Deep Learning - Predicting Accidents

Elias Gorine and Jacob Smethurst
CS 4624 - Multimedia, Hypertext, and Information Access
Dr. Edward A. Fox
Virginia Tech, Blacksburg, VA, 24061
May 6, 2020

Outline

- Overview
- Data
  - Input
  - Incident
- Preliminary Results and Analysis
- Future Plans
- Lessons Learned
- Acknowledgements
- References

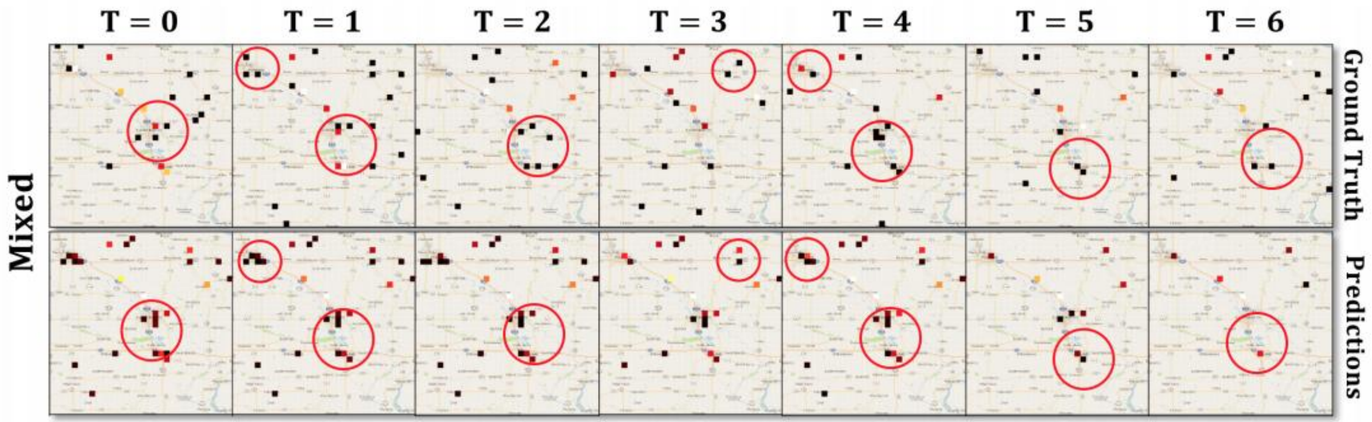
Overview

Help emergency responders respond to accidents faster

Differences From Other Works

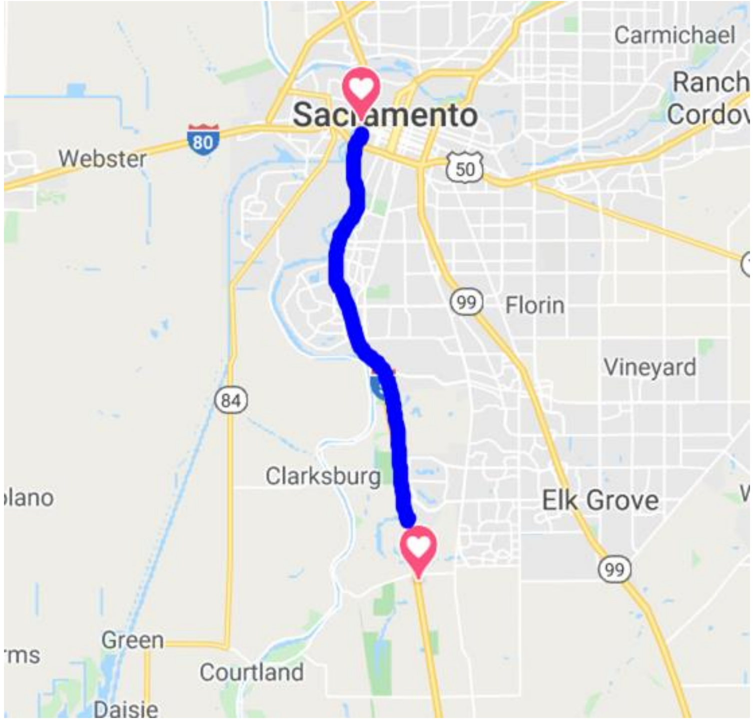
Conv-LSTM paper (Zhuoning Yuan, Xun Zhou, Tianbao Yang) we previously referenced split up Ohio into a grid and predicted the number of accidents to happen per grid square

We want to predict **time and location** of accidents as this fits better for helping emergency responders



Input data: location and time scope

- California I-5 North
- Mile markers 504 through 520
  - 30 collection stations
- July 2018 through November 2018



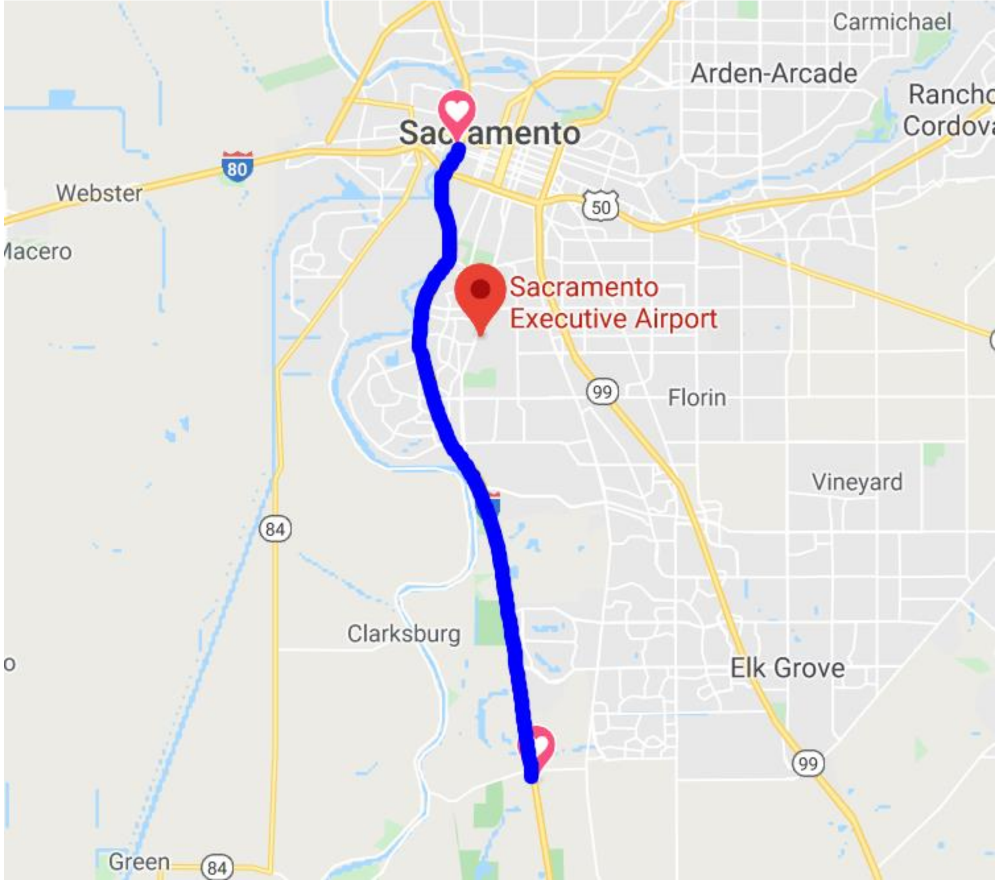
Input data: traffic

- Speed
  - MPH
- Flow
  - # of cars

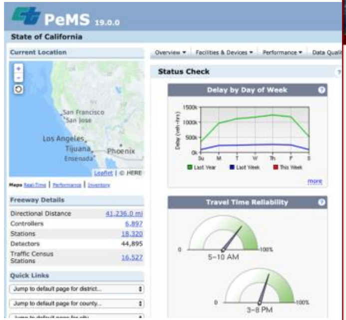
https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQdjvwSOe-1gUgJFxJ8JZ6arzVzr4H7wG6QmW7VXgnbUWmdXR0m&usqp=CAU

Input data: weather

- Temperature
  - Degrees C
- Wind
  - m/s
- Precipitation
  - mm



Input data: collection procedure




```
1710: {
    speed: 66.1,
    flow: 271,
    temp: 20,
    wind: 2.57,
    precip: 0
},
```

SQLite

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	time	s504223	f504223	t504223	w504223	p504223	s504793	f504793	t504793	w504793	p504793	s506383	f506383
2	########	68.2	56	28	3.09	0	67.8	88	28	3.09	0	68.3	6[illegible]
3	########	68.3	40	28	3.6	0	67.2	87	28	3.6	0	66.8	5[illegible]
4	########	68.4	62	28	4.12	0	66.9	91	28	4.12	0	62.5	5[illegible]
5	########	67.8	41	28	4.12	0	64.1	78	28	4.12	0	67.3	4[illegible]
6	########	67.9	48	28	4.63	0	60.7	77	28	4.63	0	61.9	4[illegible]
7	########	67.3	48	28	4.12	0	61.5	74	28	4.12	0	64	3[illegible]
8	########	67.1	30	28	4.12	0	62.5	72	28	4.12	0	66	5[illegible]
9	########	67.8	44	27	3.6	0	63.7	83	27	3.6	0	66.8	5[illegible]
10	########	67.8	48	28	3.6	0	63.4	84	28	3.6	0	66	5[illegible]
11	########	68.3	37	27	3.6	0	64.2	68	27	3.6	0	68.5	4[illegible]
12	########	67.1	36	28	4.12	0	65.4	83	28	4.12	0	69.8	5[illegible]
13	########	68.4	38	27	3.6	0	65.3	68	27	3.6	0	69.9	3[illegible]
14	########	67.4	33	27	3.6	0	65.4	70	27	3.6	0	70.4	3[illegible]
15	########	68	47	27	4.63	0	64.9	81	27	4.63	0	68.8	3[illegible]
16	########	68.4	36	27	4.12	0	63.4	74	27	4.12	0	66.8	5[illegible]
17	########	68	39	27	4.63	0	62.9	71	27	4.63	0	65.4	3[illegible]
18	########	66.4	35	26	4.12	0	62.8	70	26	4.12	0	67.3	4[illegible]
19	########	67.2	32	26	4.12	0	63	67	26	4.12	0	65.7	2[illegible]
20	########	67.5	35	26	3.09	0	60.9	72	26	3.09	0	60.7	3[illegible]

Input data: combined dataset

time	…	s514662	f514662	t514662	w514662	p514662	…
. . .							
07/01/2018 17:10		66.1	271	20	2.57	0	
. . .							

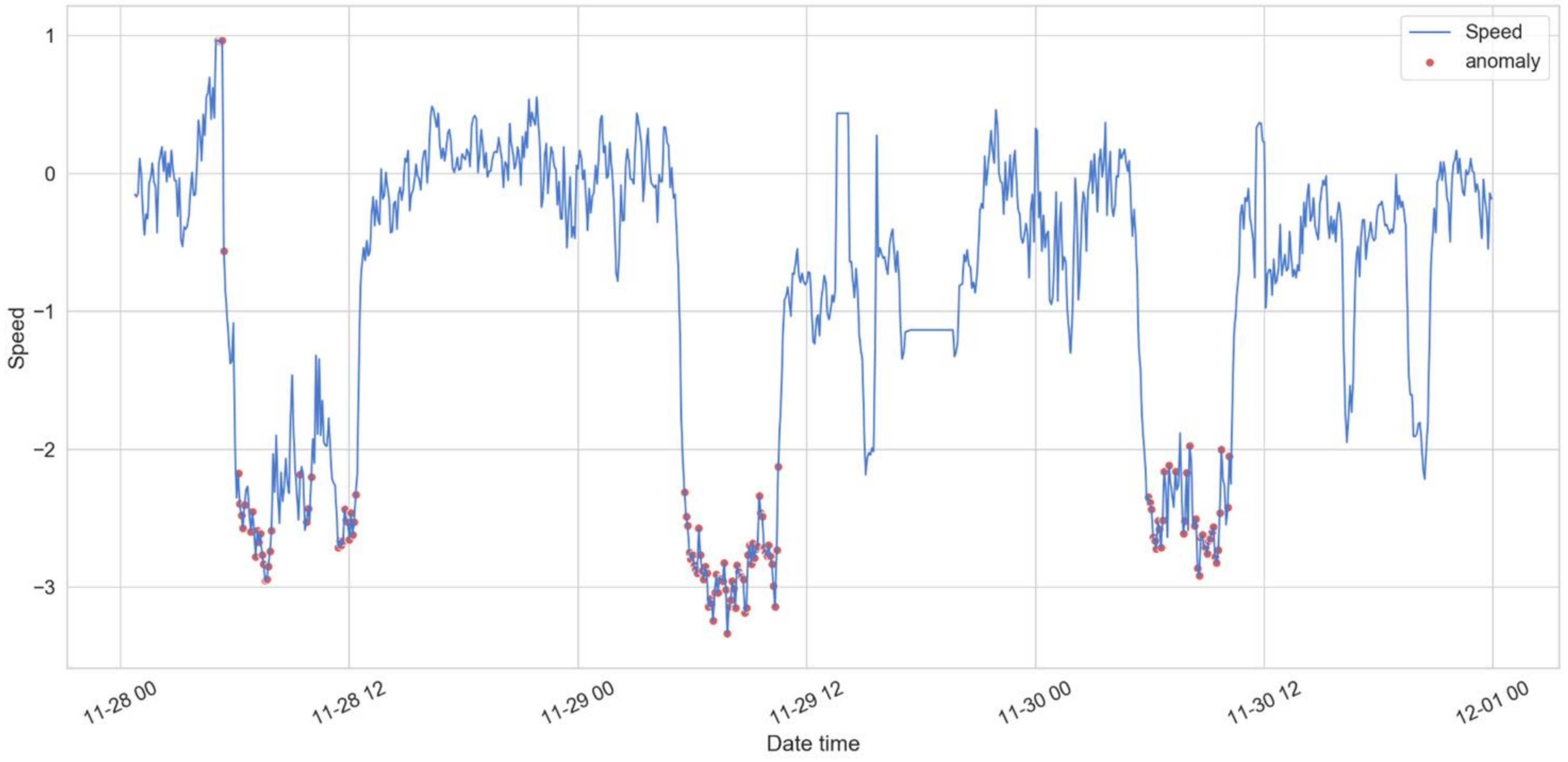
Incident data

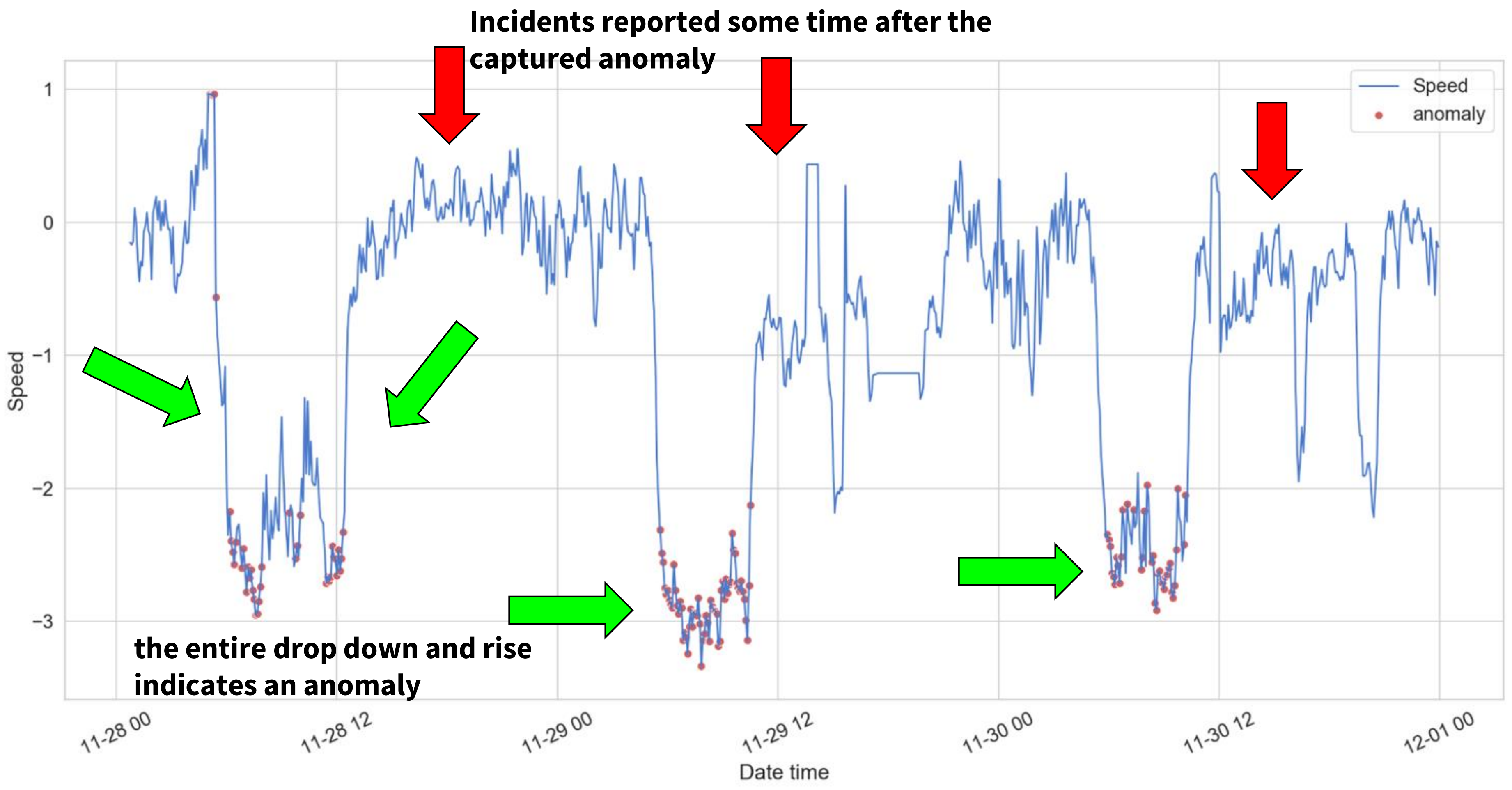
- Time
- Mile-marker location
- Incident type

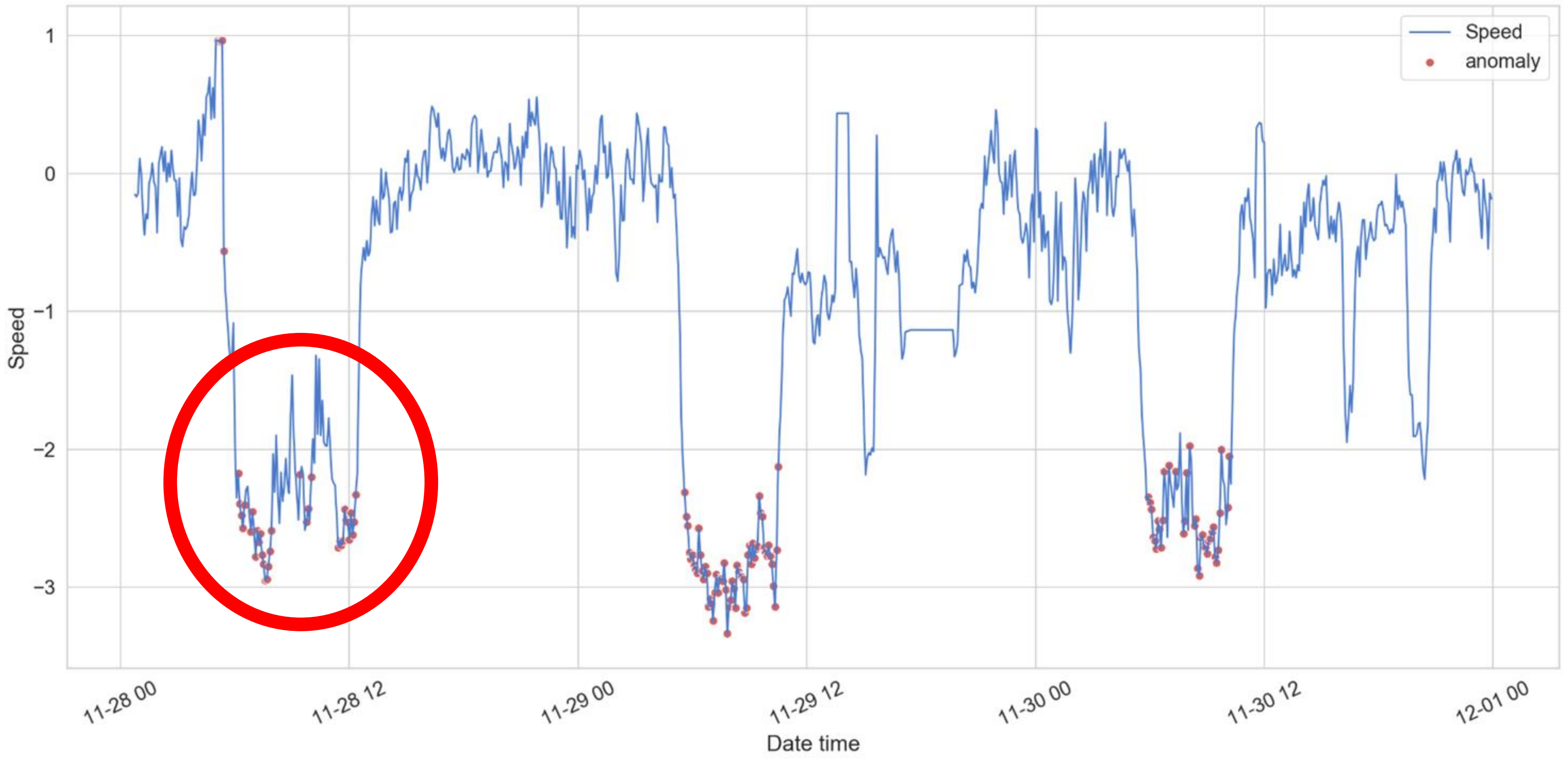
https://bloximages.chicago2.vip.townnews.com/capenews.net/content/tncms/assets/v3/editorial/e/7d/e7dbdbfc-406e-11e5-9a00-27de674c5bef/55ca678e11c12.image.jpg

Autoencoder vs. Bidirectional model

- Moved to bidirectional model because of inconsistencies
- Auto-encoder takes speed of one point as input feature and trains model accordingly
- When more features were added (like downstream speed) auto-encoder didn't work
- Bidirectional model is faster for training compared to the other model as well


Speed
anomaly
1
0
−1
−2
−3
Speed
11-28 00
11-28 12
11-29 00
11-29 12
11-30 00
11-30 12
12-01 00
Date time

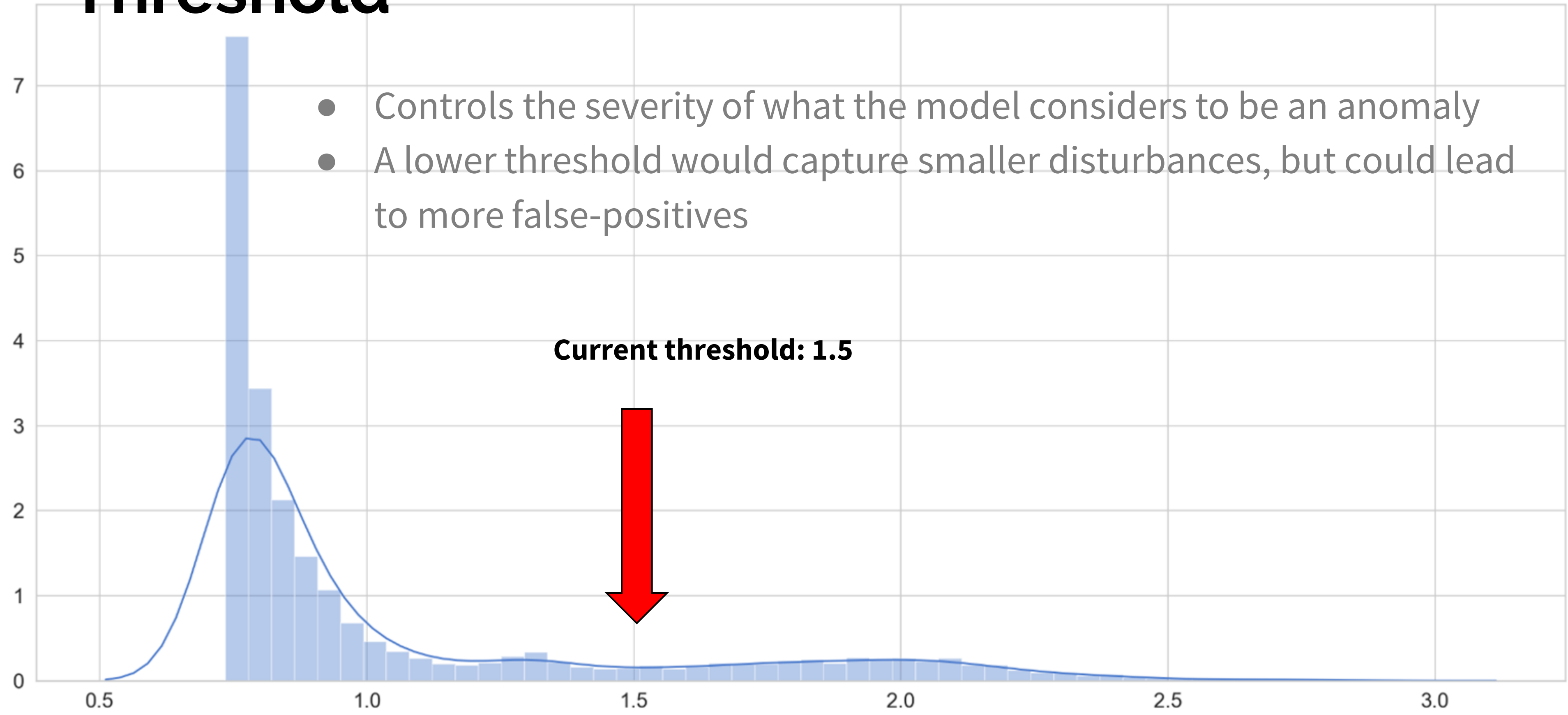

Incidents reported some time after the captured anomaly
the entire drop down and rise indicates an anomaly
Speed
anomaly
Speed
Date time
1
0
−1
−2
−3
11-28 00
11-28 12
11-29 00
11-29 12
11-30 00
11-30 12
12-01 00



Speed
anomaly
1
0
−1
−2
−3
Speed
11-28 00
11-28 12
11-29 00
11-29 12
11-30 00
11-30 12
12-01 00
Date time

Threshold

- Controls the severity of what the model considers to be an anomaly
- A lower threshold would capture smaller disturbances, but could lead to more false-positives



Bidirectional Model Overview

- Performs well in capturing extreme events - hazards not necessarily reported as anomalies
- Setting a smaller threshold allows the model to capture smaller points like road hazards
- Need to aggregate captured points and match them to incidents in the DB

Future Plans

Filter data spatially and temporally so we can build a confusion matrix to identify

true positives, false negatives, etc.

“a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one.” - Wikipedia

		Actual class	
		Cat	Non-cat
Predicted class	Cat	5 True Positives	2 False Positives
	Non-cat	3 False Negatives	3 True Negatives

Things We Learned

Databases

Jupyter

Various Python libraries

Deep Learning

Long-term projects

Working remotely

Database Schema Overview

```
CREATE TABLE incidents_raw (
    incident_id integer,
    time timestamp,
    duration integer,
    freeway text,
    ca_pm text,
    abs_pm real,
    source text,
    area text,
    location text,
    description text
    );
```

5,707 Rows

```
CREATE TABLE speed_raw (
    time timestamp,
    pm_abs real,
    pm_ca text,
    vds integer,
    agg_speed real,
    lane_points integer,
    pct_obs real
    );
```

7,615,378 Rows

```
CREATE TABLE flow_raw (
    time timestamp,
    pm_abs real,
    pm_ca text,
    vds integer,
    agg_flow real,
    lane_points integer,
    pct_obs real
    );
```

7,593,230 Rows

```
CREATE INDEX postmiles_index ON
flow_raw(pm_abs);
CREATE INDEX postmiles_index_speed ON
speed_raw(pm_abs);
CREATE INDEX postmiles_index_incidents ON
incidents_raw(abs_pm);
```

Note: Using roughly two months of data on I5-N (California)

Acknowledgements

Farnaz Khaghani - Client

Dr. Edward A. Fox

https://media-exp1.licdn.com/dms/image/C4E03AQHvTldzxkRm2Q/profile-displayphoto-shrink_200_200/0?e=1590019200&v=beta&t=goWKseC89LpVBvhg6ByFsGE-T2n4Id-OedNOHIjbo0Q

References

Zhuoning Yuan, Xun Zhou, Tianbao Yang. 2018. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data. In KDD ’18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom. ACM, New York, NY, USA, 9 pages. Retrieved April 29, 2020, from https://doi.org/10.1145/3219819.3219922

Brownlee, J. (2020, January 7). How to Develop a Bidirectional LSTM For Sequence Classification in Python with Keras. Retrieved April 29, 2020, from https://machinelearningmastery.com/develop-bidirectional-lstm-sequence-classification-python-keras/