambridge, Massachusetts
psanti@mit.edu

Sebastiano Milardo
University of Palermo
Palermo, Italy
MIT Senseable City Laboratory
Cambridge, Massachusetts
sebastiano.milardo@unipa.it

Kacem Abida
VW Group Electronics Research Laboratory
Belmont, California
kacem.abida@vw.com

Carlo Ratti
MIT Senseable City Laboratory
Cambridge, Massachusetts
ratti@mit.edu

ABSTRACT

People's driving behavior is influenced by different human and environmental factors, and several attempts to characterize it have been proposed. Nowadays, the standardization of the CAN bus and the increase of the electronic components units in modern cars offer a large availability of sensors data that make possible a more reliable and direct characterization of driving styles. In this work, we propose the concept of "Driving DNA" as a way of describing the complexity of driving behavior through a set of individual and easy-to-measure quantities. These quantities are responsible for some aspects of the driver's behavior, just as – in the metaphor – genes are responsible for the tracts of an individual.

The concept has been tested on a dataset collected from the CAN bus consisting of more than 2000 trips performed by 53 people, in a wide scenario of road types and open traffic conditions. The Driving DNAs have been calculated for each person, and a graphical visualization of their comparison is provided.

1 INTRODUCTION

The process of understanding and characterizing human driving behavior has recently gained a lot of importance due to its fundamental applications in connected autonomous vehicles, electric vehicles and artificial transportation systems [11, 17]. Moreover, as driving style influences accident risk [1] and fuel efficiency [10], technologies able to classify driving behavior can improve safety and car eco-friendliness.

Differently from earlier studies, where models were based only on GPS location [7], these can now rely on multiple layers on information coming from several hundreds of sensors and electronic control units (ECUs) embedded in the car, whose intercommunication is made feasible through the CAN bus technology [9]. This not only implies richer and higher quality data, but expands the possibilities of large-scale data collections and extensive sensing applications [12].

Driving behavior (or driving style) has no unique definition nor measure, and it is a combinations of mixed factors and components

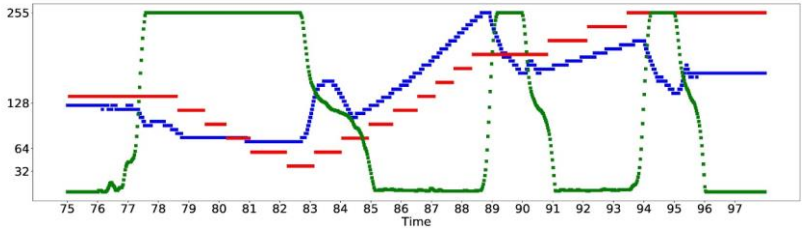
Mik a lehetőségek?

- Vezető azonosításon alapuló szolgáltatások
  - Személyre szabott vezetési élmény
  - Értesítés ha más vezeti az autónkat
  - Személyre szabott reklámok

Ami közös bennük

- Visszafejtett adatokon dolgoztak
- A vezetők fix útvonalon vezettek
- 2-15 vezető

Timestamp	CAN-ID	Req	Len	Data
1481492683.285052	0x0208	000	0x8	0x00 0x00 0x32 0x00 0x0e 0x32 0xfe 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc8 0x00 0x0f 0x03 0x00 0x92 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc9 0x00 0x0f 0x00 0x00 0x92 0x4c
1481492674.736055	0x02c4	000	0x8	0x82 0xcc 0x00 0x0f 0x08 0x00 0x92 0x5a
1497323915.123844	0x018e	000	0x8	0x03 0x03 0x00 0x00 0x00 0x00 0x07 0x3f
1497323915.112910	0x00f1	000	0x6	0x28 0x00 0x00 0x40 0x00 0x00
1481492674.736055	0x02c4	000	0x8	0x82 0xd2 0x00 0x0f 0x0c 0x00 0x92 0x5d
1481492674.736055	0x02c4	000	0x8	0x82 0xa1 0x00 0x0f 0xa1 0x00 0x92 0x4d



Kuplung
Sebesség
Fordulatszám

gyártó

A mi munkánk

Automatic Driver Identification from In-Vehicle Network Logs

Mina Remeli, Szilvia Lestyán, Gergely Acs, and Gergely Biczók
CrySyS Lab, BME-HIT, Hungary
{remeli, lestyan, acs, biczok}@crysys.hu

***Abstract*— Data generated by cars is growing at an unprecedented scale. As cars gradually become part of the Internet of Things (IoT) ecosystem, several stakeholders discover the value of in-vehicle network logs containing the measurements of the multitude of sensors deployed within the car. This wealth of data is also expected to be exploitable by third parties for the purpose of profiling drivers in order to provide personalized, value-added services. Although several prior works have successfully demonstrated the feasibility of driver re-identification using the in-vehicle network data captured on the vehicle's CAN (Controller Area Network) bus, they inferred the identity of the driver only from known sensor signals (such as the vehicle's speed, brake pedal position, steering wheel angle, etc.) extracted from the CAN messages. However, car manufacturers intentionally do not reveal exact signal location and semantics within CAN logs. We show that the inference of driver identity is possible even with off-the-shelf machine learning techniques without reverse-engineering the CAN protocol. We demonstrate our approach on a dataset of 33 drivers and show that a driver can be re-identified and distinguished from other drivers with an accuracy of 75-85%.**

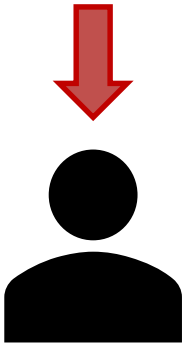
trillion market and could become five times bigger than the market for the cars themselves in the next few years[2].

Unfortunately, sharing in-vehicle network data raises serious privacy concerns. Although drivers are expected to opt-in to such data sharing[3], it is still unclear what exact personal information they would transfer then to third parties. For example, can a skilled data analyst infer the driver's identity using *only* in-vehicle network data? Despite the inherently noisy nature of this fine-grained measurement data, the feasibility of driver identification has been demonstrated in several prior works [2], [3], [4], [5]. In particular, it is well-known that drivers can be re-identified in constrained environments if they follow the same route with the same car and sensor readings are available from the captured network logs [2]. However, when following different routes, unique driving patterns are more difficult to extract due to the variable traffic conditions. Also, it is much more plausible that an adversary is able to collect CAN logs from arbitrary

A mi munkánk

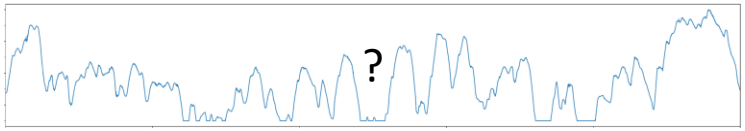
- Egyenesen a nyers adatok alapján végzünk klasszifikációt
- Különböző útvonalakon
- 32 vezetővel

Timestamp	CAN-ID	Req	Len	Data
1481492683.285052	0x0208	000	0x8	0x00 0x00 0x32 0x00 0x0e 0x32 0xfe 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc8 0x00 0x0f 0x03 0x00 0x92 0x3c
1481492674.736055	0x02c4	000	0x8	0x82 0xc9 0x00 0x0f 0x00 0x00 0x92 0x4c
1481492674.736055	0x02c4	000	0x8	0x82 0xcc 0x00 0x0f 0x08 0x00 0x92 0x5a
1497323915.123844	0x018e	000	0x8	0x03 0x03 0x00 0x00 0x00 0x00 0x07 0x3f
1497323915.112910	0x00f1	000	0x6	0x28 0x00 0x00 0x40 0x00 0x00
1481492674.736055	0x02c4	000	0x8	0x82 0xd2 0x00 0x0f 0x0c 0x00 0x92 0x5d
1481492674.736055	0x02c4	000	0x8	0x82 0xa1 0x00 0x0f 0xa1 0x00 0x92 0x4d



A mi munkánk

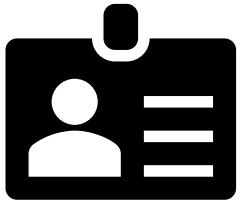
- Nem tudjuk melyik bájt mit jelent
  - 72 ismeretlen idősor
- Megoldás:

72 x 

72 x Deep Learning model

TOP-10 kombinálása

Mixture model



PRIVACY MEGFONTOLÁSOK

Privacy

~~*De hát nincs mit rejtegetnem!*~~

"My problem with [statements like these] is that they accept the premise that privacy is about hiding a wrong. It's not. Privacy is an inherent human right, and a requirement for maintaining the human condition with dignity and respect."

Bruce Schneier, IT biztonság szakértő and kriptográfus

GDPR

- Személyes adat
  - Azonosított
  - Azonosítható személyhez köthető adat
- Kötelező
  - Beleegyezést kérni
  - Informálni a felhasználókat
  - „Privacy by design” elv
- Következmények
  - Akár 20M eurós bírság...

Köszönöm a figyelmet!

Kérdések?