PURSS: Towards Perceptual Uncertainty Aware Responsibility Sensitive Safety with ML

Rick Salay,¹ Krzysztof Czarnecki,¹ Ignacio Alvarez,² Maria Soledad Elli,² Sean Sedwards,¹ Jack Weast²

¹Dept. Electrical and Computer Engineering, Univ. of Waterloo
²Intel Corporation, Automated Driving Group
Automated Driving Systems (ADS)

Sensing → Perception → World model → Planning & control → Actuation
Automated Driving Systems (ADS)
Traditional Safety Assurance

ADS Specification

verify

ADS

Sensing

Perception

World model

Planning & control

Actuation
Automated Driving Systems (ADS)
Traditional Safety Assurance

- ADS Specification
- RSS Responsibility Sensitive Safety

ADS
- Sensing → Perception
- World model → Planning & control
- Planning & control → Actuation

verify
Responsible Sensitive Safety (RSS)

Formalizes
“common sense safety”

e.g., Do not hit the car in front
Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[ d_{\text{min}} = v_r \rho + \frac{1}{2} a_{\text{max,accel}} \rho^2 + \left( \frac{v_r + \rho a_{\text{max,accel}}}{2a_{\text{min,brake}}} \right)^2 - \left( \frac{v_f^2}{2a_{\text{max,brake}}} \right) + \]

- Safe actions maintain distance \( d_{\text{min}} \)
- If \( d_{\text{min}} \) is breached, “proper response” is safe action

Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[ d_{\text{min}} = \left[ v_r \rho + \frac{1}{2} a_{\text{max, accel}} \rho^2 + \frac{(v_r + \rho a_{\text{max, accel}})^2}{2 a_{\text{min, brake}}} - \frac{v_f^2}{2 a_{\text{max, brake}}} \right] + \]

Problem: Assumes perfect perception

Misperception -> wrong action
-> safety risk!
Automated Driving Systems (ADS)
Traditional Safety Assurance

ADS Specification

??

RSS
Responsibility
Sensitive Safety

ADS

verify

verify

Sensing

Perception

World model

Planning & control

Actuation
Automated Driving Systems (ADS)  
Traditional Safety Assurance

ADS Specification

Hard to specify!

RSS  
Responsibility  
Sensitive Safety

ADS

Sensing

Perception

World model

Planning & control

Actuation
Automated Driving Systems (ADS)
Traditional Safety Assurance

- ADS Specification
  - Hard to specify!

- ADS
  - Sensing
  - World model
    - Planning & control
      - Why we need ML!
        - RSS Responsibility Sensitive Safety

10
Automated Driving Systems (ADS)

Traditional Safety Assurance

Is there another approach to perceptual safety?

RSS
Responsibility
Sensitive Safety

Planning & control

Actuation
Perceptual Uncertainty

• Uncertainty of perceptual component is cause of misperception
  – many factors*: poor labeling, inadequate dataset coverage, etc.

• ML components can report their own uncertainty!
  – as long as they are calibrated...

PURSS

PURSS = perceptual uncertainty (PU) + RSS

Safety Idea:

Use perceptual uncertainty measure to make RSS rules appropriately cautious and limit safety risk

PURSS formalizes this idea
Precise World Model

Real-world situation

True state (unknowable)

Pedestrian speed = 0
activity = standing

Perception (+ PU)

Pedestrian speed = 0.1
activity = walking

Accuracy

Misperception: precise but inaccurate
Perceptual Uncertainty Handling via Imprecise World Models

Real-world situation

Perception (+PU)

PU -> Imprecise World Model (\(\alpha\))

True state (unknowable)

Accuracy

Probability \(\alpha\) that true world model is in the set

Covers a “credible set” of world models with conf. level \(\alpha\)

- Pedestrian speed = 0, activity = standing
- Pedestrian speed = 0.1, activity = walking
Perceptual Uncertainty Handling via Imprecise World Models

Real-world situation

Perception (+PU)

PU -> Imprecise World Model (\(\alpha\))

Covers a “credible set” of world models with conf. level \(\alpha\)

Safety parameter \(\alpha\) is set to desired level of safety

Probability \(\alpha\) that true world model is in the set

True state (unknowable)

Pedestrian speed = 0  activity = standing

Pedestrian speed = 0.1  activity = walking

Pedestrian speed = 0  activity = standing

Pedestrian speed = 0.1  activity = walking

...
Perceptual Uncertainty Handling via Imprecise World Models

Real-world situation

RSS rules are “lifted” to accept imprecise world models

Result: exercises caution by limiting actions to those safe for any covered world model

Probability $\alpha$ that true world model is in the set

Covers a “credible set” of world models with conf. level $\alpha$
Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[
d_{\text{min}} = \left[ v_r \rho + \frac{1}{2} a_{\text{max, accel}} \rho^2 + \frac{(v_r + \rho a_{\text{max, accel}})^2}{2 a_{\text{min, brake}}} - \frac{v_f^2}{2 a_{\text{max, brake}}} \right] +
\]

Lifting: replace values with credible intervals corresponding to \( \alpha \)
Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[ d_{min} = \left[ v_r \rho + \frac{1}{2} a_{max, accel} \rho^2 + \frac{(v_r + \rho a_{max, accel})^2}{2a_{min, brake}} - \frac{v_f^2}{2a_{max, brake}} \right] + \]

Lifting: replace values with credible intervals corresponding to \( \alpha \)

e.g.,

precise: \( v_f = 30 \, m/s \)

PU: \( \sigma_f^2 = 1 \, m/s \)

lift to imprecise:

\( \alpha = 68\%: v_f = [29,31] \, m/s \)

\( \alpha = 95\%: v_f = [28,32] \, m/s \)
Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[
d_{\text{min}} = \left( v_r \rho + \frac{1}{2} a_{\text{max, accel}} \rho^2 + \frac{(v_r + \rho a_{\text{max, accel}})^2}{2a_{\text{min, brake}}} - \frac{v_f^2}{2a_{\text{max, brake}}} \right) +
\]

Lifting: replace values with credible intervals corresponding to given uncertainty \( \sigma_f^2 \).

- Precise: \( v_f = 30 \text{ m/s} \)
- PU: \( \sigma_f^2 = 1 \text{ m/s} \)
- Lift to imprecise:
  - \( \alpha = 68\%: v_f = [29,31] \text{ m/s} \)
  - \( \alpha = 95\%: v_f = [28,32] \text{ m/s} \)

Given uncertainty \( \sigma_f^2 \), increasing confidence \( \alpha \Rightarrow \) decreasing precision of \( v_f \Rightarrow \) larger \( d_{\text{min}} \) to be more cautious.
Responsible Sensitive Safety (RSS)

Do not hit the car in front

\[
d_{\text{min}} = \left[ v_r \rho + \frac{1}{2} a_{\text{max, accel}} \rho^2 + \frac{(v_r + \rho a_{\text{max, accel}})^2}{2a_{\text{min, brake}}} - \frac{v_f^2}{2a_{\text{max, decel}}} \right]
\]

Lifting: replace values with credible intervals corresponding to increasing confidence \( \alpha \Rightarrow \) decreasing precision of \( v_f \Rightarrow \) larger \( d_{\text{min}} \) to be more cautious

\( v_f = [29,31] \text{ m/s} \)

\( a_{\text{max, decel}} = 95\%: v_f = [28,32] \text{ m/s} \)
Benefits and Costs

• Benefit: Safety parameter $\alpha$ can be increased to get as safe as you want
  – RSS rules become correspondingly more cautious
• Cost: More cautious behaviour may negatively impact progress
• Important future work: negotiating the trade-off
Summary

• RSS provides a spec on planning & control
  – supports traditional safety assurance
• Perception is hard to specify and needs ML
  – different safety approach is needed
• PURSS approach to safety
  – Set desired level of safety \( (\alpha) \)
  – Perceptual uncertainty \( \rightarrow \alpha \) imprecise world models
  – Lift RSS rules to be correspondingly cautious
• Much further work coming!