Integrating Learning with Game Theory for Societal Challenges

Fei Fang
Carnegie Mellon University
feifang@cmu.edu

Security Challenges

Lack of Security Resources

Sustainability Challenges

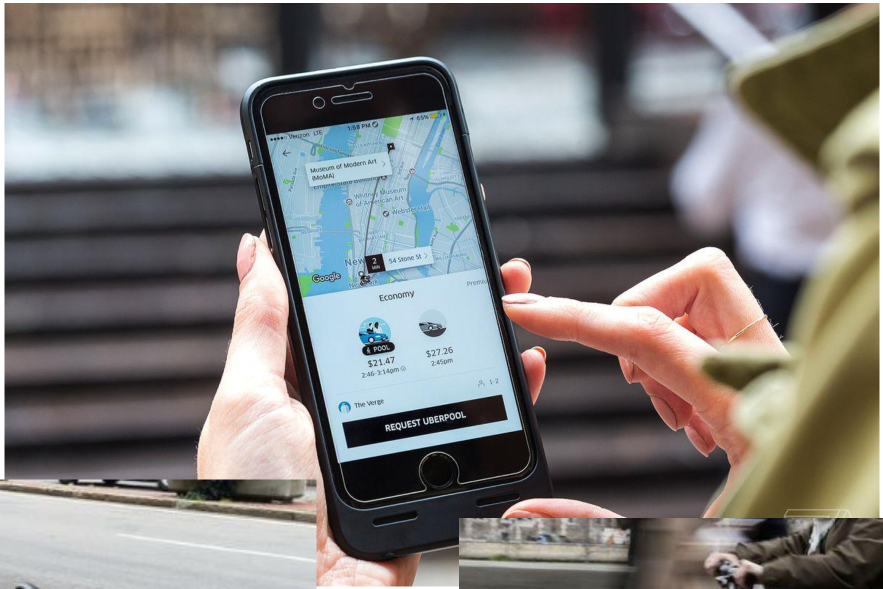
Lack of Ranger/Conservation Resources

100 years ago ≈ 60,000 tigers

Today ≈ 3,200 tigers

Mobility Challenges

Inefficiencies in Matching/Dispatching



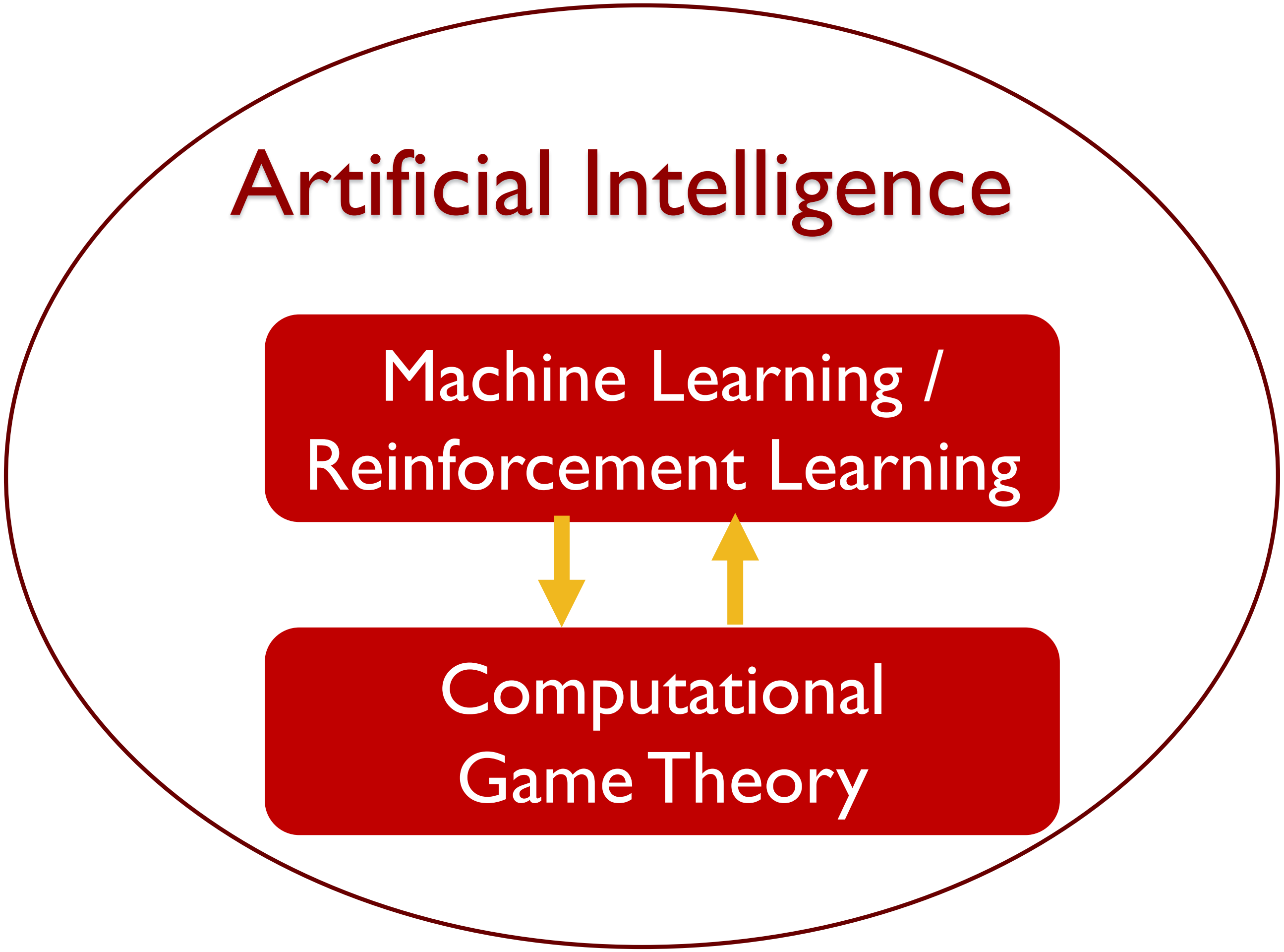
Societal Challenges

Security & Safety

Environmental Sustainability

Mobility

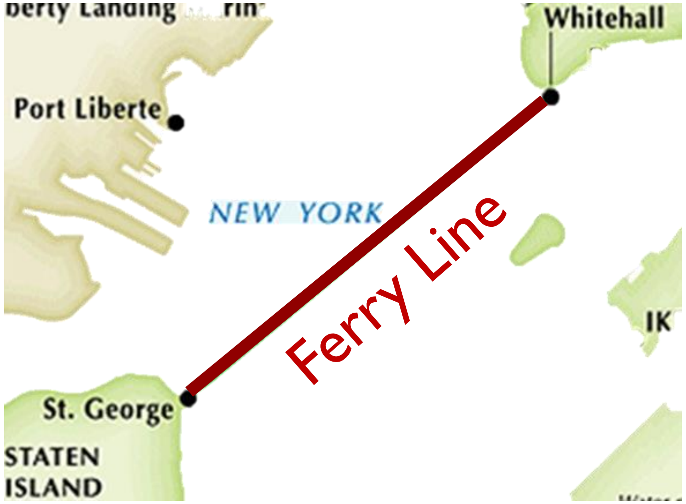
Integrate Learning with Game Theory



Outline

- Security games
- Integrating learning and game theory
  - Data-based game theoretic reasoning
  - Learning-powered equilibrium computation
  - End-to-end learning in games
- Summary

Protecting Staten Island Ferry



Protecting Staten Island Ferry

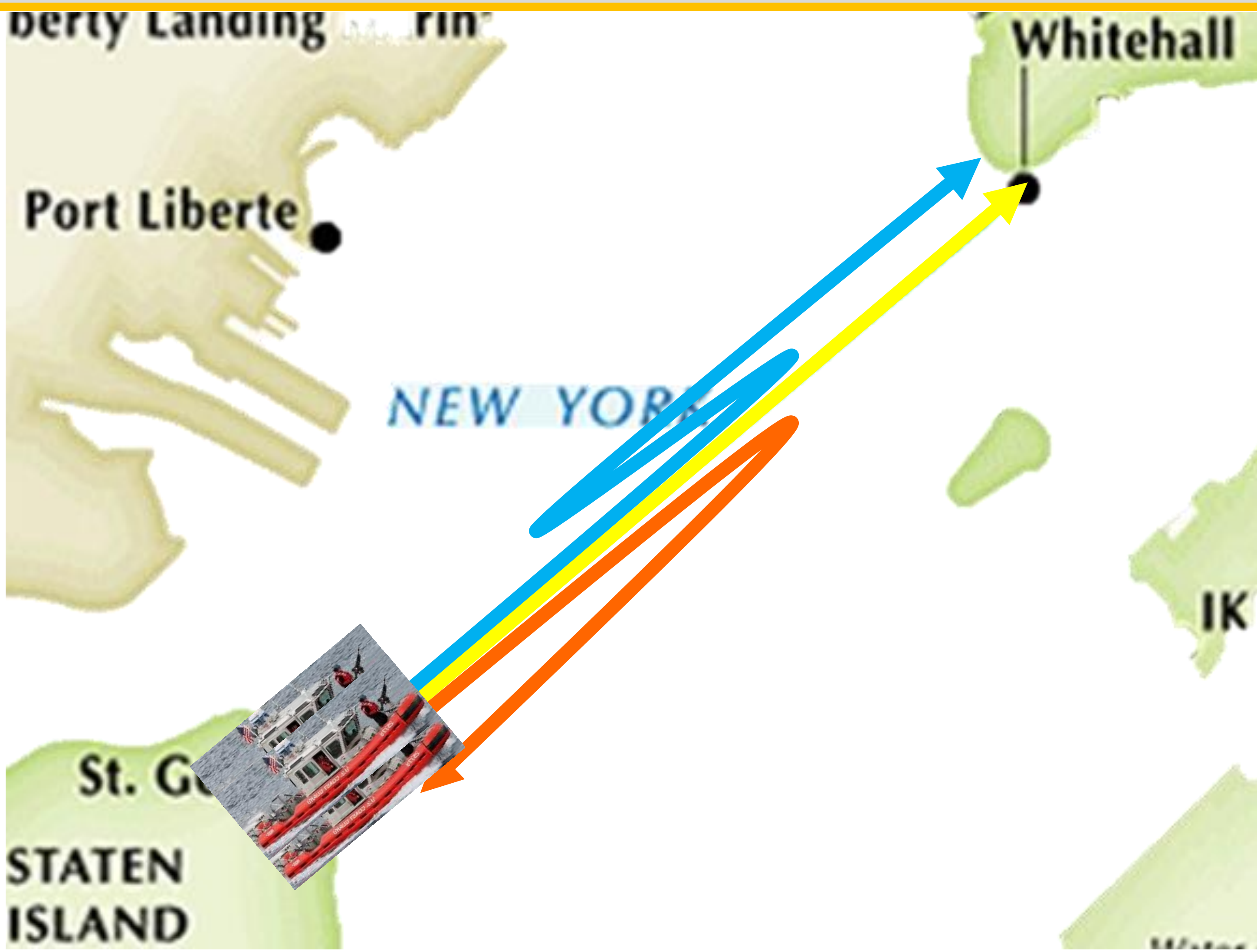
Protecting Staten Island Ferry

Previous USCG Approach

Protecting Staten Island Ferry

In collaboration with US Coast Guard (Contact: Joe Direnzo, Craig Baldwin)

Problem



Game Model and Linear Programming-based Solution

- Stackelberg game: Leader – Defender, Follower – Attacker
- Attacker's payoff: $u_i(t)$ if not protected, 0 otherwise
- Zero-sum → Strong Stackelberg Equilibrium=Nash Equilibrium =Minimax (Minimize Attacker's Maximum Expected Utility)

$$\min_{p_r, v} v$$

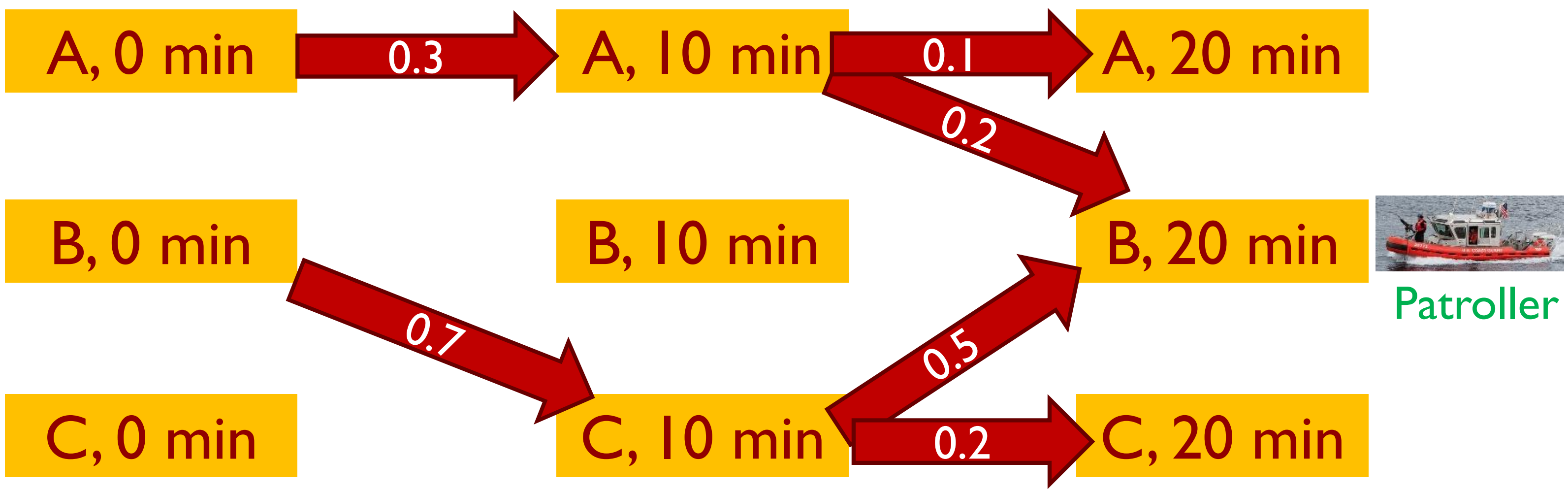
$$\text{s.t. } v \geq \mathbb{E}[U^{att}(i,t)] \; = u_i(t) \times \mathbb{P}[unprotected(i,t)], \forall i, t$$

Adversary

Defender p_r		10:00:00 AM Target 1	10:00:01 AM Target 1	…	10:30:00 AM Target 3	…
30%	Purple Route	-5, 5	-4, 4		0, 0	
40%	Orange Route					
20%	Blue Route					
	……					

Advanced Solution

- Flow-based Representation + Critical Time Points



$$\min_{\boxed{f}} v$$

Prob. flow over feasible edges

s.t. $\sum_{e\in(i,t)\to} f(e) = \sum_{e\in\to(i,t)} f(e)$

$$\sum_{e\in(*,t)\to} f(e) = 1$$

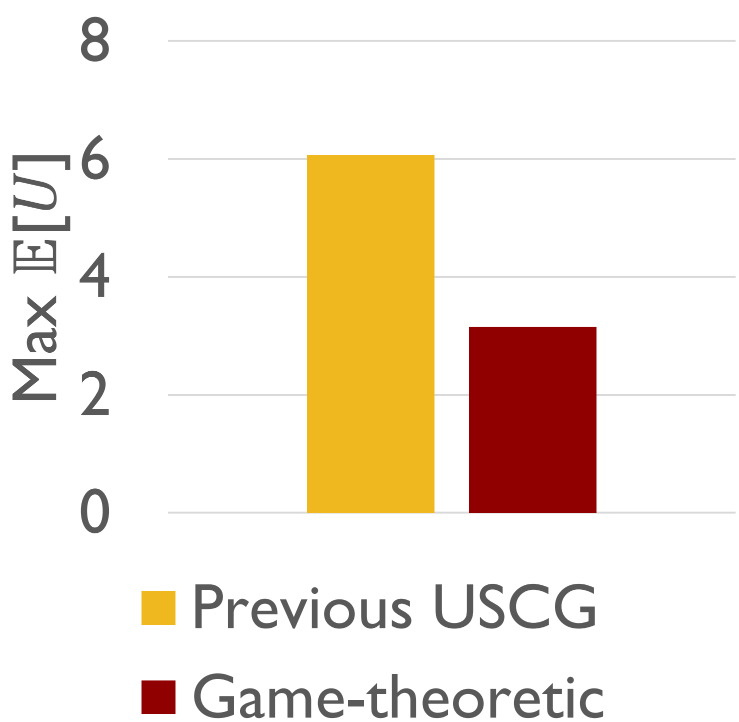
f is a unit flow

$$v \geq \mathbb{E}[U^{att}(i,t)], \forall i, \forall t \in \{t^*\}$$

Best response

Evaluation: Simulation & Real-World Feedback

Reduce potential risk by 50%



- Deployed by US Coast Guard
- USCG evaluation
  - Point defense to zone defense
  - Increased randomness
- Professional mariners:
  - Apparent increase in Coast Guard patrols

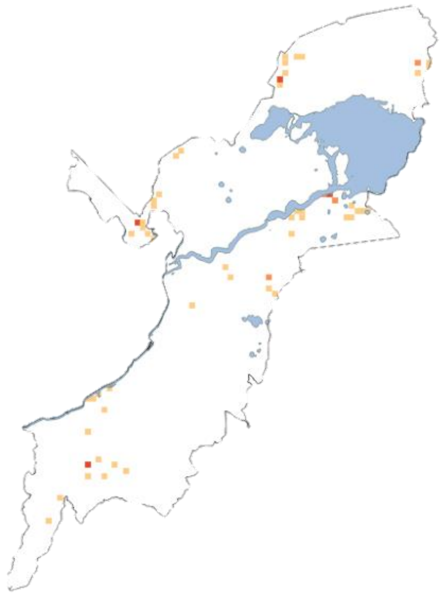
Outline

- Security games
- Integrating learning and game theory
  - Data-based game theoretic reasoning
  - Learning-powered equilibrium computation
  - End-to-end learning in games
- Summary

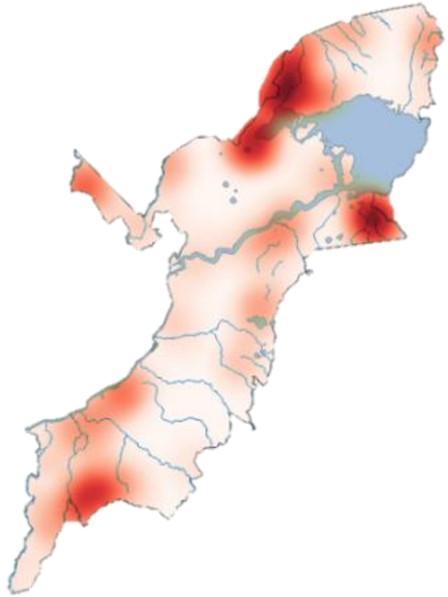
Repeated and Frequent Poaching Activities

Learn Poacher Behavior Model from Real-World Data

12yr of poaching data from Queen Elizabeth National Park in Uganda

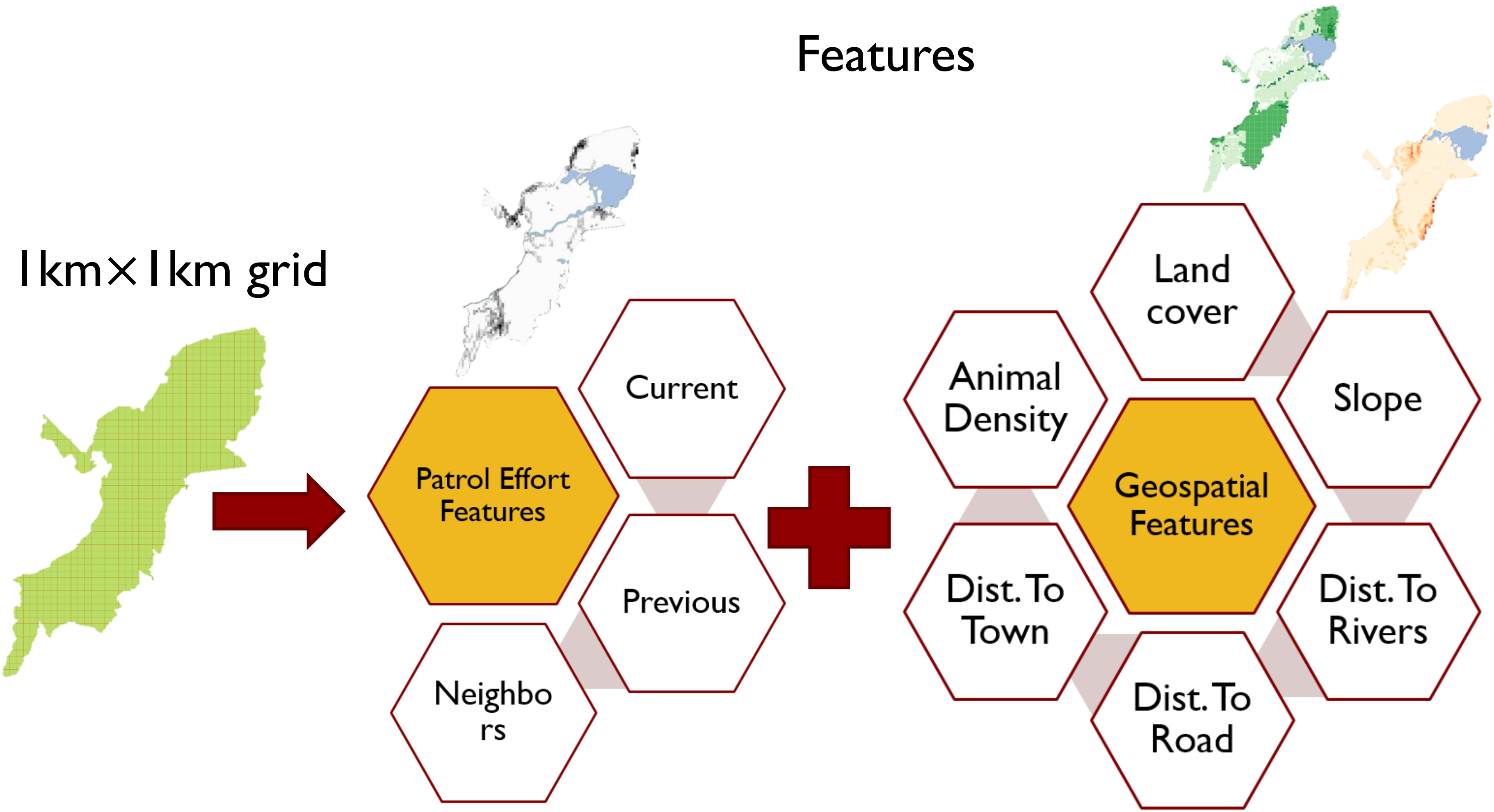


Prob. of snaring or Prob. of detection



With Uganda Wildlife Authority and Wildlife Conservation Society

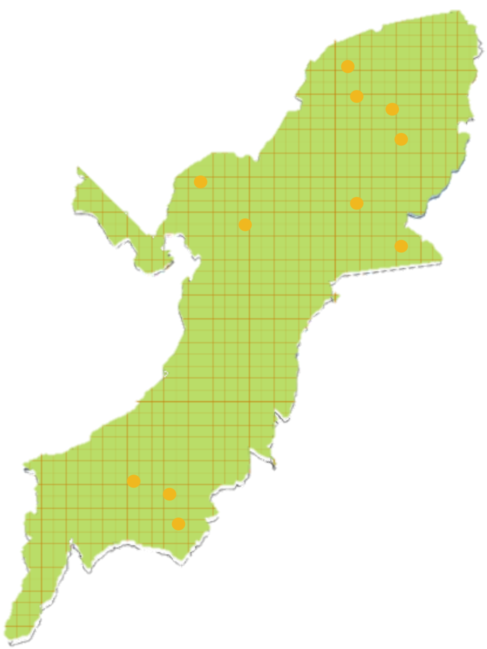
Learn Poacher Behavior Model from Real-World Data



Collaborators: Uganda Wildlife Authority and Wildlife Conservation Society

Learn Poacher Behavior Model from Real-World Data

Labels

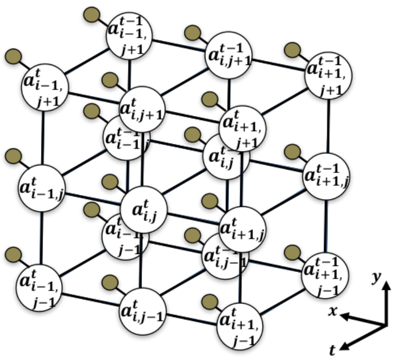


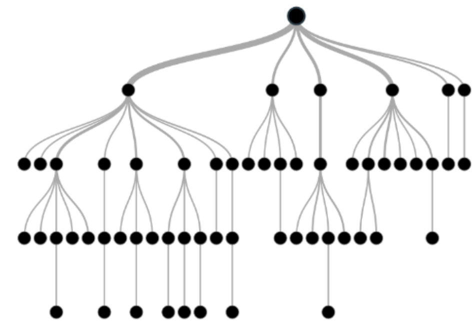
+: Has Poaching

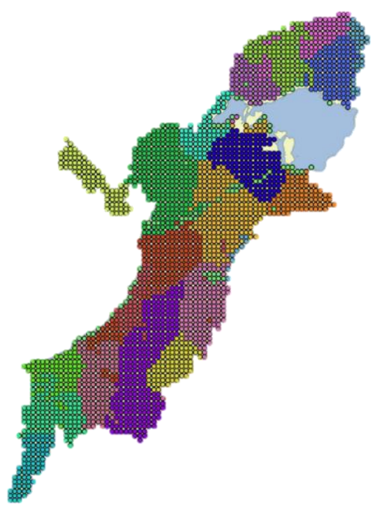
−: No Poaching

Taking it for a Test Drive: A Hybrid Spatio-temporal Model for Wildlife Poaching Prediction Evaluated through a Controlled Field Test. Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga. In ECML-PKDD 2017

Learn Poacher Behavior Model from Real-World Data

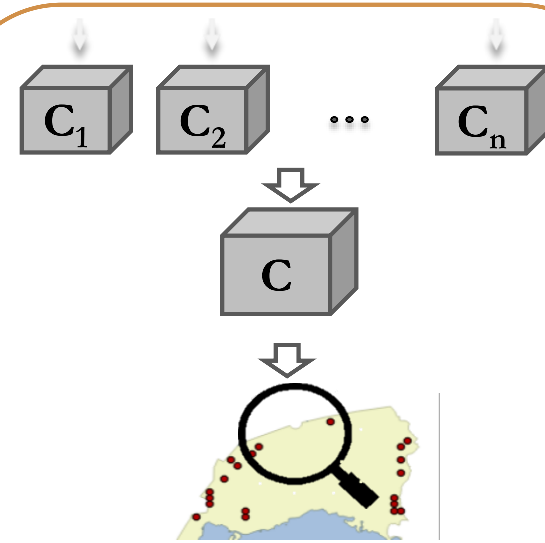






Markov Random Fields

+



Bagging of Decision Trees

Taking it for a Test Drive: A Hybrid Spatio-temporal Model for Wildlife Poaching Prediction Evaluated through a Controlled Field Test. Shahrzad Gholami, Benjamin Ford, Fei Fang, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga. In ECML-PKDD 2017

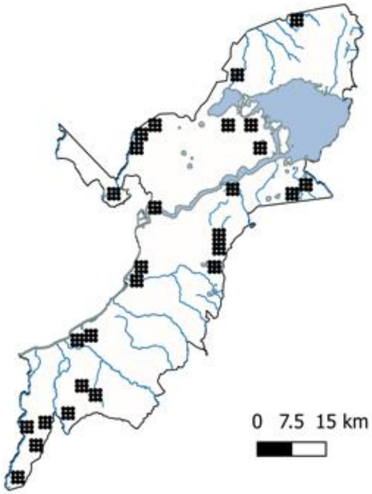
1 Month Field Test

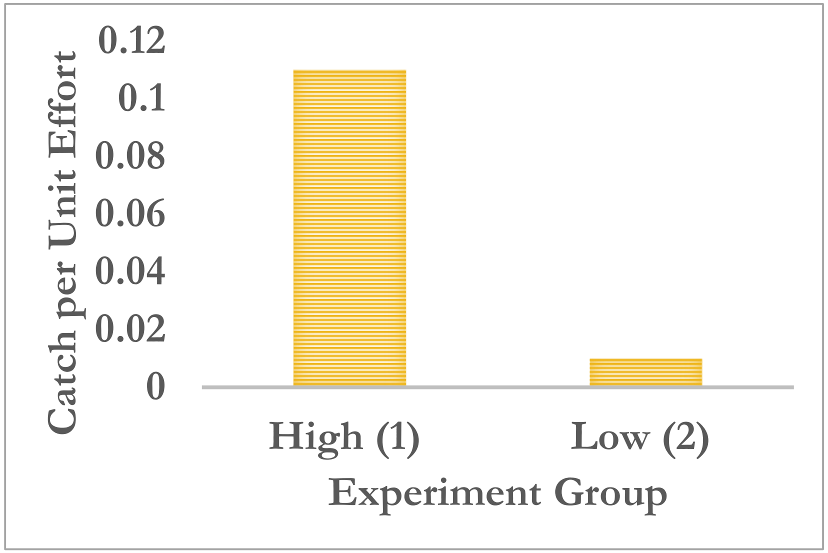
- Two 9-sq. km areas
  - Infrequent patrols
  - Predicted hotspot
- Findings
  - 19 litter, ashes, etc.
  - 1 poached elephant
  - 1 active snare
  - 10 antelope snares
  - 1 roll of elephant snares
- Snaring hit rates
  - Outperform 91% of historical months

Historical Base Hit Rate	Our Hit Rate
Average: 0.73	3

8 Month Field Test

- 27 areas, 9-sq km each
- 2 experiment groups
  - High: 5 areas
  - Low: 22 areas
- 452 km patrolled in total
- Catch Per Unit Effort (CPUE)
  - Unit Effort = km walked
  - Historical CPUE: 0.03
- Can differentiate H/L threat areas





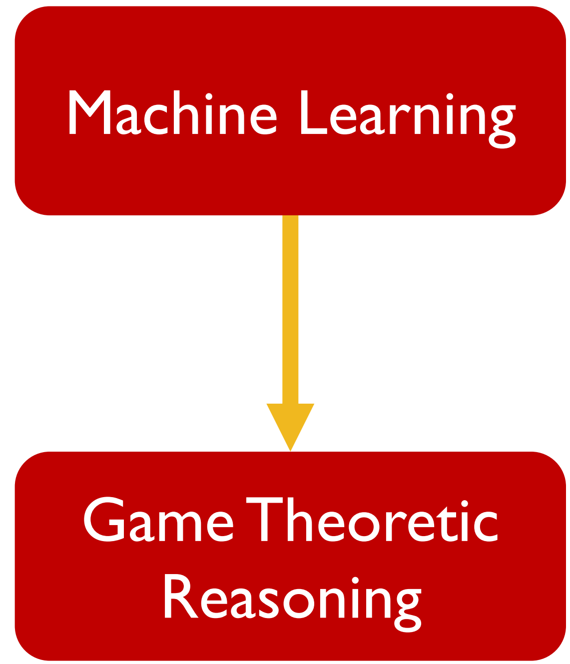
Further Enhancement

- Augment dataset based on experts' knowledge
- Field Test in China
  - Two-day field test in October 2017: 22 snares
  - 34 patrols from November 2017 to February 2018: 7 snares

	Cluster Value in "predict_heatmap_i10.tif"	Estimated Snaring Threat (0~10)
1	1	7
2	2	3
3	3	8
4	4	7
5	5	3
6	6	2
7	7	8
8	8	3
9	9	0
10	10	0

Exploiting Data and Human Knowledge for Predicting Wildlife Poaching. Swaminathan Gurumurthy, Lantao Yu, Chenyan Zhang, Yongchao Jin, Weiping Li, Xiaodong Zhang, Fei Fang. In COMPASS-18

Machine Learning + Game Theoretic Reasoning



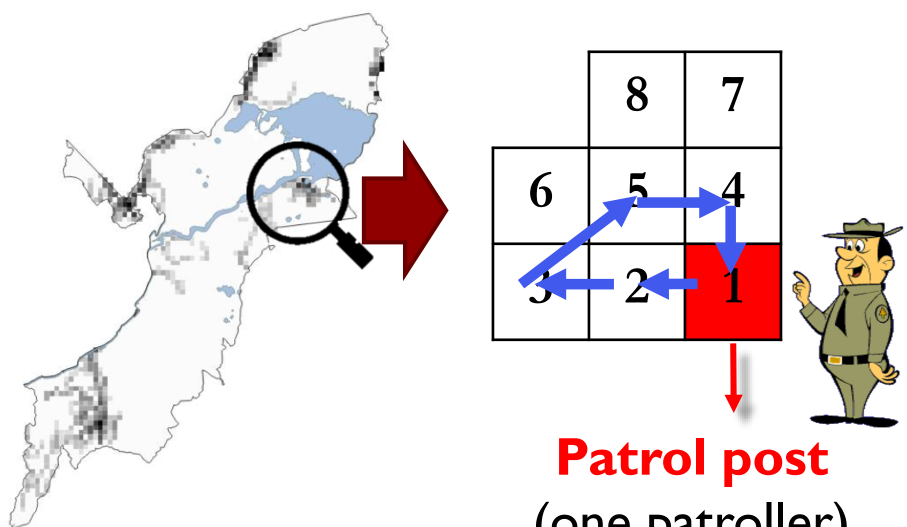
Game Theoretic Reasoning Based on Learned Model

For each cell *i*:

Machine Learning Model

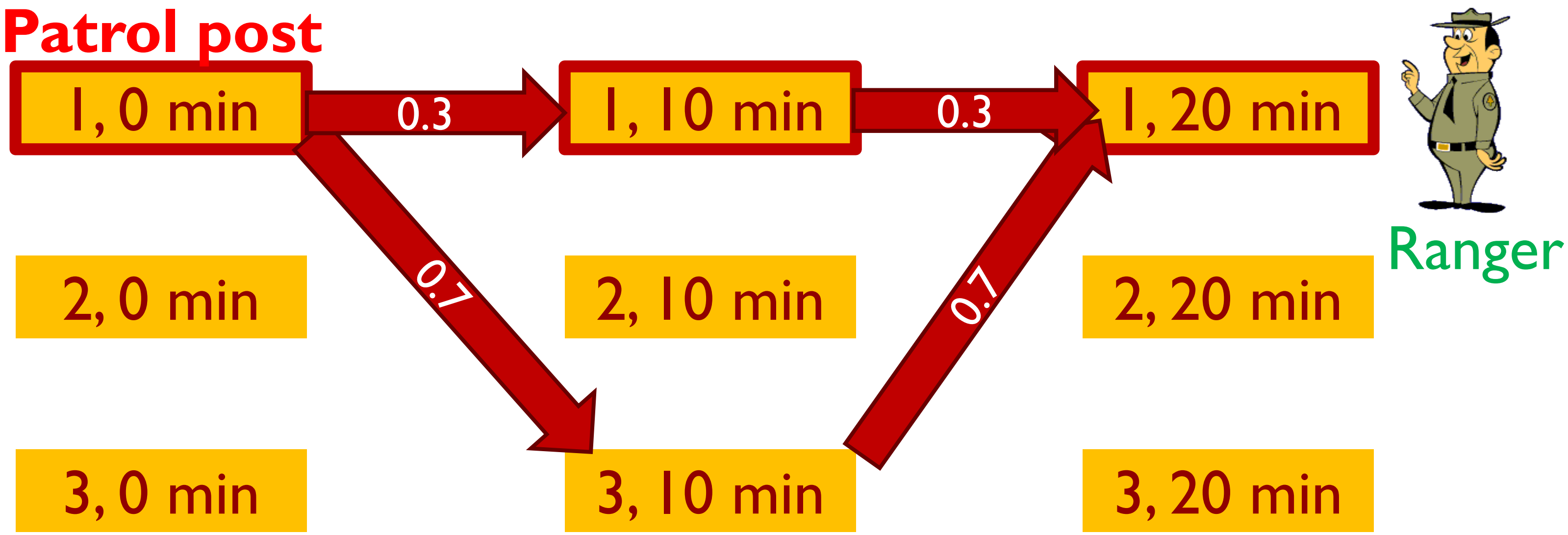
x_i: Current patrol effort at i → g_i → y_i: Prob. of snaring or detection at i in current period

- Maximize $\sum_i y_i$
- Scheduling constraint



Patrol post
(one patroller)

Game Theoretic Reasoning Based on Learned Model

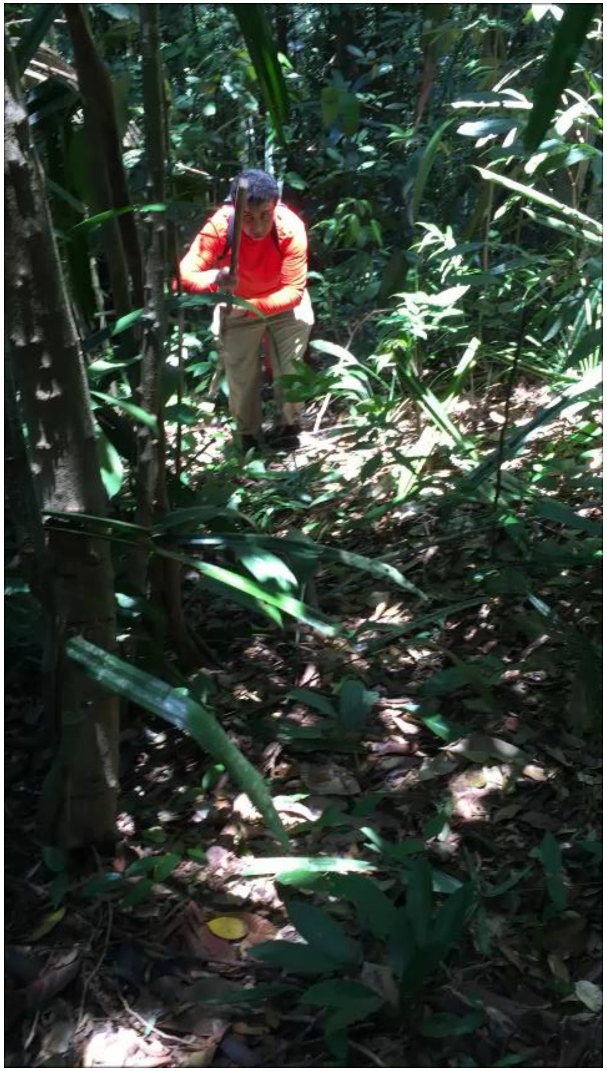


$$\max_{x_i} \sum_i g_i(x_i)$$

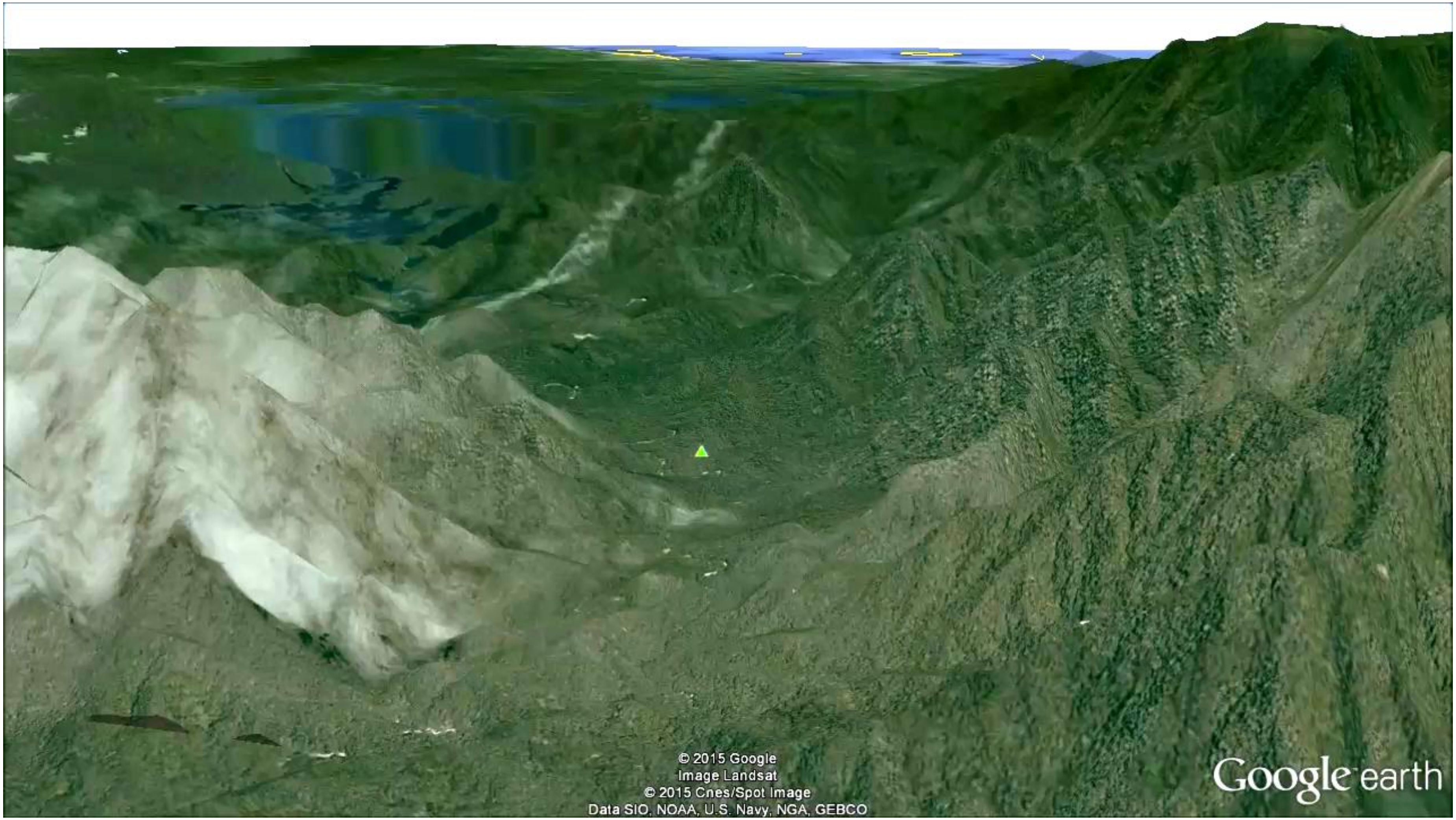
s.t. $x_i = \sum_t \sum_{e \in (i,t)\rightarrow} f(e)$

f is a unit flow

Complex Terrain

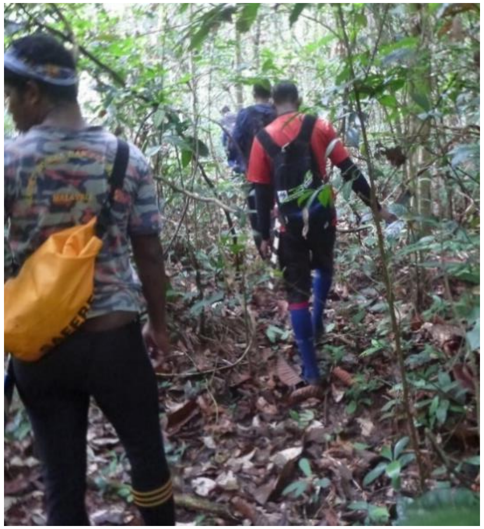


Model Terrain to Get Virtual Street Map



Field Test in Malaysia

- In collaboration with Panthera, Rimba
- Regular deployment since July 2015 (Malaysia)

Real-World Deployment

Animal Footprint

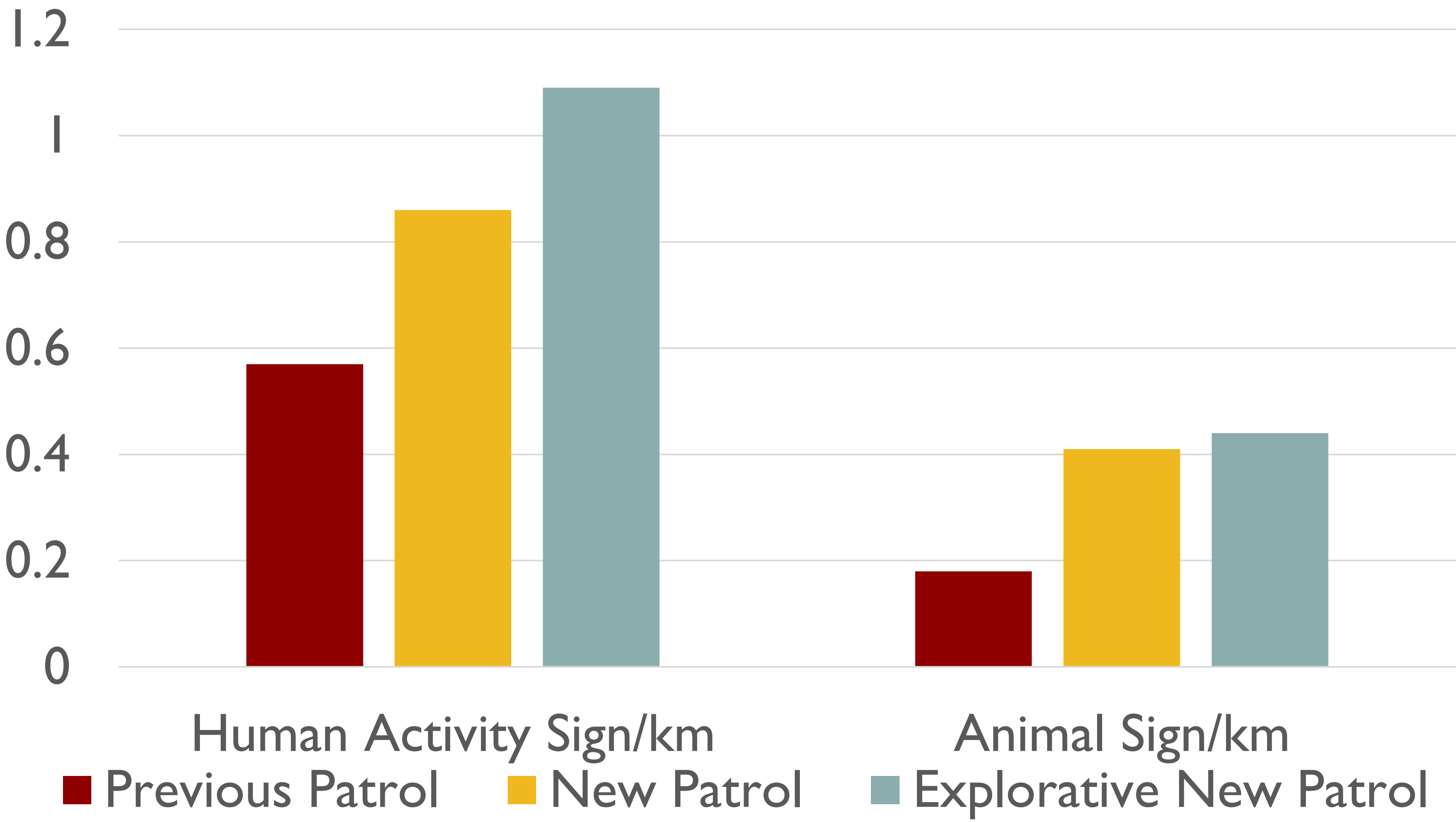
Tree Mark

Camping Sign

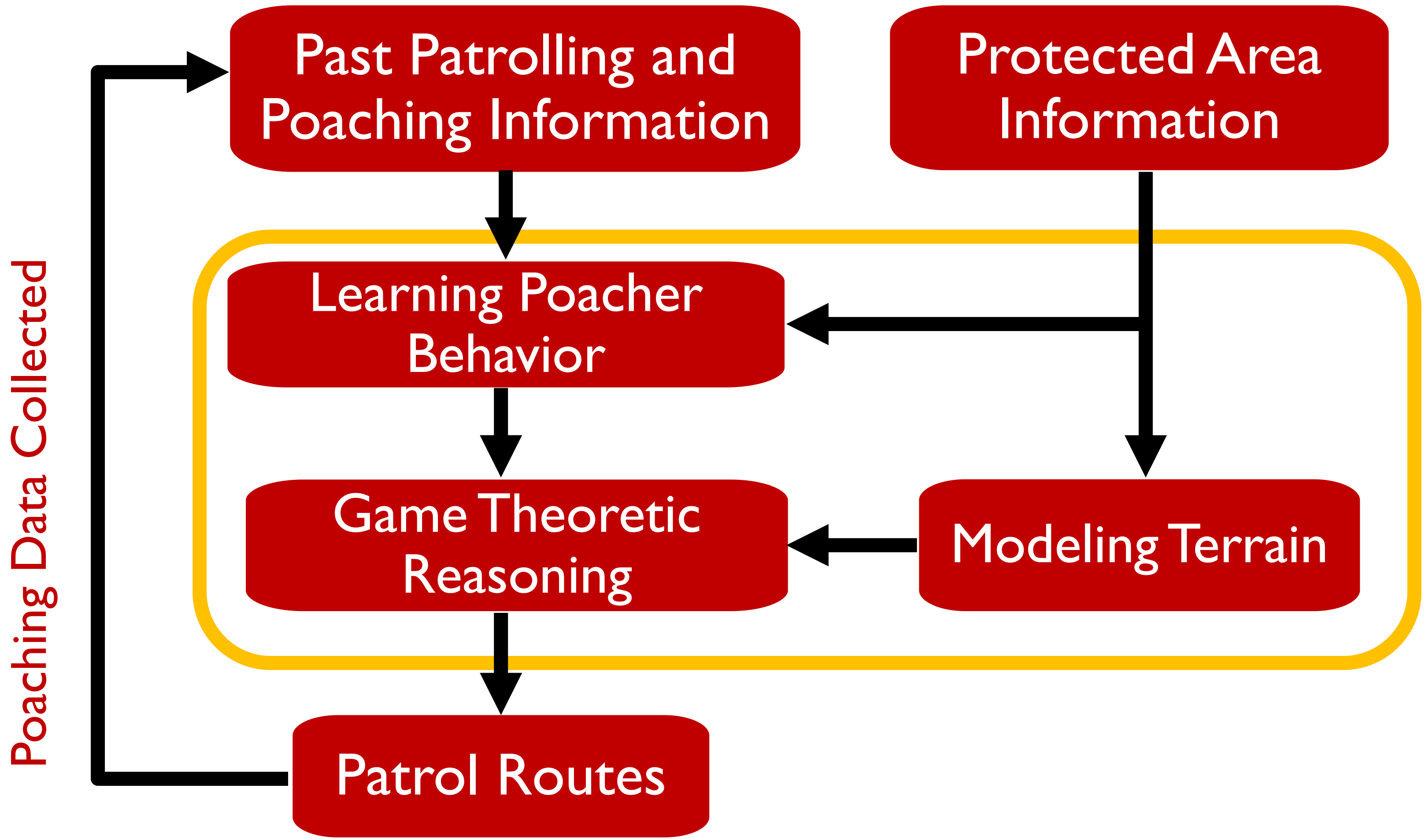
Tiger Sign

Lighter

Real-World Deployment

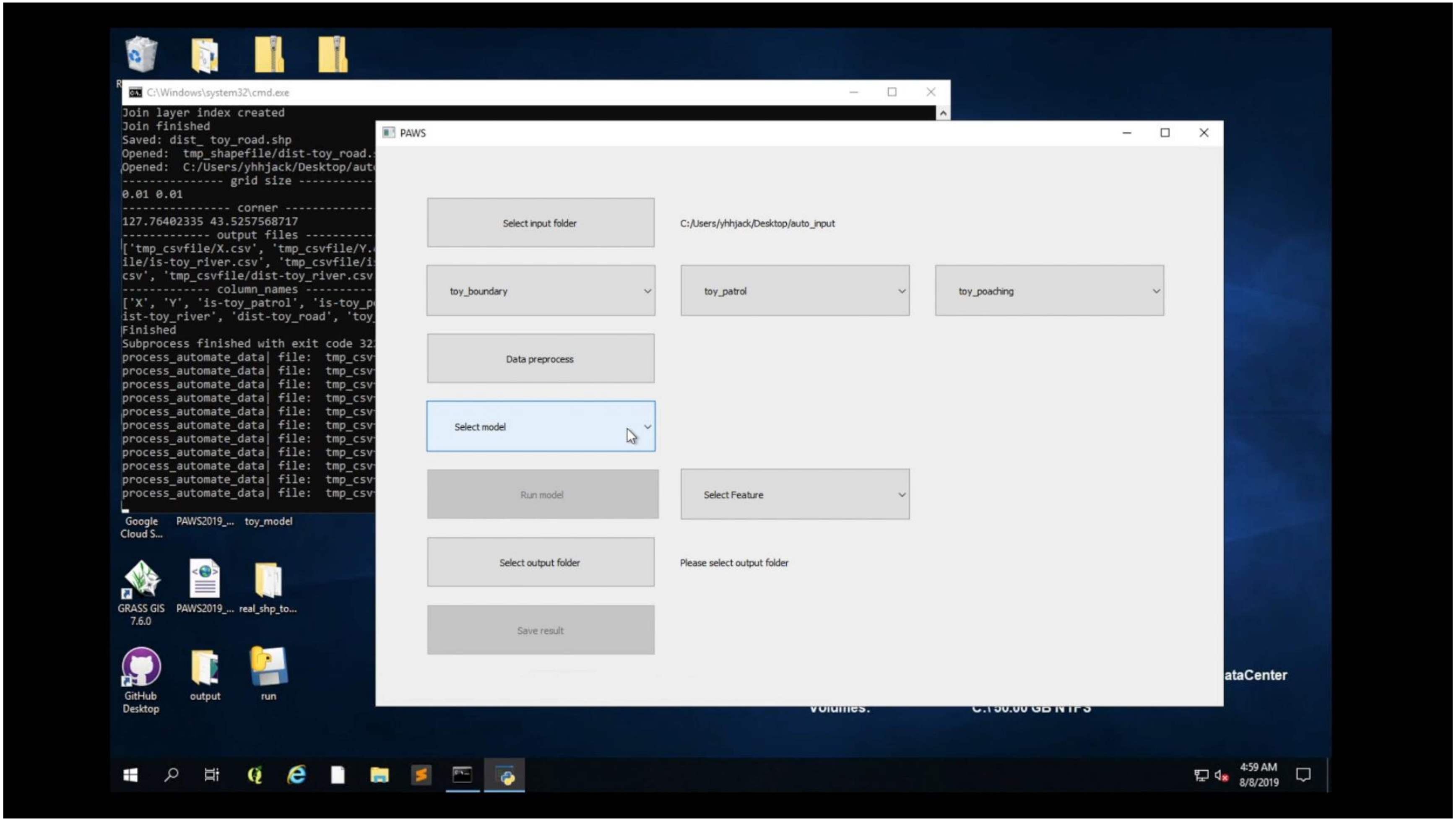


PAWS: Protection Assistant for Wildlife Security



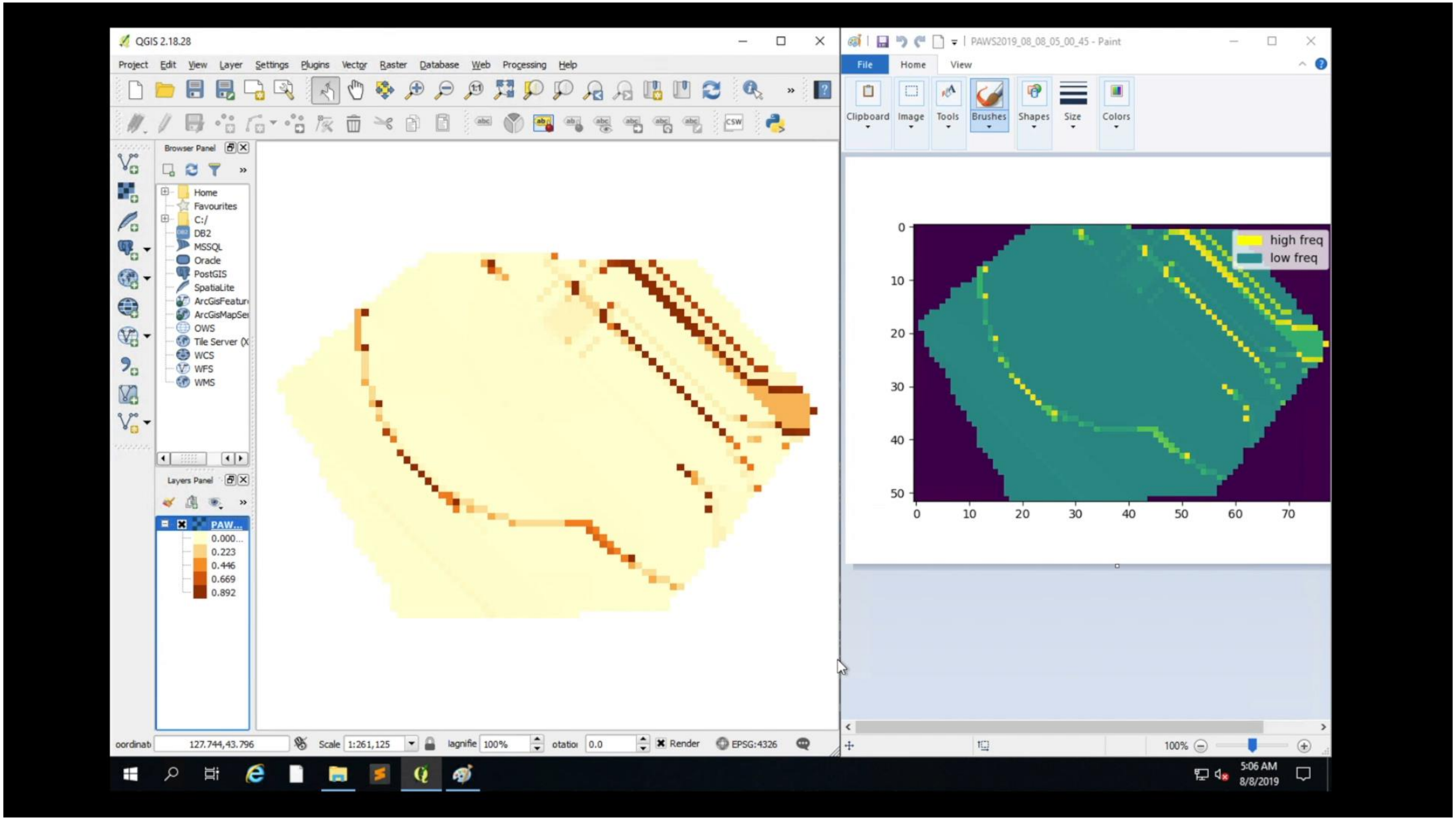
Deploying PAWS: Field Optimization of the Protection Assistant for Wildlife Security
Fei Fang, Thanh H. Nguyen, Rob Pickles, Wai Y. Lam, Gopalasamy R. Clements, Bo An, Amandeep Singh, Milind Tambe, Andrew Lemieux
In IAAI-16: The Twenty-Eighth Annual Conference on Innovative Applications of Artificial Intelligence, February 2016

Usable Software Tools



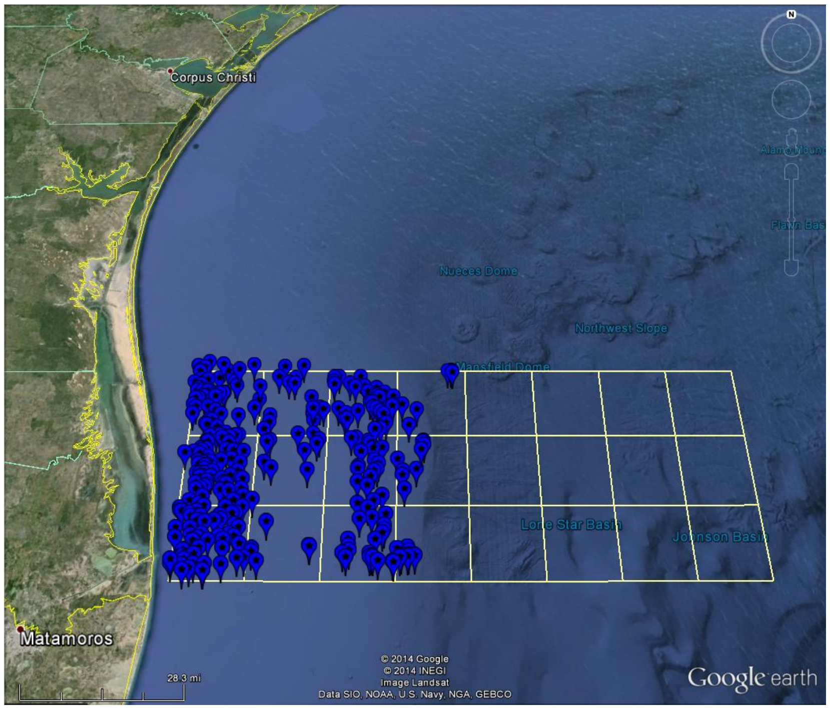
https://github.com/AIandSocialGoodLab/PAWS

Usable Software Tools

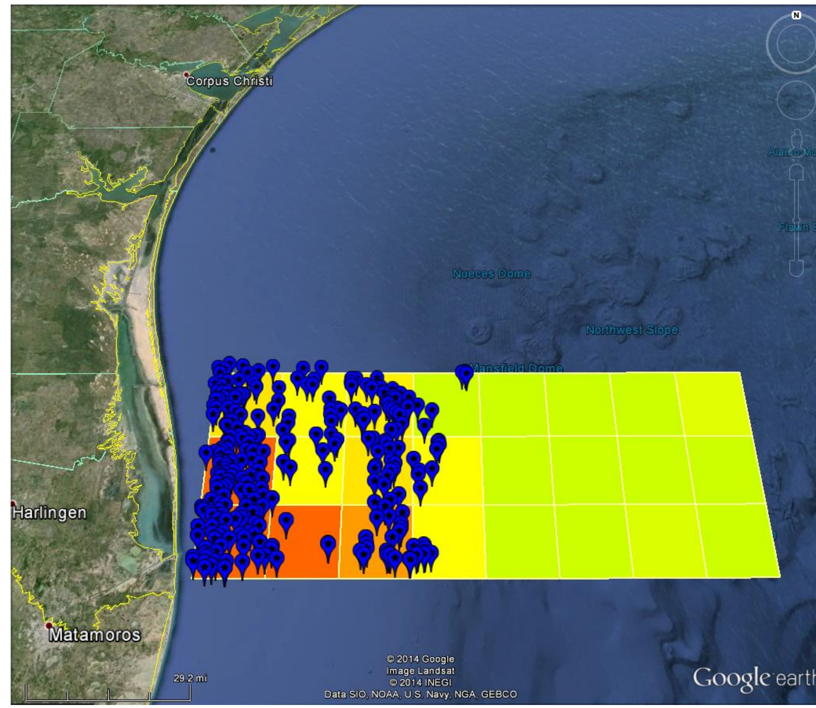


https://github.com/AIandSocialGoodLab/PAWS

Other Domains: Reduce Overfishing



Fish Data



Optimal Patrol Strategy

Outline

- Security games
- Integrating learning and game theory
  - Data-based game theoretic reasoning
  - Learning-powered equilibrium computation
  - End-to-end learning in games
- Summary

Patrol with Real-Time Information

- Sequential interaction
  - Players make flexible decisions instead of sticking to a plan
  - Players may leave traces as they take actions

- Example domain: Wildlife protection

Footprints

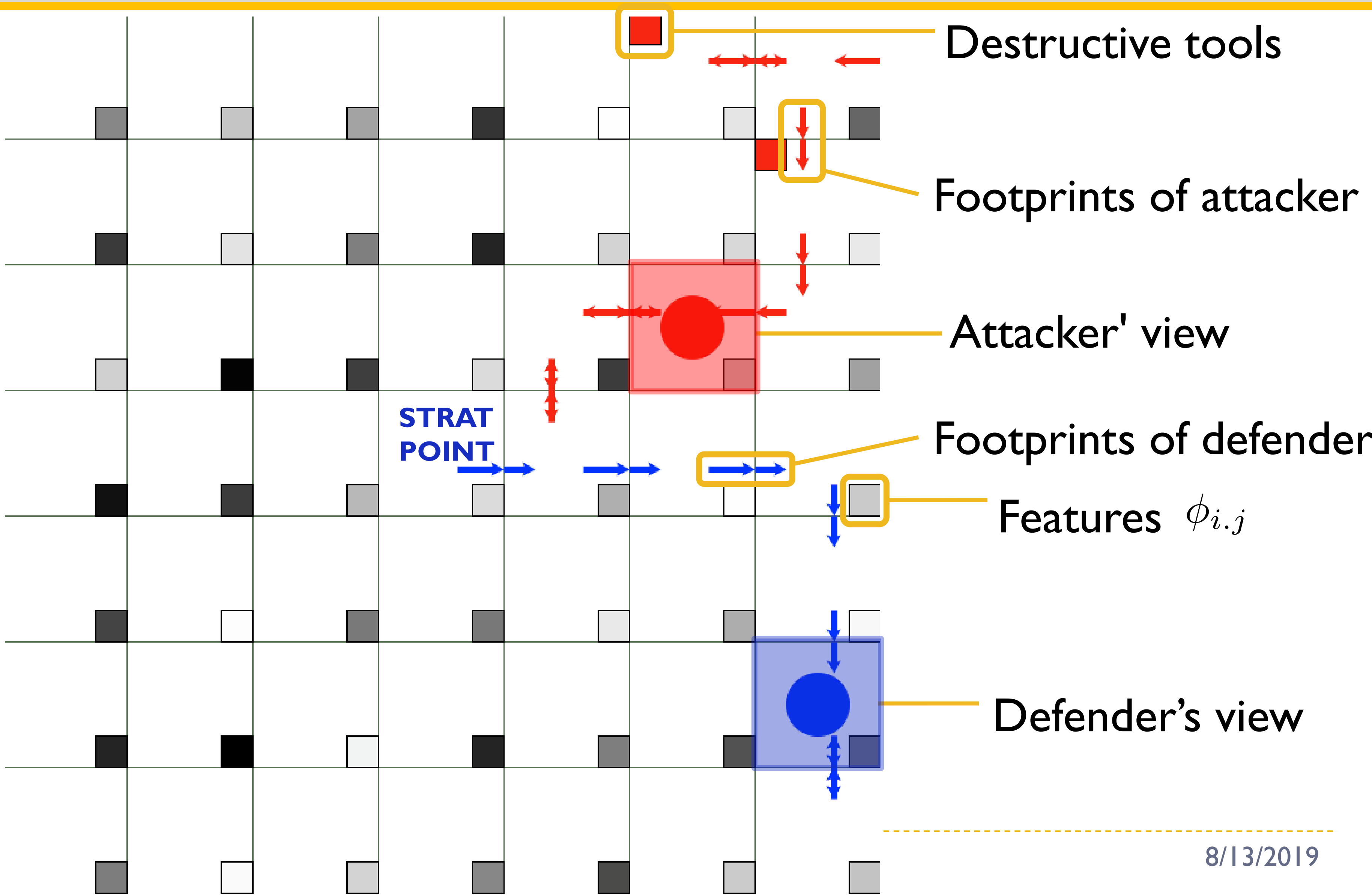
Lighters

Poacher camp

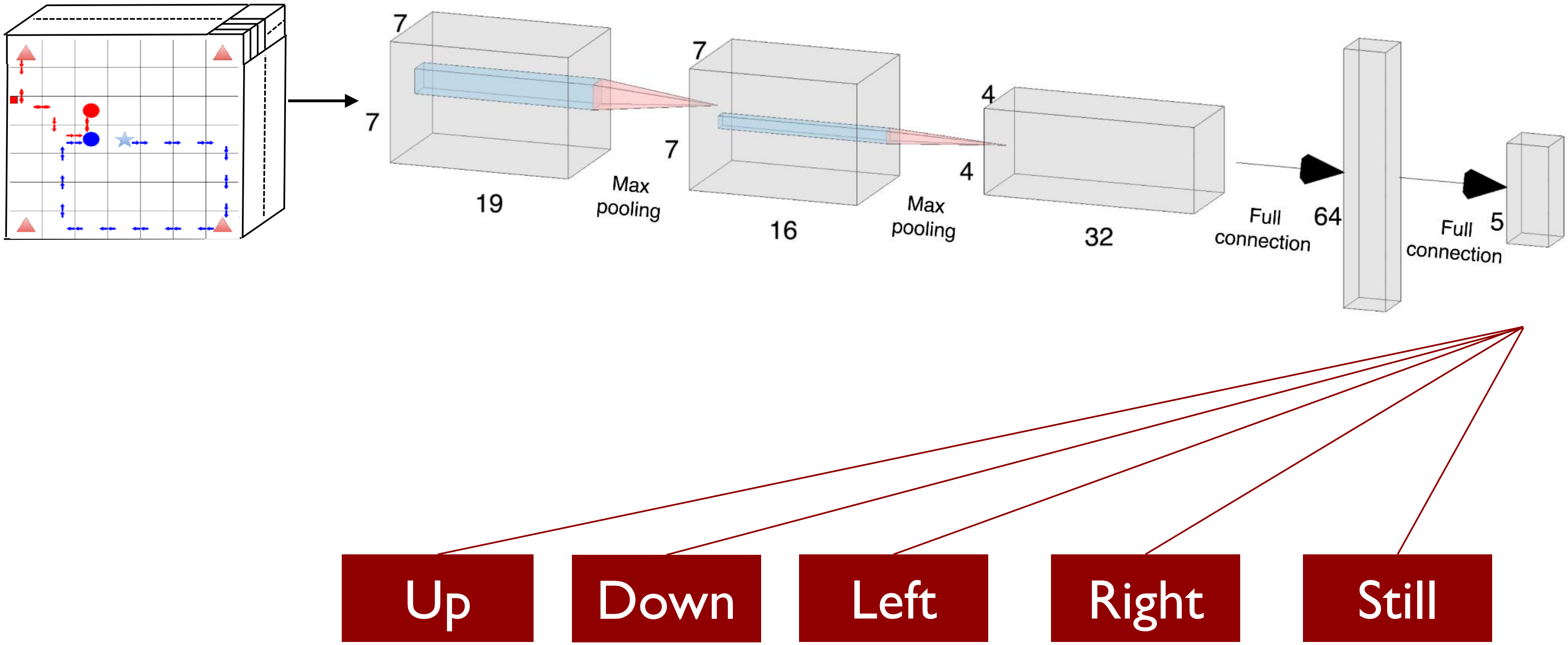
Tree marking

Deep Reinforcement Learning for Green Security Games with Real-Time Information Yufei Wang, Zheyuan Ryan Shi, Lantao Yu, Yi Wu, Rohit Singh, Lucas Joppa, Fei Fang In AAAI-19

Patrol with Real-Time Information

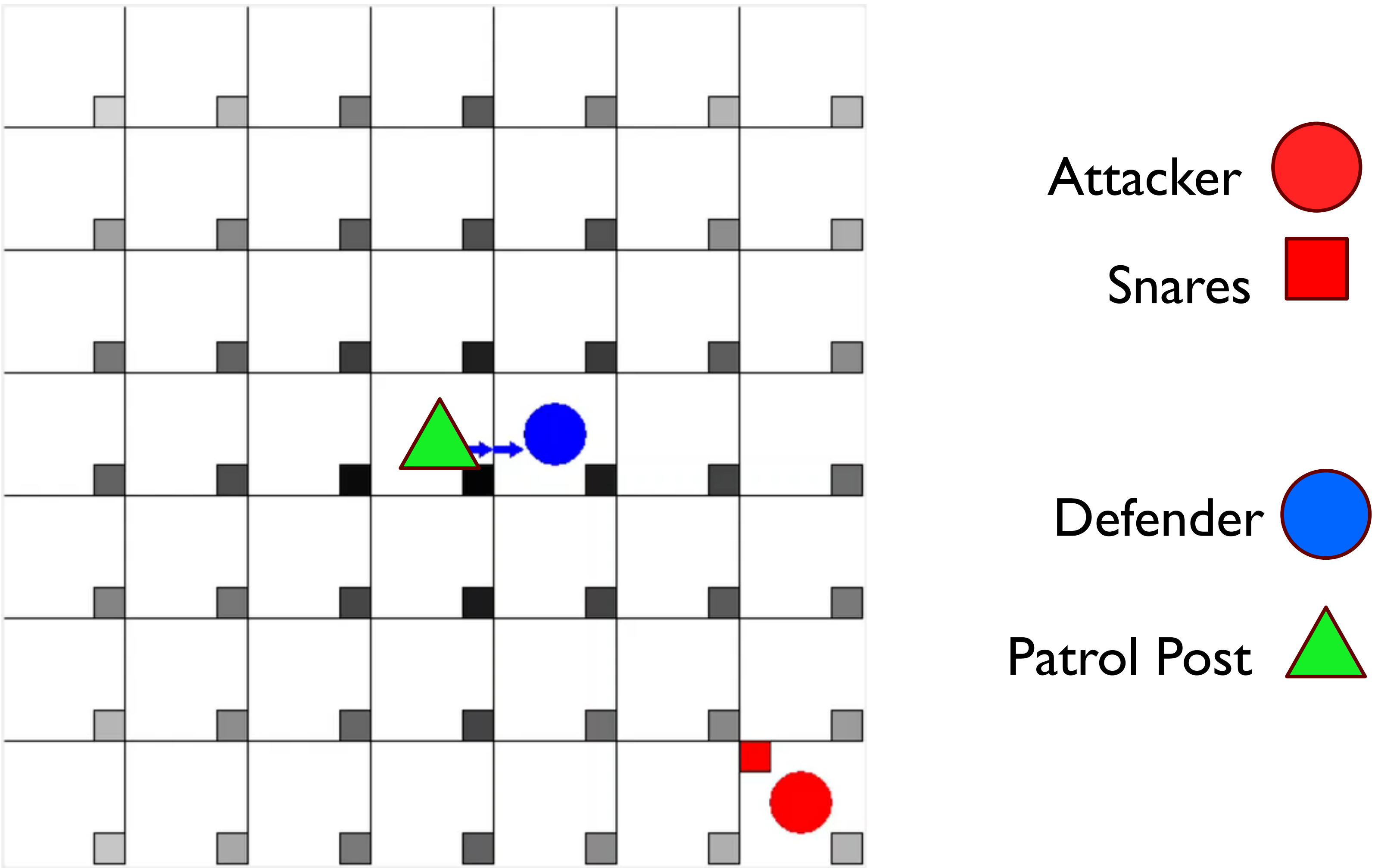


DQN Defender Trained Against Heuristic Attacker



- Q Network: Game state → Q-value

DQN Defender Trained Against Heuristic Attacker



Compute Equilibrium: DQN + Double Oracle

Compute $\sigma^d, \sigma^a = Nash(G^d, G^a)$

Compute Nash/Minimax

Train $f^d = DQN(\sigma^a)$

Find Best Response to attacker's strategy

Train $f^a = DQN(\sigma^a)$

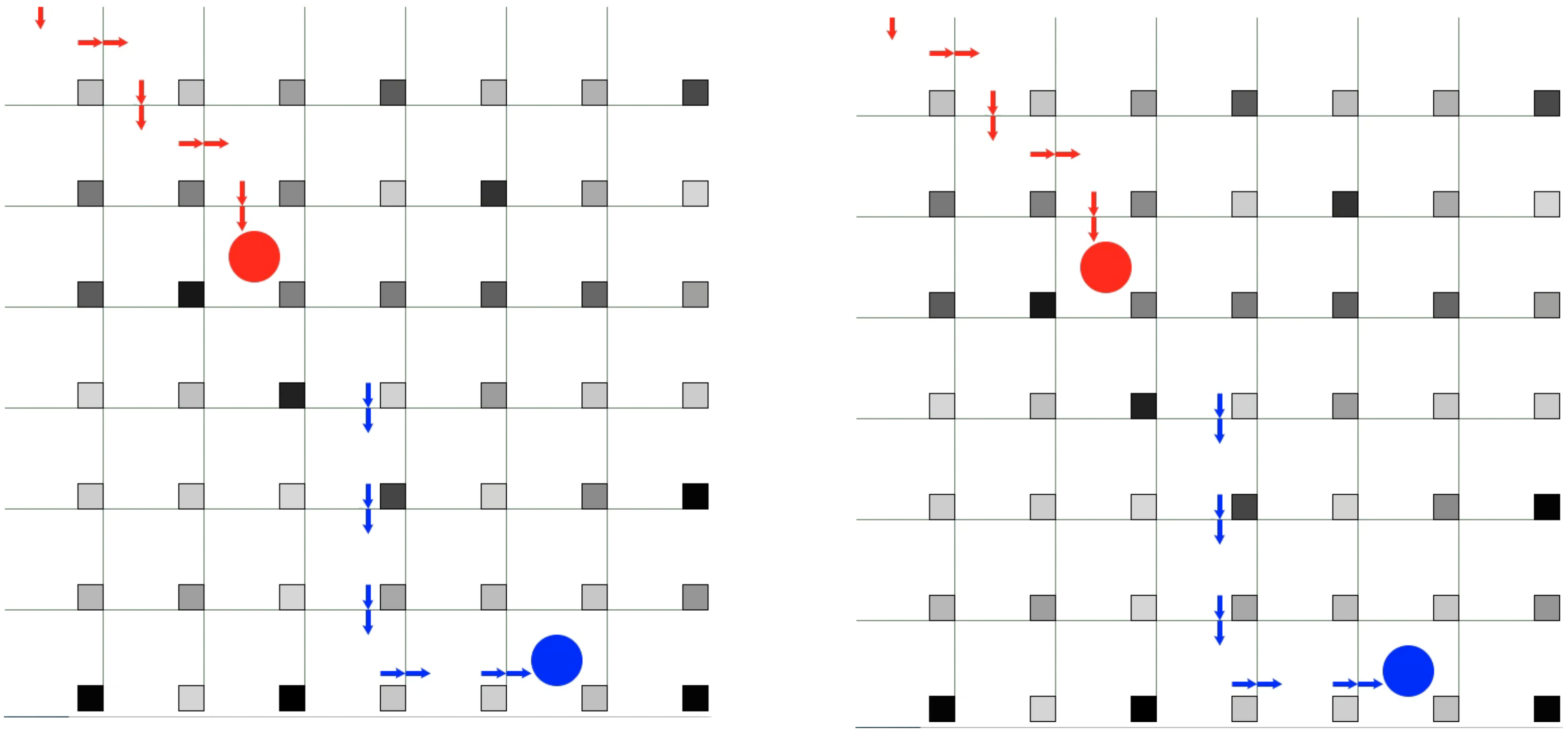
Find Best Response to defender's strategy

Add f^d, f^a to G^d, G^a

Update bags of strategies

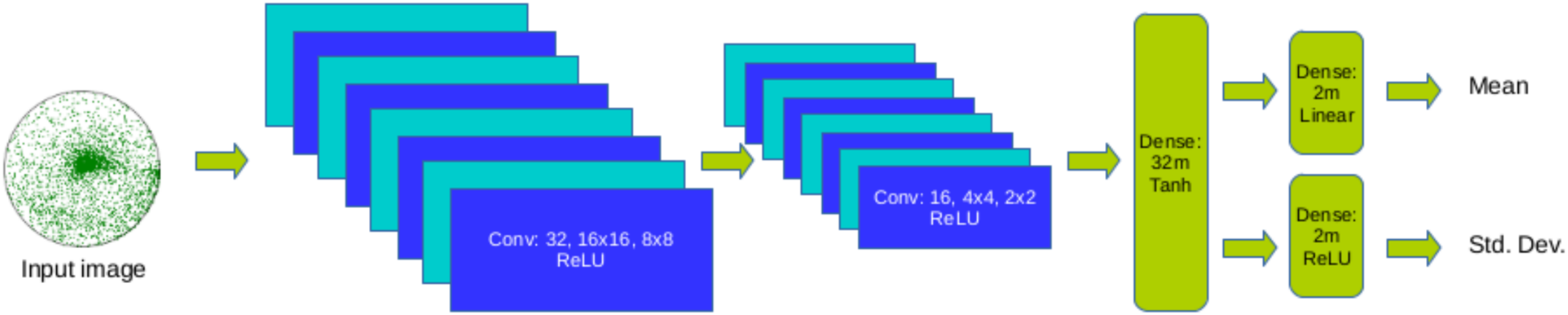
Enhancements

- Use local modes for efficient and parallized training
- Start with domain-specific heuristic strategies

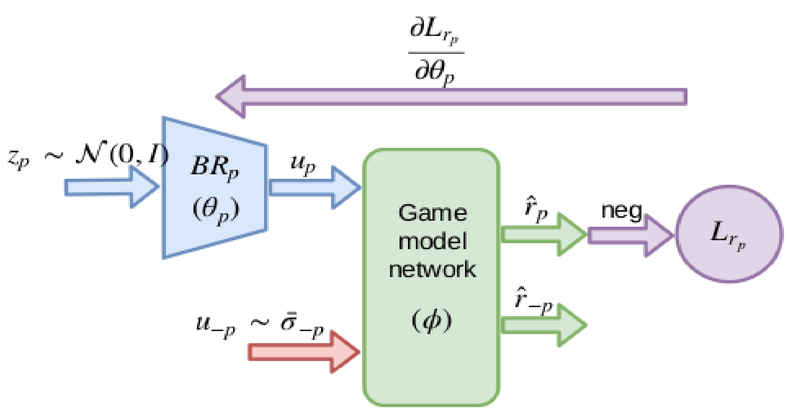


Other Domains: Patrol in Continuous Area

OptGradFP: CNN + Fictitious Play



DeepFP: Generative network + Fictitious Play

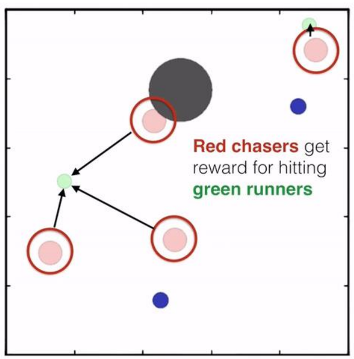


Policy Learning for Continuous Space Security Games using Neural Networks. Nitin Kamra, Umang Gupta, Fei Fang, Yan Liu, Milind Tambe. In AAAI-18

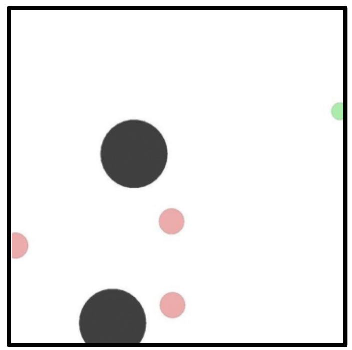
DeepFP for Finding Nash Equilibrium in Continuous Action Spaces. Nitin Kamra, Umang Gupta, Kai Wang, Fei Fang, Yan Liu, Milind Tambe. In GameSec-19

Other Domains: Multiple Agents

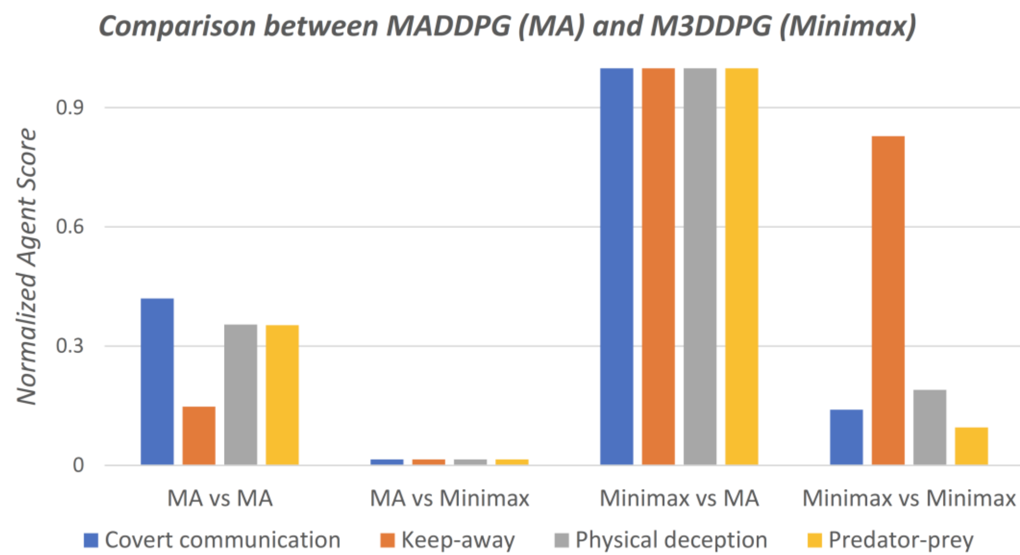
Multiple Chasers vs Runners



MADDPG



MADDPG (red) vs M3DDPG (green)



MADDPG + **Minimax** ≈ **Minimax MADDPG (M3DDPG)**

Adversarial Learning

Robust Multi-Agent Reinforcement Learning via Minimax Deep Deterministic Policy Gradient. Shihui Li, Yi Wu, Xinyue Cui, Honghua Dong, Fei Fang, Stuart Russell. In AAAI-19

Outline

- Security games
- Integrating learning and game theory
  - Data-based game theoretic reasoning
  - Learning-powered equilibrium computation
  - End-to-end learning in games
- Summary

What game are we/they playing?

- Common criticism: game parameters are fully known
- How to learn parameters of **2-player zero sum games** from opponents' or players' actions?

			
	0	b_1	$-b_2$
	$-b_1$	0	b_3
	b_2	$-b_3$	0

Learn

i.i.d samples from equilibrium strategies

$a^{(1)} =($  ,  $)$

$a^{(2)} =($  ,  $)$

$a^{(3)} =($  ,  $)$

What game are we playing? End-to-end learning in normal and extensive form games. Chun Kai Ling, Fei Fang, J. Zico Kolter. In IJCAI-ECAI-18

Differentiable Learning

			
	0	b_1	$-b_2$
	$-b_1$	0	b_3
	b_2	$-b_3$	0

Learn

i.i.d samples from equilibrium strategies

$a^{(1)}$ =(, )

$a^{(2)}$ =(, )

$a^{(3)}$ =(, )

- Guess the value of b_i
- Compute equilibrium of guessed game — Forward pass
- Check if the computed equilibrium consistent with data
- Adjust the value of b_i to increase consistency (update b)
- Repeat until satisfied

$$b_i := b_i - \frac{\partial L}{\partial b_i}$$ Backward pass

Learning of normal form games

- QRE = solution of min-max convex-concave problem

$$\min_u \max_v u^T P v - \sum_i v_i \log v_i + \sum_i u_i \log u_i$$

s.t.

$$1^T u = 1$$
$$1^T v = 1$$

- KKT conditions:

$$Pv + \log(u) + 1 + \mu 1 = 0$$
$$P^T u - \log(v) - 1 + \nu 1 = 0$$
$$1^T u = 1,\ 1^T v = 1$$

Learning of normal form games

- Forward pass: Apply Newton's Method

$$\begin{bmatrix} diag(\frac{1}{u}) & P & 1 & 0 \\ P^T & -diag\left(\frac{1}{v}\right) & 0 & 1 \\ 1^T & 0 & 0 & 0 \\ 0 & 1^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta v \\ \Delta \mu \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} Pv + \log(u) + 1 + \mu 1 \\ P^T u - \log(v) - 1 + \nu 1 \\ 1^T u - 1 \\ 1^T v - 1 \end{bmatrix}$$

- Backward pass: Implicit function theorem

$$\nabla_P L = y_u v^T + u y_v^T,$$

where

$$\begin{bmatrix} y_u \\ y_v \\ y_\mu \\ y_\nu \end{bmatrix} = \begin{bmatrix} \text{diag}(\frac{1}{u}) & P & 1 & 0 \\ P^T & -\text{diag}(\frac{1}{v}) & 0 & 1 \\ 1^T & 0 & 0 & 0 \\ 0 & 1^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla_u L \\ -\nabla_v L \\ 0 \\ 0 \end{bmatrix}$$

Learning in the presence of features

			
	0	$-b_1(x)$	$b_2(x)$
	$b_1(x)$	0	$-b_3(x)$
	$-b_2(x)$	$b_3(x)$	0

Learn

i.i.d samples from equilibrium strategies

$a^{(1)}$ =( , )

$a^{(2)}$ =( , )

$a^{(3)}$ =( , )

...

Context

$$x^{(1)} = [0.1, 0.5]$$

$$x^{(2)} = [0.3, 0.7]$$

$$\ldots$$

End-to-end learning

Algorithm 1: Learning parameters Φ using SGD

Input: training data $\{(x^{(i)}, a^{(i)})\}$, learning rate η, Φ_{init}

for *ep in* $\{0, \ldots, ep_{max}\}$ **do**

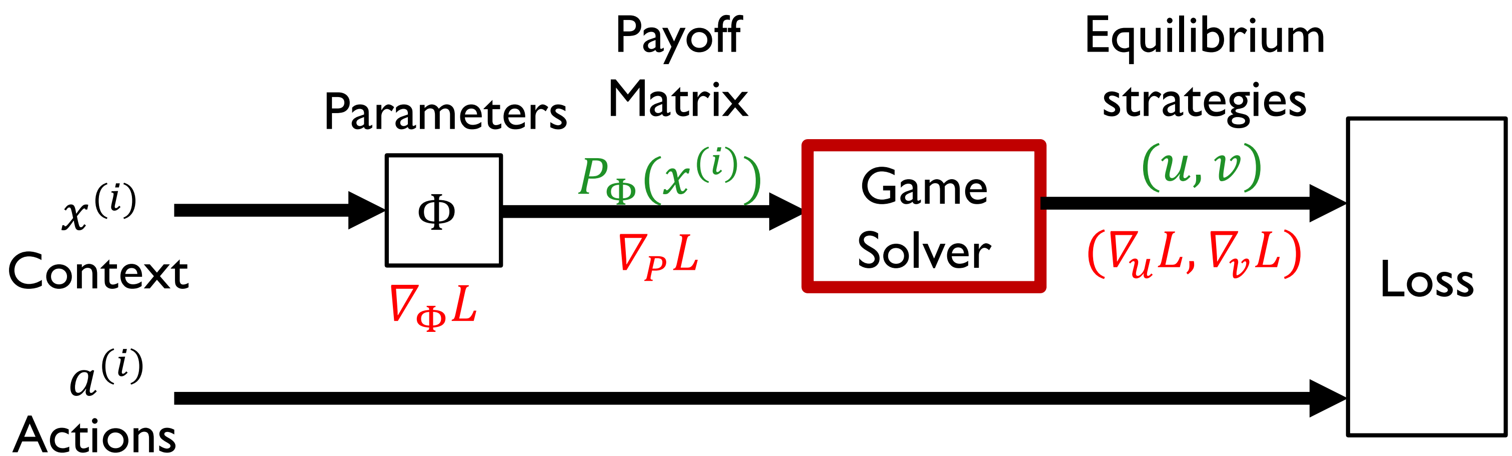
Sample $(x^{(i)}, a^{(i)})$ from training data;

Forward pass: Compute $P_\Phi(x^{(i)})$, QRE (u, v) and loss $L(a^{(i)}, u, v)$;

Backward pass: Compute gradients $\nabla_u L, \nabla_v L, \nabla_P L, \nabla_\Phi L$;

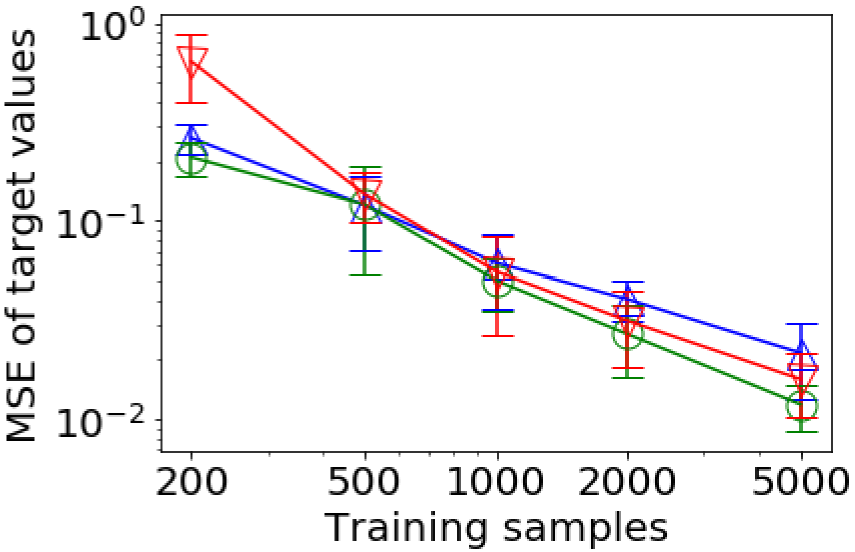
Update parameters: $\Phi \leftarrow \Phi - \eta \nabla_\Phi L$;

end

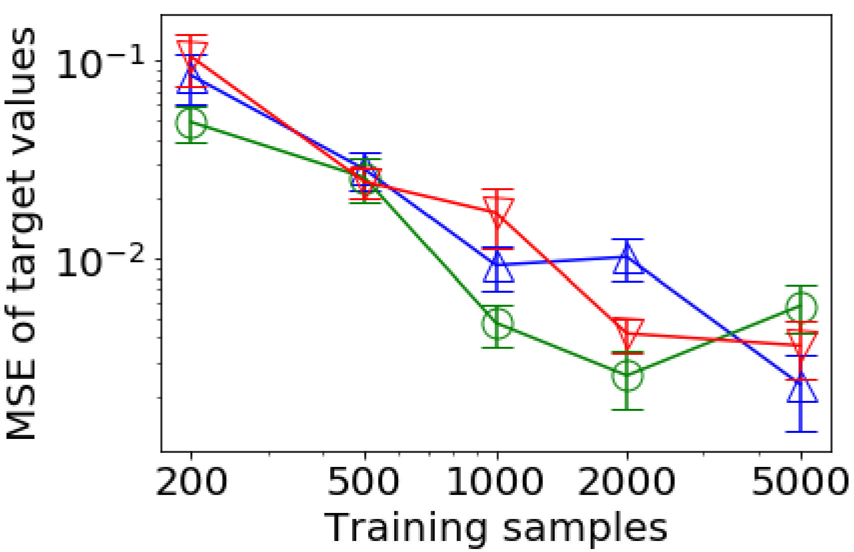


Resource Allocation Security Game

- $n = 2, r = 5$



T= 1



T= 2

Extend to extensive form Games

- Equilibrium is expressed as solution using *dilated entropy regularization*

$$\min_u \max_v u^T P v - \sum_i \sum_a v_a \log(\frac{v_a}{v_{p_i}}) + \sum_i \sum_a u_a \log(\frac{u_a}{u_{p_i}})$$
$$Eu = e, Fv = f$$

$$\nabla_P L = y_u v^T + u y_v^T ,$$

where

$$\begin{bmatrix} y_u \\ y_v \\ y_\mu \\ y_\nu \end{bmatrix} = \begin{bmatrix} -\Xi(u) & P & E^T & 0 \\ P^T & \Xi(v) & 0 & F^T \\ E & 0 & 0 & 0 \\ 0 & F & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla_u L \\ -\nabla_v L \\ 0 \\ 0 \end{bmatrix}$$

Improve Scalability using FOM

- Both the problem in the forward pass and backward pass can be converted into the following format

$$\min_{Ex=x_0} \max_{Fy=y_0} x^T P y + \boxed{\mathcal{E}(x) - \mathcal{F}(y)}$$

strictly convex functions

- This problem can be solved using various first-order iterative methods

Input: $x^{(0)}, y^{(0)}, P, \tau, \sigma$
for i *in* $\{0, \ldots\}$ **do**
  $\tilde{y} = y^{(i)}$;
  $x^{(i+1)} = \boxed{\mathrm{BR}_x(x^{(i)}, \tilde{y}; P, \tau);}$
  $\tilde{x} = 2x^{(i+1)} - x^{(i)}$;
  $y^{(i+1)} = \boxed{\mathrm{BR}_y(y^{(i)}, \tilde{x}; P, \sigma);}$
end

BR is smoothed best response

$$\mathrm{BR}_x(\bar{x}, \tilde{y}) = \arg\min_{Ex=x_0} x^T P \tilde{y} + \mathcal{E}(x) + \frac{1}{\tau} D_x(x, \bar{x})$$

$$\mathrm{BR}_y(\bar{y}, \tilde{x}) = \arg\min_{Fy=y_0} -\tilde{x}^T P y + \mathcal{F}(y) + \frac{1}{\sigma} D_y(y, \bar{y}).$$

Large Scale Learning of Agent Rationality in Two-Player Zero-Sum Games. Chun Kai Ling, Fei Fang, Zico Kolter. In AAAI-19

Takeaway 1: AI Has Great Potential for Social Good

Artificial Intelligence

Machine Learning / Reinforcement Learning

Computational Game Theory

Societal Challenges

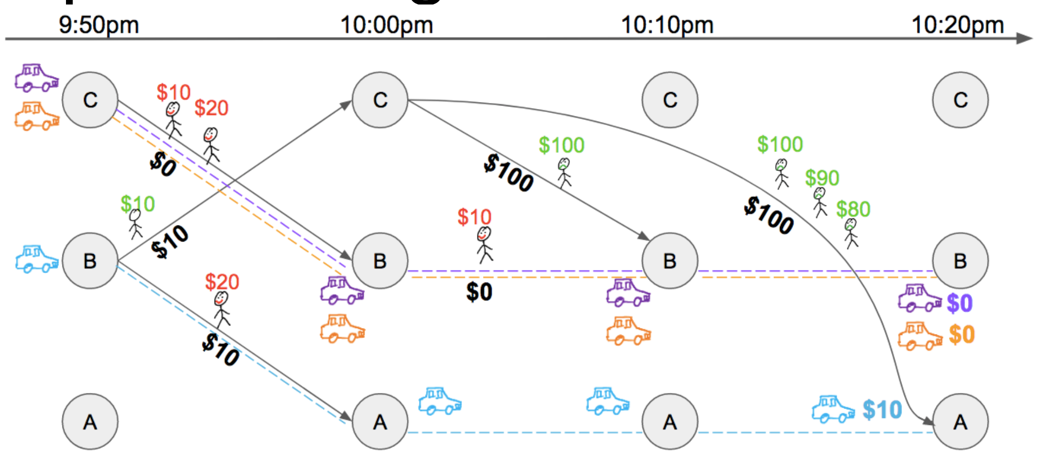
Security & Safety

Environmental Sustainability

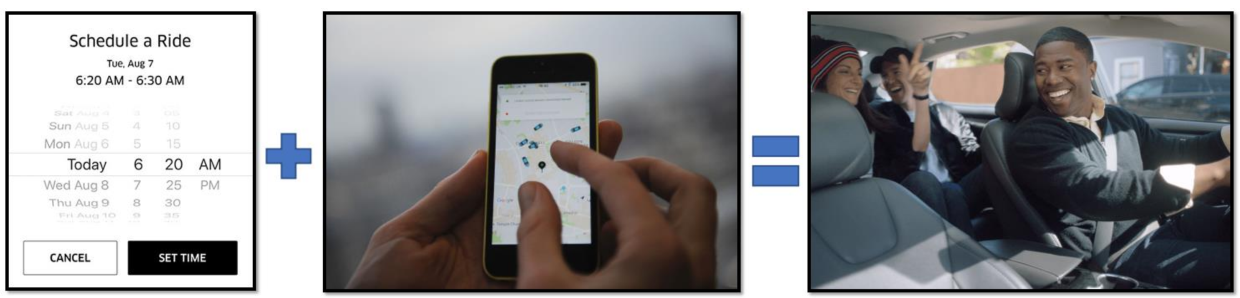
Mobility

Design Dispatching and Pricing Scheme in Ridesharing

- Spatial-Temporal Pricing



- Two-Stage Dispatching



Stage 1: Scheduled Requests

Stage 2: Real-Time Requests

Successful Dispatching

Spatio-Temporal Pricing for Ridesharing Platforms. Hongyao Ma, Fei Fang, David C. Parkes. In EC-19

Dynamic Trip-Vehicle Dispatch with Scheduled and On-Demand Requests. Taoan Huang, Bohui Fang, Xiaohui Bei, Fei Fang. In UAI-19

Takeaway 2: Ways to Integrate Learning and Game Theory

- Data-based game theoretic reasoning
- Learning-powered equilibrium computation
- End-to-end learning in games
- More?!

Acknowledgment

- Advisors, students and all co-authors!





- Collaborators and partners

- Funding support