Learning an Optimal Policy for Police Resource Allocation on Freeways

Brian Jackson, Taylor Howell, Olaoluwa Shorinwa {bjack205, thowell, shorinwa}@stanford.edu

Motivation

Each year over 32,000 driving-related deaths and 2 million injuries occur on US roads[1]. We posit that automatic traffic surveillance and efficient allocation of law enforcement resources to apprehend anomalous drivers can improve road safety.

Objectives

1. Identify anomalous drivers from a real-world dataset
2. Optimally allocate police to apprehend anomalous drivers

Dataset

We used the NGSIM dataset for US Highway 101. This data set captures about 6000 vehicle trajectories over 45 minutes of congested morning traffic. After filtering, we extracted six features averaged over 15-second intervals, resulting in 64,000 samples of driver behavior.

Features

Each driver is defined by six features: *number of lane changes, average velocity, maximum velocity, average acceleration, average deviation from lane centers, standard deviation from lane centers.* Principal Component Analysis (PCA) was used to map theses features into new feature spaces of size 2, 3, and 4.

Methodology

Anomaly Detection – The driving trajectories were modeled using a Mixture of Gaussians $p(x,z;\theta)$ with the model parameters obtained using the EM algorithm. The feature space of each driver was compressed using PCA.

EM:
- E-step: $Q_i(z^{(i)}) = p(z^{(i)}|x^{(i)};\theta)$
- M-step: $\theta = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) log \frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$

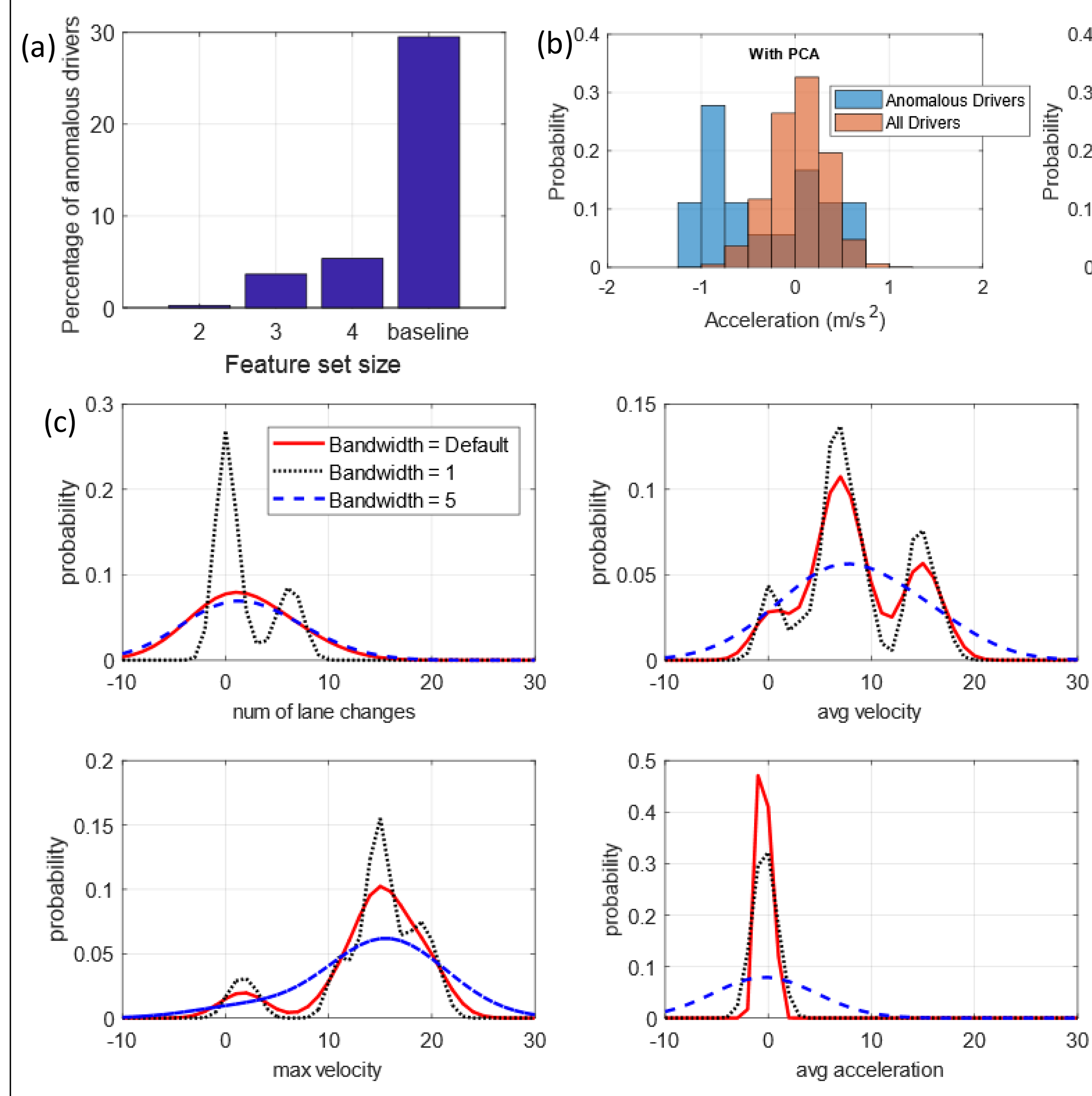
PCA: $u = \underset{u:u^T u=1}{\operatorname{argmax}} \frac{1}{m} \sum_{i=1}^{m} (x^T u)^2$

Police Allocation – The task was formulated as a discounted infinite-horizon Markov Decision Process (MDP) and solved using model-based Reinforcement Learning (RL).

$$U^*(s) = \underset{a}{\operatorname{argmax}} R(s,a) + \gamma \sum_{s'} T(s'|s,a) U^*(s')$$

Results

Anomaly Detection – Anomalous drivers were identified using the PCA models with the threshold $p(x) < 0.0025$. The performance of the PCA models were evaluated against the baseline model.

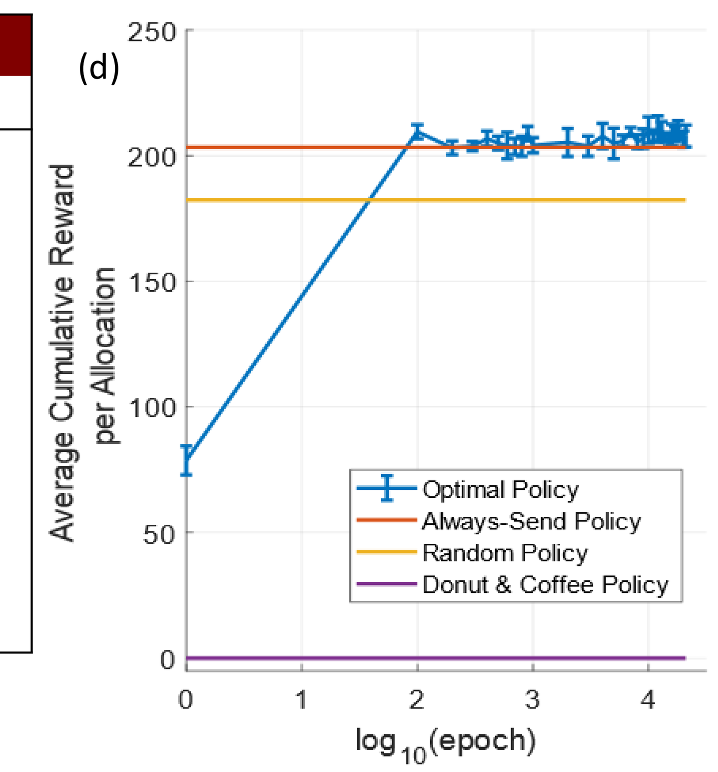


(a) Percentage of anomalous drivers after compressing to different feature spaces with PCA. ***(b)*** Histogram showing effect of PCA (two-feature space) on identification of anomalous drivers. ***(c)*** Kernel distribution fitting for 4 features. The bandwidth of 1 was used to model the distributions for the MDP simulator. ***(d)*** MDP Learning curve

Optimal Policy –The Bellman equation was solved using asynchronous value iteration. The optimal policy (ϵ-greedy w/ ϵ =0.1) learned to only allocated police for the two highest reward drivers. The optimal policy was compared with an always-send (A) and a randomly-send (R) policy.

State			Value	Policy		
Driver	Driver	Police	U*	Opt.	A	R
1	1	0	333.2	0	1	0
2	1	0	341.5	0	1	0
3	1	0	372.8	1	1	1
4	1	0	454.7	1	1	1
2	2	0	341.1	0	1	0
3	2	0	377.1	1	1	1
4	2	0	475.9	1	1	0
3	3	0	378.0	1	1	1
4	3	0	478.4	1	1	0
4	4	0	496.2	1	1	1

Table 1: Comparison of the Optimal (Opt.) policy to a "greedy" always-send policy (A) and a randomly-send policy (R).



MDP Formulation

We consider a case with 1 police and 2 anomalous drivers. At each timestep, the police can be allocated if available and can be allocated again after 5 timesteps. Each driver is identified as: no citation, speeder, weaver, or speeder and weaver. The rewards for citing a driver identity are: 50, 100, 150, 250; if a police is allocated but not available, the reward is -100.

State Space – $|S| = 96$

s = [P D] where: $P \in \{0,..,5\}$ and $D = [d_1\ d_2]$, $d_i \in \{1,..,4\}$

Action Space – $|A| = 2$

$a \in \{0,1\}$

Rewards

$R(d_i) = [50, 100, 150, 250]^{\mathrm{T}}, R(P \neq 0, a = 1) = -100.$

The transition probabilities $T(s'|s,a)$ and rewards $R(s,a)$ are modeled using maximum likelihood estimates. Discount factor $\gamma = 0.9$ and the MDP model is updated every 100 timesteps.

Discussion

The application of PCA reduced the impact of noisy observations, improving the accuracy of the anomaly detection model. This finding was expected as PCA retains valuable variations among the training data while discarding less important noisy variations. We were also able to obtain an optimal policy for police allocation which performed better than both the always-send and randomly-send policies. As expected, the agent learned not to allocate police vehicles in states with lower citations, reserving those police to apprehend drivers with higher citations.

Future Steps

The driver features are not independent as assumed in our formulation. We will train a generative adversarial network to generate new drivers for the simulator. We expect the optimal policy behavior (wait to send police for higher rewards) to scale as the number of police and drivers is increased and we plan to explore a state space with more police and drivers.

References

1. SAUBER-SCHATZ, E. K., EDERER, D. J., DELLINGER, A. M. AND BALDWIN, G. T. Vital Signs: Motor Vehicle Injury Prevention — United States and 19 Comparison Countries
2. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR),* vol 41. no 3m p. 15, 2009.
3. N. Adler, A. S. Hakkert, J. Kornbluth, and M. Sher, "Location-allocation models for traffic police patrol vehicles on interurban network," *Springer,* pp 9-31, 2014
4. B.T. Morris and M. M. Trivendi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE transactions on circuits and systems for video technology,* vol. 18, no. 8, pp. 1114-1272, 2008.