Neural Net Validation

Classical CS and High School Maths to the Rescue

Kai Engelhardt

Ghost Locomotion
Mountain View, CA, USA and Sydney, AU

Two Worlds

Physical	Perceived
car on road	sampling (IMU[1], GPS, CAN)
road, lane markers	sampling (camera), RMS NN
other cars & objects	sampling (cameras), KFE NN

Problem: How to relate the two worlds?
How to do so measurably and verifiably?

[1] Confused by acronym bingo? ▸ Check the glossary.

Partial Answers

- Emerging mathematical traffic models and definitions of socially acceptable driving behaviour [Shalev-Shwartz et al., 2018] indicate how much we need to know about the physical world to make acceptable driving decisions.
- Classical sampling theory tells us how often and how accurately we have to sample the signals given assumptions, eg about their rates of change.
- Samplers and controllers can be validated (or even formally verified).
- Reliability can be improved with the usual techniques (redundancy and/or ASIL-certified COTS).

Problem: How do we tame NNs, measurably and verifiably?

Verified Realisation

Let I, O, and C be sets. Let

$$
\begin{aligned}
f &: I \longrightarrow O && \text{(ground truth)} \\
n &: I \longrightarrow O \times C && \text{(neural net)} \\
v &: I \longrightarrow O \times C \longrightarrow \mathbb{B} && \text{(verifier)}
\end{aligned}
$$

be functions. We say that v *verifies that* n *realises* f if

$$
n(i) = (o, c) \quad \Rightarrow \quad v(i)(n(i)) \quad \Rightarrow \quad f(i) = o \ ,
$$

for all $i \in I$, $o \in O$, and $c \in C$.

Somewhat similar to the **P** vs **NP** distinction, f is generated by a classical (**P**) algorithm but way too slow, whereas n realising f (sometimes) produces the same outputs plus certificates we can efficiently check with v.

This has also been discovered by Jackson et al. [2021].

Example: Verified Realisation

Let's try a simple RMS.

I	camera image
O	set of lane marker shapes and locations
C	shape and location of the road ahead, shape and location of lane markers, and a grid of non-road and non-lane marker areas to prove that what's suggested as detected is all there is
v	checks C against I and the relevant highway code for the possible shapes of lane marker on the road ahead

Why is verified realisation often unrealistic?

Problem: Outputs of f and n hardly ever agree exactly.

Instead, we aim for an n that produces outputs that are close enough.

Metric Space

Let X be a set. Let $d : X^2 \longrightarrow \mathbb{R}_{\geq 0}$. We call d a *metric* (on X) and (X, d) a *metric space* whenever d satisfies all of:

$$\forall x, y \in X \, (d(x, y) = 0 \Leftrightarrow x = y) \tag{\textbf{id}}$$

$$\forall x, y \in X \, (d(x, y) = d(y, x))$$

$$\forall x, y, z \in X \, (d(x, z) \leq d(x, y) + d(y, z))$$

Without too big a loss, the identity of indiscernibles (**id**) can be weakened to

$$\forall x \in X \, (d(x, x) = 0) \tag{\textbf{id}'}$$

to accommodate irrelevant detail in the input space.

Lipschitz Continuity

Let (X, d_X) and (Y, d_Y) be metric spaces. Let $f : X \longrightarrow Y$. If there exists a $\gamma \in \mathbb{R}_{\geq 0}$ such that

$$\forall x, y \in X \, (\gamma \cdot d_X(x, y) \geq d_Y(f(x), f(y)))$$

then f is *Lipschitz continuous*. The smallest such γ is f's *Lipschitz constant*.

Lipschitz continuous functions map close sources to close targets.

Lemma

Composition ("$;$" as well as "$\|$") preserves Lipschitz continuity.

Example: Lipschitz Continuity

Let's try driving.

I scene descriptions (some canonical rep. of lanes, objects, trajectories)

$O = M \times A$ driving decisions comprising a manœuvre and target values for long. and lat. acceleration

$M = \{\text{keep lane}, \text{change lane left}, \ldots, \text{emergency stop}, \ldots\}$

A e.g. vector of floats

Problem: An $f : I \longrightarrow O$ that computes driving decisions can hardly be meaningfully Lipschitz continuous because M is discrete.

Answer: Change O to distributions over driving decisions.

Verified Approximate Realisation

Let (I, d_I) and (O, d_O) be metric spaces. Let C be a set. Let

$$
\begin{aligned}
&\epsilon > 0 && \\
&f : I \longrightarrow O && \text{(Lipschitz continuous g.t.)} \\
&n : I \longrightarrow O \times C && \text{(certifying NN)} \\
&v : I \longrightarrow O \times C \longrightarrow \mathbb{B} && \text{(verifier)}
\end{aligned}
$$

v *verifies that* n ϵ*-realises* f if

$$n(x) = (y, c) \Rightarrow v(x)(n(x)) \Rightarrow d_O(f(x), y) \leq \epsilon \ ,$$

for all $x \in I$, $y \in O$, and $c \in C$.

Here, a certificate lets us validate that the NN's output is close enough to ground truth.

Example: Verified Approximate Realisation

Let's try driving again.

I scene descriptions with certainty scores for individual elements and their trajectories

$O = M \longrightarrow [0,1] \times A$ manœuvres mapped to their likelihood and target values for long. and lat. acceleration

C for each manœuvre $m \in M$, a justification of its score and the chosen target values

Eg, if $n(i) = (o, c)$ and $o(\text{“change lane left”}) = (0.9, \vec{a})$ then c should indicate one or more objects in i that mandate such a lane change and, moreover, attest to the safety of it (there is a lane on the left and we can move safely into it by following $\vec{a}$).

Lipschitz Continuous NNs?

Suppose v verifies that n ϵ-realises the Lipschitz continuous f and that γ is f's Lipschitz constant. Let $x_1, x_2 \in I$ and let $(y_i, c_i) = n(x_i)$ for $i = 1, 2$.

$$\begin{aligned} d_O(y_1, y_2) &\leq d_O(y_1, f(x_1)) + d_O(f(x_1), f(x_2)) + d_O(f(x_2), y_2) \\ &\leq d_O(y_1, f(x_1)) + \gamma d_I(x_1, x_2) + d_O(f(x_2), y_2) \\ &\leq 2\epsilon + \gamma d_I(x_1, x_2) \end{aligned}$$

Lipschitz Continuous NNs?

Suppose v verifies that n ϵ-realises the Lipschitz continuous f and that γ is f's Lipschitz constant. Let $x_1, x_2 \in I$ and let $(y_i, c_i) = n(x_i)$ for $i = 1, 2$.

$$\begin{aligned} d_O(y_1, y_2) &\leq d_O(y_1, f(x_1)) + d_O(f(x_1), f(x_2)) + d_O(f(x_2), y_2) \\ &\leq d_O(y_1, f(x_1)) + \gamma d_I(x_1, x_2) + d_O(f(x_2), y_2) \\ &\leq 2\epsilon + \gamma d_I(x_1, x_2) \end{aligned}$$

For discrete I, define $\mu = \min(d_I(I^2) \setminus \{0\})$, i.e., the smallest non-zero distance in I. Use that to define $\delta = \frac{2\epsilon}{\mu} + \gamma$.

$$\leq \delta d_I(x_1, x_2)$$, if indeed

we also have that n does not distinguish more inputs than d_i, that is, $d_I(x_1, x_2) = 0 \Rightarrow d_O(y_1, y_2) = 0$. (A non-issue if we stuck the the stricter **(id)**.)

References I

Daniel Jackson, Valerie Richmond, Mike Wang, Jeff Chow, Uriel Guajardo, Soonho Kong, Sergio Campos, Geoffrey Litt, and Nikos Arechiga. Certified control: An architecture for verifiable safety of autonomous vehicles, 2021. URL https://arxiv.org/abs/2104.06178.

Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. On a formal model of safe and scalable self-driving cars, 2018. URL https://arxiv.org/abs/1708.06374.

Glossary

ASIL *automotive safety integrity level (ASIL)*, a risk classification scheme

CAN *controller area network*, vintage robust vehicle bus (Bosch)

COTS *Commercial off-the-shelf*, products that are commercially available and can be bought “as is”

GPS *global positioning system*, a satellite-based radionavigation system

IMU *inertial movement unit*, a motion sensor

KFE *kinetic field estimator*, an NN to detect objects and their trajectories in movies

NN *neural network*

RMS *road marker system*, an NN to detect road markers in images