Real-time Collision Prediction

Sam Premutico, Neel Ramachandran, Chris Lucas

Mentor: Vincent Duong, viaduct.ai

Project Overview

Question: Given a driving sequence of 5 seconds, can we predict whether a driver will crash in 5 to 10 seconds?

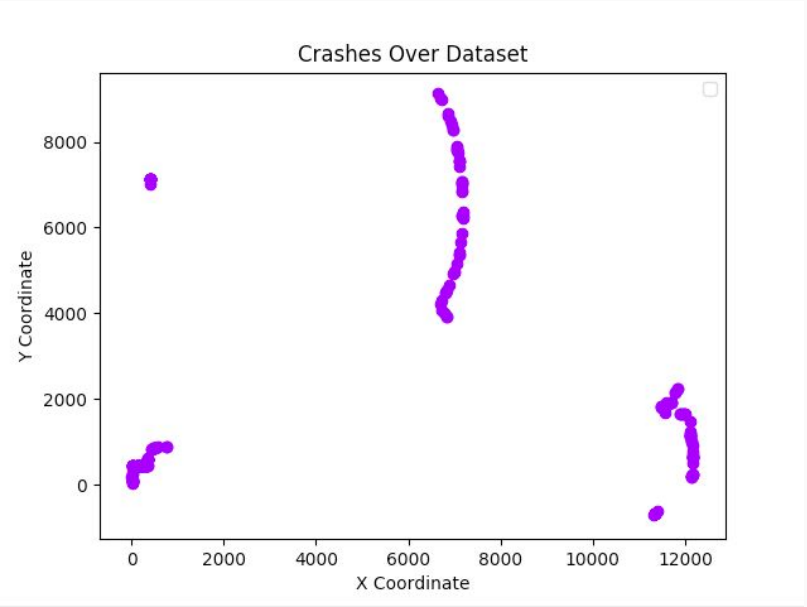
- How much streaming data do we need to store?
- How far ahead in time can we predict risky behavior?

Applications: Insurance companies, triggering in-car warnings, etc.

Collision Segmentation

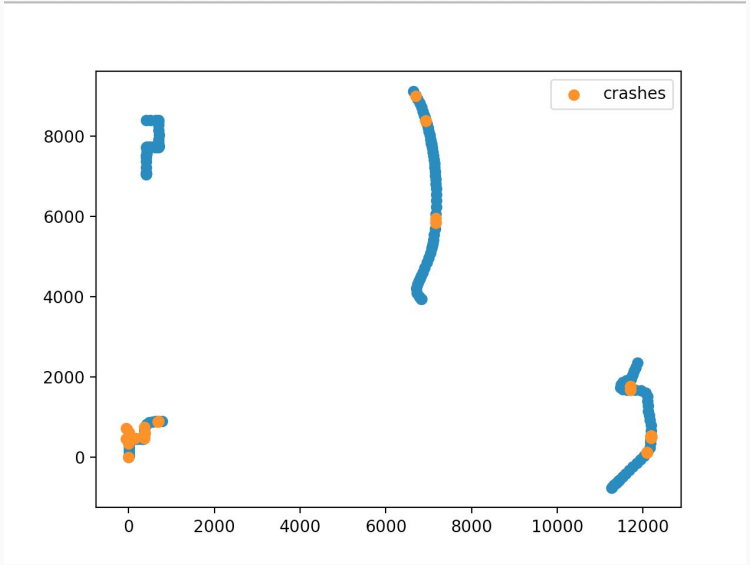
Collisions are ubiquitous

- We segment drives into collision and non-collision sequences
- 9 drivers produce 179 collision segments over course
  - ~20 collisions per driver in < 1 hour: clearly unrealistic
  - Drives are difficult, risk profile is reduced in a simulation



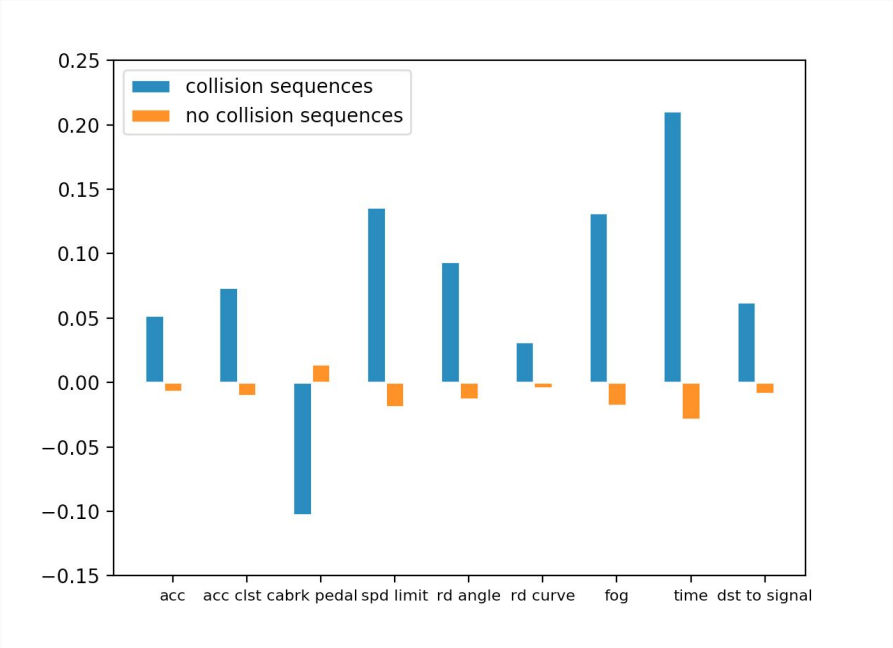
Mining Tiny Data Sets

- 9 users' driving simulations, each < 1 hr
- Segment into collision and no-collision sequences
  - Largest sequence size: 2613
  - Smallest sequence size: 657
- Dimensionality of a single sensor reading: 76
- ..so we stick to simple models



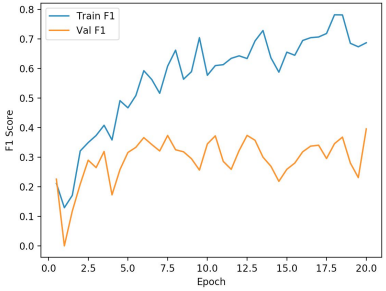
Data Exploration

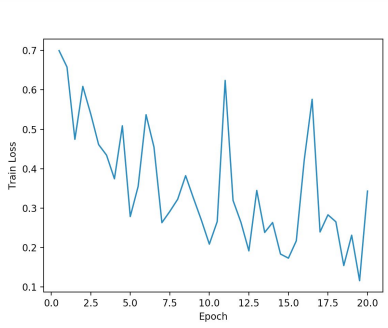
- We see driving factors and environmental factors at play



Methodology

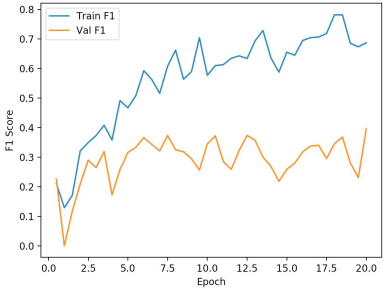
- Baselines: Logistic Regression, RFs
- Deep learning architectures
  - LSTM: 60 dim hidden state
  - 1D-CNN: 2 x (conv → relu → max-pool) → FC
    - Conv layers user 80 & 40 kernels, filter size of 3
    - Max-pool w/ kernel_size = 2
  - Cross-Entropy Loss, ADAM Optimizer
  - Loss weights to address imbalanced classes
- We see noisy loss due to loss weights, overfitting (~1500 total data points)
- Run random search on hyperparameters
  - CNN performs slightly better across the board

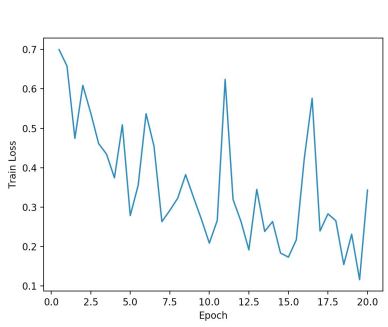




Methodology

- 9 “settings”
  - 3 sequences lengths: 3, 5, 10 seconds
  - 3 window sizes: (1, 6), (5, 10), (15, 20) seconds from present





Results (F1 scores)

Window Size

Sequence Length	(1, 6)	(5, 10)	(15, 20)
3	0.17	0.29	0.39
5	0.24	0.30	0.52
10	0.40	0.49	0.63

Results (F1 scores)

Window Size

Sequence Length	(1, 6)	(5, 10)	(15, 20)
3	0.17	0.14	0.09
5	0.24	0.19	0.11
10	0.40	0.33	0.20

Challenges/Future Work

- Simulation with little/no risk -- how applicable are these results?
  - Assessing our model's accuracy is difficult in this context
- We experimented with training in one scenario and testing in another, however results not meaningful given dataset size
  - Proof of concept: analyze whether a model can predict collisions in one driving scenario giving training data in another
- We'd like to further study how different sequence/window lengths affect predictions in different environments
  - Hypothesis: shorter sequences more predictive in urban environments than rural environments and vice versa

Thank you!