From Model T to Waymo: Cars have gotten better - has transportation data science?

February 17, 2017

Carol Flannagan, Ph.D.
Brief History of Crash Data and Analysis

Detroit traffic in the early days of the automobile (the last transportation revolution)

Brief History of Crash Data and Analysis

Started counting crashes, injuries, fatalities

• In 1917, 75% of fatalities were pedestrians

• First countermeasure—drive 5 mph in cities—resulted in congestion

• James Couzens (of Couzens Hall fame) implemented crosswalks, lane lines, stop signs, and eventually traffic lights

Brief History of Crash Data and Analysis

Crash data have been the source of safety information for 100 years

• In 1975, NHTSA established the Fatality Analysis Reporting System (FARS), a census of fatal crashes in the U.S.

• Since 1979, two national samples have been collected annually:

  1. National sample of police-reported crashes (for general estimates of crashes)

  2. National sample of investigated crashes (focused on occupant protection)
Brief History of Crash Data and Analysis

Top 10 dangerous intersections in Washtenaw County

December 7, 2013 by Steven Gursten

Washtenaw County’s most dangerous intersection: Michigan Avenue at Platt Road

Today I wanted to share this new information on the most dangerous intersections in Washtenaw County for motor vehicle accidents with our clients and friends from the Ann Arbor and Washtenaw County area.

It’s good timing to post this list, as
As a result of having good data on injuries and context, occupant protection (seatbelts, airbags, crush zones, etc.) has gotten really good

- Crash datasets are ideal for understanding outcome *given* a crash

...Which used to be most of what could be improved with technology...
Brief History of Crash Data and Analysis

INSURANCE INSTITUTE FOR HIGHWAY SAFETY

Car–to–car crash test
1959 Chevrolet Bel Air
2009 Chevrolet Malibu
80 mi/h closing speed
50 percent overlap

CF09012
September 9, 2009
Brief History of Crash Data and Analysis

Natural technological progression to crash *avoidance* technologies

Hey, what if we just didn’t get into crashes in the first place?
Brief History of Crash Data and Analysis

BUT, how do you count crashes that DON’T happen?

And wouldn’t it be great to know what happened before a crash?

Solution: Collect data while people are driving rather than only after they crash
Data Collection Strategies

Information (variables) ➔

Cases/Events/Drivers ➔

Instrumented Vehicle Dataset
- Few cases
- Extensive info
- Data stored in the veh, physical download later

Triggered Dataset
- Lots of cases
- Limited, targeted info
- Data stored in module, transferred over air (cell)
WHY NOT THIS?

Information (variables) →

ALL THE DATA

Cases/Events/Drivers
WHY NOT THIS?

- World-wide, vehicles travel more than 1 light-year per year
- Over-the-air data transfer is required for any large-scale collection
  - Airtime is still expensive (in very large quantities)
- Manufacturers may (or may not) have all data for their vehicles, but even then, it is usually stored on the vehicle for later download
  - Physical download requires access to vehicle, limiting possible sample size
- Production cars are not over-built for either bandwidth or storage
  - Cuts into profit margin
Instrumented Vehicle Data

- Naturalistic driving studies and Field Operational Tests (FOTs)
Instrumented Vehicle Data Use Example

Performance Characterization and Safety Effectiveness of Collision Mitigation Braking and Forward Collision Warning Systems for Medium/Heavy Commercial Vehicles

Report Number UMTRI-2011-36

John Woodrooffe
Daniel Blower
Shan Bao
Scott Bogard
Carol Flannagan
Paul E. Green
David LeBlanc

The University of Michigan Transportation Research Institute

June 2012
Instrumented Vehicle Data Example: Estimate safety benefits of truck automatic braking system

- **t0**: Object tracked
- **t1**: Potential rear end collision detected
- **t2**: Collision warning: Visual and Audible
- **t3**: Avoidance maneuver not possible
- **t4**: Hard braking required to prevent collision
- **t5**: Automatic braking for collision prevention or mitigation
- **t6**: Crash prevented or mitigated

**System Reactions**

- **Warning Tone and Lamp**
- **Engine Torque Limitation**
- **Brake Activation**
Baseline Simulation Methodology

• Start with truck driving data (from instrumented trucks)
  • Find events where truck driver brakes because of actions of a lead vehicle
    – These are typically not near-crash or dangerous situations
    – In dataset, drivers responded to the situation appropriately
  • Retain lead-vehicle kinematics and starting speed and range of truck; remove actual truck braking
  • Successively delay simulated braking and play out scenario until vehicles “crash"
Effect of Delay Time on Severity

Initial Conditions:
LV Speed = 40 mph
FV Speed = 38 mph
Range = 55 m

LV Decel Profile

FV Decel Profile

Increasing Delay Time

Increasing Crash Severity

Crash

Driver Delay Time to Braking, 3 s
Driver Delay Time to Braking, 4 s
Driver Delay Time to Braking, 4.2 s
Driver Delay Time to Braking, 4.3 s
Driver Delay Time to Braking, 6.8 s
Effect of Delay Time on Severity

Initial Conditions:
FV Speed = 40 mph
LV Speed = 38 mph
Range = 55 m

LV Decel Profile

FV Decel Profile

Increasing Crash Severity

Increasing Delay Time
Range vs Range-rate With CMB

Initial Conditions:
FV Speed = 40 mph
LV Speed = 38 mph
Range = 55 m

Pov Decel Profile

Acceleration Profile used in CMB1 and CMB2 Simulations

- Measured Ax (from test-track)
- Average Measured Ax
- Ax Profile for Simulation

Increasing Crash Severity

Crash
## Results: crashes avoided

<table>
<thead>
<tr>
<th>Device</th>
<th>Lead-Vehicle Decelerating and Stopped</th>
<th>Lead-Vehicle Fixed</th>
<th>Lead-Vehicle Slower</th>
<th>Cut-In</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCW Alone</td>
<td>15.8%</td>
<td>9.7%</td>
<td>20.3%</td>
<td>27.5%</td>
</tr>
<tr>
<td>CMB1 Alone</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>CMB2 Alone</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>CMB3 Alone</td>
<td>25.4%</td>
<td>52.7%</td>
<td>6.0%</td>
<td>18.6%</td>
</tr>
<tr>
<td>FCW + CMB1</td>
<td>15.8%</td>
<td>9.7%</td>
<td>20.3%</td>
<td>27.6%</td>
</tr>
<tr>
<td>FCW + CMB2</td>
<td>15.8%</td>
<td>9.7%</td>
<td>20.3%</td>
<td>27.9%</td>
</tr>
<tr>
<td>FCW + CMB3</td>
<td>36.6%</td>
<td>58.5%</td>
<td>25.1%</td>
<td>40.0%</td>
</tr>
</tbody>
</table>
Results….injury reductions  
(example results)

Injury Reductions for Avoidance + Mitigation (LV Decel and Stopped):

<table>
<thead>
<tr>
<th>Device</th>
<th>K</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCW ONLY</td>
<td>37.5%</td>
<td>35.5%</td>
<td>31.9%</td>
<td>26.3%</td>
<td>9.0%</td>
</tr>
<tr>
<td>CMB1 ONLY</td>
<td>6.7%</td>
<td>5.8%</td>
<td>4.1%</td>
<td>1.9%</td>
<td>-1.6%</td>
</tr>
<tr>
<td>CMB2 ONLY</td>
<td>11.0%</td>
<td>10.0%</td>
<td>8.0%</td>
<td>5.0%</td>
<td>-3.3%</td>
</tr>
<tr>
<td>CMB3 ONLY</td>
<td>53.9%</td>
<td>52.7%</td>
<td>50.5%</td>
<td>46.1%</td>
<td>14.0%</td>
</tr>
<tr>
<td>FCW + CMB1</td>
<td>38.8%</td>
<td>36.7%</td>
<td>32.9%</td>
<td>26.9%</td>
<td>8.6%</td>
</tr>
<tr>
<td>FCW + CMB2</td>
<td>40.6%</td>
<td>38.5%</td>
<td>34.7%</td>
<td>28.4%</td>
<td>7.9%</td>
</tr>
<tr>
<td>FCW + CMB3</td>
<td>66.4%</td>
<td>65.0%</td>
<td>62.2%</td>
<td>57.1%</td>
<td>25.1%</td>
</tr>
</tbody>
</table>
Triggered Data Collection: Onboard Modules

- Study conducted in collaboration with GM and OnStar
- Funded by NHTSA
- Evaluated driver behavior in conjunction with two warning systems
Triggered Data Collection: Onboard Modules

Study data per ignition cycle:

1. **Trip aggregated statistics**
   E.g., Distance, speeds, ranges, night/day, time, etc.

2. **Alert-triggered data**:
   Kinematics, alerts, settings, system states, lane position, brake status.

Data Collection

- **Pre-Alert** data: 3-6 sec
- **Alert** data: 4 sec
- **Post-Alert** data & key data between Alert & Post-Alert
Triggered Data Collection: Onboard Modules

- Participants opted in; no further contact after opt-in email
- Data collected telematically

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Vehicles</td>
<td>1,958</td>
</tr>
<tr>
<td>Total Trips</td>
<td>2,463,142</td>
</tr>
<tr>
<td>Total Miles of Driving</td>
<td>18,815,458</td>
</tr>
<tr>
<td>Total Hours of Driving</td>
<td>615,054</td>
</tr>
<tr>
<td>Total LDW Alerts</td>
<td>10,058,567</td>
</tr>
<tr>
<td>Total FCA Tailgating Alerts</td>
<td>1,830,501</td>
</tr>
<tr>
<td>Total FCA Imminent Alerts</td>
<td>260,756</td>
</tr>
</tbody>
</table>
### Triggered Data Collection: Onboard Modules

<table>
<thead>
<tr>
<th>Measure</th>
<th>Vehicles always share lane</th>
<th>False</th>
<th>Lateral motion of a vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slowing</td>
<td>Stopped</td>
<td>Other</td>
</tr>
<tr>
<td>Percent of alerts:</td>
<td>19%</td>
<td>0.40%</td>
<td>31%</td>
</tr>
<tr>
<td>How often drivers don't respond:</td>
<td>19%</td>
<td>24%</td>
<td>54%</td>
</tr>
<tr>
<td>Does experience reduce the rate?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Analysis using prior data from UMTRI studies using vehicles equipped with cameras and radars allows us to identify these scenarios from the very sparse OnStar data.

“From empirical data to models”
Triggered Data Collection: Event Data Recorders (EDRs)

- Event Data Recorders are the “black boxes” of cars
- Triggered by a crash (based on acceleration/airbag sensor)
- Initially in race cars
- First in production in late 90’s/early 2000’s
- Contents (minimum) standardized by NHTSA in 2004
Triggered Data Collection: Event Data Recorders (EDRs)

### System Status at Event (Most Recent Frontal/Rear Event, TRG 3)

<table>
<thead>
<tr>
<th></th>
<th>Complete</th>
<th>3</th>
<th>5000 or greater</th>
<th>0</th>
<th>Buckled</th>
<th>Unbuckled</th>
<th>Not Occupied</th>
<th>Rearward</th>
<th>Drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recording Status, Front/Rear Crash Info.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRG Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time From Previous TRG (msec)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time from Pre-Crash to TRG (msec)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckle Switch, Driver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckle Switch, Passenger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupancy Status, Passenger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seat Position, Driver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shift Position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Pre-Crash Data, -5 to 0 seconds (Most Recent Frontal/Rear Event, TRG 3)

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>0 (TRG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Speed (MPH [km/h])</td>
<td>48.5 [78]</td>
<td>48.5 [78]</td>
<td>48.5 [78]</td>
<td>48.5 [78]</td>
<td>46 [74]</td>
<td>46 [74]</td>
</tr>
<tr>
<td>Brake Switch</td>
<td>OFF</td>
<td>OFF</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
</tr>
<tr>
<td>Accelerator Rate</td>
<td>OFF</td>
<td>OFF</td>
<td>Middle</td>
<td>Middle</td>
<td>OFF</td>
<td>OFF</td>
</tr>
<tr>
<td>Engine RPM (RPM)</td>
<td>1,600</td>
<td>1,600</td>
<td>1,600</td>
<td>1,600</td>
<td>1,200</td>
<td>1,200</td>
</tr>
<tr>
<td>Pre-Crash Data Status *</td>
<td>Valid</td>
<td>Valid</td>
<td>Valid</td>
<td>Valid</td>
<td>Valid</td>
<td>Valid</td>
</tr>
</tbody>
</table>

* "Invalid" may be set for M/T vehicle
Triggered Data Collection: Event Data Recorders (EDRs)

Longitudinal Crash Pulse (Most Recent Frontal/Rear Event, TRG 3 - table 1 of 2)

Max Longitudinal Delta-V (MPH [km/h]) | -9.5 [-15.3]

[Graph showing Longitudinal Delta-V over time]
EDR Analysis of Rear-End Crashes

- National crash dataset that includes EDR from some vehicles (N=2297)
- Identified lead and following vehicles in rear-end crashes where EDR was available (N=364); only 6 were from same crash
- Used pre-crash data to look at speed, deceleration, and brake pedal use just prior to a crash

Collaborators: Dave LeBlanc, Dan Blower, Mark Gilbert, Steve Stachowski, Steve Karamihas, Shan Bao
Funder: Toyota Collaborative Safety Research Center
EDR Analysis of Rear-End Crashes

Analysis of lead and following vehicles in rear-end crashes

- 2/3 of lead vehicles in “stopped” cases were moving 1-2 sec prior to crash
- 40% of following vehicles never brake
- Braking generally starts 1-2 sec before event, if at all
- Ave. deceleration 1 sec before crash is 0.35g (for drivers who brake at all)
  - Hard braking on dry pavement can reach 1g
Back to Automation and the Original Question

The Original Question: How can I tell whether self-driving cars are safer than human-driven cars?

Which is equivalent to…

How can I tell if I should be upset when Will Smith takes the wheel?

Which is equivalent to…

What should safety data look like as automated vehicles enter the fleet?
Specific challenges with AVs:

- Crashes are rare events
  - Initial samples will be small
  - Billions of miles needed for traditional hypothesis testing approaches to see significant differences in injury crashes (Kalra & Paddock, 2016)

- AVs are new actors—we have no idea what they will likely get wrong
  - We know a LOT about what people do wrong

- AV algorithms are a moving target (which is good)
  - But our data analysis methods need to track that
Safety Data and Analysis Requirements for the Near Future

Crash data
- Yes, still
- …and for awhile (20 yrs to turn over fleet)

Event-triggered data
- But what data?
Safety Surrogates

• Surrogate measures often stand in for crashes (or enhance sample size of adverse events) in analysis

• Traffic Conflict Theory (Hydén)

Safety Surrogates

Analysis of SHRP2 Data to Understand Normal and Abnormal Driving Behavior in Work Zones: Phase I Final Briefing

November 14, 2016

Carol Flannagan*, Robert Thomson†, Selpi†, Jordanka Kovaceva†, and Andrew Leslie*

*University of Michigan, Ann Arbor, MI, US
†Chalmers University of Technology, Gothenburg, Sweden
Principal Components Analysis (PCA)

Time series data:
Principal Components Analysis (PCA)

Three 2D views (easier to visualize)

Brown is baseline, blue is SCEs; 10 seconds prior to precip event
Probabilistic Topic Modeling (PTM)

- Method developed in context of text analysis
- Has been adapted for driving data
- Strength is handling many categorical variables and identifying recurrent patterns
Probabilistic Topic Modeling (PTM)

How do we turn driving into text???

- Divide each event into shorter (e.g., 2-4 sec) segments
- **Letter** => Variable value
- **Word** => Combination of variable values present in one segment
- **Document** => Event (described by a sequence of words)
- **Topic** => A set of words commonly found together in documents (i.e., a repeated pattern in events)
Probabilistic Topic Modeling (PTM)

• Goal of PTM is to find Topics, or patterns of words that go together within events
• Modeling done separately for baseline and SCEs
Visualization to highlight continuous measures:
- Topics run vertically and contain additional information not shown
- Topics for baseline and SCEs can be completely different
- Order from most common (#1) to least common (#20) of top 20
Baseline topics typically:
higher speed
keeping acceleration/deceleration low (in all topics)
pre-incident maneuvers (dominated by straight constant speed)
maneuvers judgement - safe
driver behavior (speeding)
traffic density - free flow

SCE topics include:
lower speed
higher deceleration/acceleration
pre-incident maneuver - dynamic and more merging
maneuvers judgement - unsafe and illegal maneuvers are slightly more dominant
driver behavior - inexperience driving and unfamiliar with road, aggressive
traffic density - flow with some restrictions
In addition to defining the data, we need to address analysis techniques.

Basic question: Are AVs safer than human drivers? We mean: If I want to take a trip and have choice between an AV and a human-driven vehicle, which should I get into? RISK

We would like to compare rates, but…

- Driving miles for AVs and human-driven vehicles are not apples to apples
- Highest-speed roads have lowest fatality risk *per mile*
- AVs will make different mistakes—how do we look at this holistically?
You can do a lot of things with these data…

- Driver intent modeling
- Detecting driver state
- Injury prediction (triage algorithm)
- Safety benefits estimation

Questions are wide open and we need more people with varied background to bring their perspective and talents…
The Last Slide

BECOME A TRANSPORTATION DATA SCIENTIST!!

UMTRI
Thank you