CSC2125: Safety and Certification of Autonomous Vehicles

Instructor: Prof. Marsha Chechik
TA: Ramy Shahin
Careful supervision by: Dr. Rick Salay
http://www.cs.toronto.edu/~chechik/courses19/csc2125
<table>
<thead>
<tr>
<th>Time Period</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-1990</td>
<td>Static analysis of programs, state machine specifications</td>
</tr>
<tr>
<td>2000s</td>
<td>Software model-checking (Yasm, UFO)</td>
</tr>
<tr>
<td>2010s</td>
<td>Runtime analysis of web service interactions</td>
</tr>
<tr>
<td>Now</td>
<td>Safety and compliance of software-based systems (with GM)</td>
</tr>
</tbody>
</table>

- Model-checking (Xcheck) and formal specifications, para-consistent logics
- Reasoning about incomplete and inconsistent systems
- Modeling and reasoning about variability, product lines
- Safety of self-driving cars
Team Behind the Course

Ramy Shahin, Ph.D. student, Toronto, course TA

Dr. Rick Salay, Waterloo/Toronto, expert in safety of ML and in modeling

Prof. Mark Lawford, McMaster University, CERT. Expert in safety and assurance, teaching a similar course in McMaster now.

Prof. Krzysztof Czarnecki, Univ. of Waterloo. Software Engineer turned expert in self-driving, safety, standards. Lots of material borrowed from his course, Winter 2017.
Other Resources

• MIT course on self-driving (with deep learning) https://selfdrivingcars.mit.edu

• A verification seminar in Aarchus (ML meets FM) https://moves.rwth-aachen.de/teaching/ss-18/fvmml/

• A recent library of testing and verification papers for ML https://sdle2018.github.io/SDLE/V1.1/en/Repository.html
Introductions

• Your name, department, area of research, year in the program, are you planning to take the course for credit
• And what you are hoping to get out of the course
The Dream of Self-Driving
Why Self-Driving Cars?

• Quite possibly, the first wide reaching and profound integration of personal robots in society.
  • **Wide reaching**: 1 billion cars on the road.
  • **Profound**: Human gives control of his/her life directly to robot.
  • **Personal**: One-on-one relationship of communication, collaboration, understanding and trust.
Self-Driving Cars  
(aka driverless cars, autonomous cars, robocars)

• Utopian view
  • Save lives (1.3 million die every year in manual driving)
    • 4D’s of human folly: drunk, drugged, distracted, drowsy driving
  • Eliminate car ownership
    • Increase mobility and access
    • Save money
  • Make transportation personalized, efficient, and reliable

• Dystopian view
  • Eliminate jobs in the transportation sector
  • Failure (even if much rarer) may not depend on factors that are human interpretable or under human control
  • Artificial intelligence systems may be biased in ways that do not coincide with social norms or be ethically grounded
  • Security
The 5 levels of driving automation

For on-road vehicles

<table>
<thead>
<tr>
<th>Level</th>
<th>Human driver monitors the road</th>
<th>Automated system monitors the road</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NO AUTOMATION</td>
<td>Steering and acceleration/deceleration</td>
</tr>
<tr>
<td>1</td>
<td>DRIVER ASSISTANCE</td>
<td>Monitoring of driving environment</td>
</tr>
<tr>
<td>2</td>
<td>PARTIAL AUTOMATION</td>
<td>Fallback when automation fails</td>
</tr>
<tr>
<td>3</td>
<td>CONDITIONAL AUTOMATION</td>
<td>Automated system is in control</td>
</tr>
<tr>
<td>4</td>
<td>HIGH AUTOMATION</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>FULL AUTOMATION</td>
<td></td>
</tr>
</tbody>
</table>
Safety

Absence of *unreasonable* risk of mishap

Risk → Severity → Likelihood
Disclaimer: Vehicles are getting safer

• That doesn’t mean we can’t do better!
Self-driving: safety

Uber Self-Driving Car Fatality Reveals the Technology’s Blind Spots

The ride-sharing company has halted its autonomous vehicle testing while it investigates the accident in Arizona
Why is Safety a problem?

- Object identification is very useful
- Can help predict and plan in addition to help partially meet some safety goals
- Pedestrian detection is an example of how ML fails badly with the key safety requirement: “Don’t hit things!”

**AI/ML Version:**

“I don’t know what it is so it’s not there.”

**vs**

**Safety Version:**

“I don’t know what it BUT IT’S THERE!”
If ML Doesn’t Recognize It, It’s Not There
ADS Hazard Sources

Mature best practices

- Mechanical faults
- Electrical faults
- Computer HW faults
- Computer SW faults

ISO 26262

01100
10110
11110

(ISO / PAS 21448)

- Sensor noise & limitations
- Machine learning errors
- Inadequate driving behavior
- DDT fallback failures

SAE J3061

- Cyber attacks
Traffic Lights in Toronto (St. Clair)
Bad Weather
Ethics and Autonomous Driving

http://moralmachine.mit.edu/
Explainability of Decisions
Goal of the Course

ADS (autonomous driving system)

Challenges of Safe ML

What is Safety of ML components?

ML

Safety

Assuring Safety of ML components through testing, verification or synthesis (safe by construction)
Course Plan (3-hour lectures but intend to use less time in the beginning)

• January 14: Introduction to self-driving: architecture, challenges, failures

• January 21: A quick ML primer and background on testing and verification, to collect safety evidence from ML-based components

• January 28: Safety lifecycle of “conventional” automotive software, failures, assurance cases, challenge of self-driving, accidents and analysis

• February 5: Guest lecture by Dr. Rick Salay. Safe development with ML

• February 12: proposal for presentations and early proposal for projects due
  • Can be about an area of ML safety lifecycle
  • Can be about exploration of generation of evidence for ML-based components
    • Via testing
    • Or via verification (many approaches available)
  • Can be about explanation of ML component
  • Instead of a formal lecture, this is an extended “office hour” to discuss these proposals
Course plan (tentative)

• February 19: family day, no class. Plan for rest of the term posted
• February 25 – paper presentations, 4 per class
• March 4 – paper presentations, 4 per class
• March 11 – paper presentations, 4 per class
• March 18 (Marsha away, Ramy and Rick filling in) – paper presentations, 4 per class
• March 25 - paper presentations, 4 per class
• April 1 – no class
• April 22 – projects due
• April 29 – presentation of projects (probably for more than 2 hours, details TBD)
Mark composition

• 20% course participation
  • Including critiques of the other presentations
• 40% paper presentation
• 40% project
Reading List

• Partial reading list posted at

• This area is extremely fast moving so the reading list is, by definition, out of date. Part of your job is to identify interesting papers in this space