What Makes It Testable?
Practical approach to quantifying the safety of Self-Driving cars
11/02/2017

Edward Schwalb, Ph.D
Lead Scientist
Machine Learning Group
Safety Statistics
How Good are Human Drivers?

Humans drive >100 million miles between catastrophic events

2015 Records show:
- 35,000 deaths
- 2.5 million injuries
- 6 million crashes

100 million miles:
- 1.1 fatalities
- 77 injuries
- 190 crashes

Despite Human competence, we see a large number of deaths and injuries because we drive lots of miles …

A.V. must be better!

<table>
<thead>
<tr>
<th>Likelihood vs Risk</th>
<th>Negligible</th>
<th>Marginal</th>
<th>Critical</th>
<th>Catastrophic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent</td>
<td>&lt;10^{-3}</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Probable</td>
<td>&lt;10^{-5}</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Remote</td>
<td>&lt;10^{-7}</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Improbable</td>
<td>&lt;10^{-9}</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Cannot use “traditional” metrics: 99.999% “accuracy” is **not** sufficient !!!

100 million miles = 10^{12} frames!

(10^{4} f/m)
Multi-Pronged Testing Approach
Road + Track + Simulation Safety Quotient

Industry Consensus:
Cannot cover all scenarios in track nor road tests.
Simulation is a necessary for Autonomous Vehicle Development.
Risk Management Best Practices
Systematically Characterize and Prioritize Risk

Characterization of Risk + Prioritization of Risk

Safety Critical Systems approach to AV development …
Quantifying Safety
Statistically Significant Testing using Simulation

Safety Case

Scenarios

Tasks

Driver Logs

Skill Level Quantification

Frame-by-frame Safety Quotient

Scenario Analyzer

Scenario Level Aggregation

Safety-Case Aggregation (Report)

Scenario Generator (Risk Model)

Scenario Simulator (Environment)
Extensive FHWA Approach

Context, Events, Tasks, Metrics

15-30 seconds of driving:

**Context:** Weather, Road Type, Traffic Conditions, Dynamic Vehicle Model, Sensors

**Events:** Route starts/ends, cut-ins, accidents, police intervention, vehicle system failure

**Tasks:** Turn left, onramp, exit, turn right, park

**Metrics:** Probability of Crashes, Rollovers, Near-misses
Degree of comfort (e.g., speed bumps)
Duration to completion (e.g., timid AI driver)
Compostable Safety Quotient
Quantifying Safety for A Sequence of Frames

Safety Quotient
\[ \mathcal{N} = \prod_{i=0}^{n} S_i^{wi} \]

Sequence of Frames
- Object Clearance: \( S_1 \)
- Lane Keeping: \( S_2 \)
- Traffic Light: \( S_3 \)
- Stop Sign: \( S_4 \)

Aggregate for Scenario

Aggregate for Safety Case
Beyond Copying Human Driving
From Superhuman Gamers to Superhuman Drivers

**Strategy:** Copying Humans does not improve on their capabilities. Deep Reinforcement Learning *does*!

**AlphaGo Zero:** Tabula-Rasa RL achieved better than Human capabilities.
Anatomy of a testable AI Driver
Decouple Perception from Decision

Human Comprehensible
Situation Fusion
Dashboard

Scenario Database

Perception → Decision

Perception NN
→ AI Driver Dashboard

Decision NN
→ AI Driver Dashboard

Multi-Agent

Perception Neural Network Development Cycle

Decision Neural Network Development Cycle

Reward
Preserve Skills Gained over Billions of Miles
As perception improves, must avoid retraining over billions of miles

Decision Component
Trained using a billion simulated miles

Situation Dashboard

Multi-Modal Input
Evolving Sensors Technologies

LiDAR #1
LiDAR #2

Infrastructure Communications #1
Infrastructure Communications #2

FOV#1
Frame Rate #1

FOV#2
Frame Rate #2
Regression is Very Real
Confirmation bias can mask major risk

Regression for 1%’tile is much more significant than regression for averages.

Reward distributions during regression show clearly more than a single mode of operation.

Max Reward
DNN Loss
Average Reward
5% of Reward
1% of Reward
Variance
How to Achieve and Quantify Operational Safety
Train and Test driving during Malfunctions and Emergencies
Integration of Vehicle Dynamics
What is the best driving decision?

Steering Angle, Velocity, and Torque

Acceleration Lon., Lat., Ver., Roll, Pitch, Yaw

Displacement and Velocity Lon., Lat., Ver., Roll, Pitch, Yaw

Longitudinal Displacement

Lateral Displacement

Yaw Angle
Testable Safe Approach
Quantify Distributions and Summary Statistics for Safety Quotient

Basic Approach
- Define Common Scenarios
- Record Sensor Data
- Create Labels
- Train
- Road Test

Safe Approach
- Generate Scenarios
- Generate Environment
- Train on Safety Quotient
- Summary Statistics
- Validate

Assemble libraries of safety case scenarios and identify specific tasks

Assemble libraries of environments and organize according to tasks

Train models using Safety Quotient test on both naturalistic and synthetic data

Quantify confidence per scenario evolve by improving on tasks

Quantify rate of failure and regression distribution of Safety Quotient for tasks

Compare Human vs AI drivers for Safety Cases Scenario and Task
Backup Slides
Avoiding Collision in Intersection
Safety Case Decomposition Example

Collision with a **moving** vehicle

- Vehicle in the same or adjacent lane
  - Controlled by **Increasing Clearance**

  - Same direction of other vehicle (HWY)
  - Opposite direction of other vehicle (passing)
  - Arbitrary direction, same lane (intersection)

    - Freeway scenarios
    - Passing scenarios
    - Crossing Intersections
    - Turning at Intersections

  - Running a Red light
    - Running a Stop Sign
    - Running a Wrong Direction Sign
    - Failure to Keep Lane
      - Failure to Merge In Lane
        - Failure to break when Vehicle In Front Brakes

  - Failure to identify Sufficient Gap in Traffic
    - Failure to break when Vehicle Enters Your Lane
Safety Quotient Estimation
Expected Error Distribution

Target error distribution

Skewed distribution observed due to poor textures e.g., for poor visibility simulation

Flat wide distribution observed when testing on naturalistic models trained on synthetic

Multi-modal error distribution observed due to combination of overfitting of ensemble components and poor distribution of scenarios
Safety Quotient Estimation: Observed Error Distribution
Safety Quotient Estimation
Observed Error Distribution

Training on Sunny
Testing on Snow
Training and Testing on Sunny Weather

Train on n=640
Sunny

Sunny
Dusk
Snow

Training and Testing on Sunny Weather
Training on Sunny Testing on Dusk
Training on Sunny Testing on Snow