Artificial Intelligence Methods for Social Good

M3-2 [Machine Learning]:

Predicting Poaching Activity and Urban Crime

08-537 (9-unit) and 08-737 (12-unit)
Instructor: Fei Fang
feifang@cmu.edu
Wean Hall 4126

Outline

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

Learning Objectives

- Describe a few models for real-world spatio-temporal prediction tasks such as predicting poaching activity and urban crime
- Answer the representation, inference, learning questions w.r.t. the models
- Describe several evaluation metrics for these models
- Describe methodologies of field tests for these tasks

Wildlife Protection in Uganda

Forest Area: QEPA

- Covers 2520 sq. km
- Divided into a grid of 1km×1km

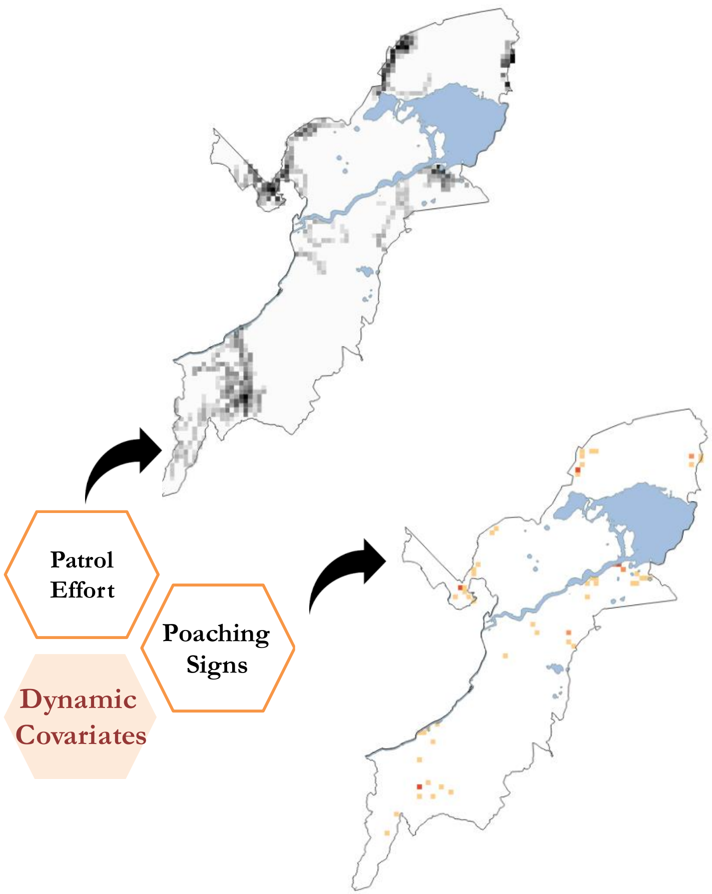
Poachers: Set trapping tools (e.g., snare)

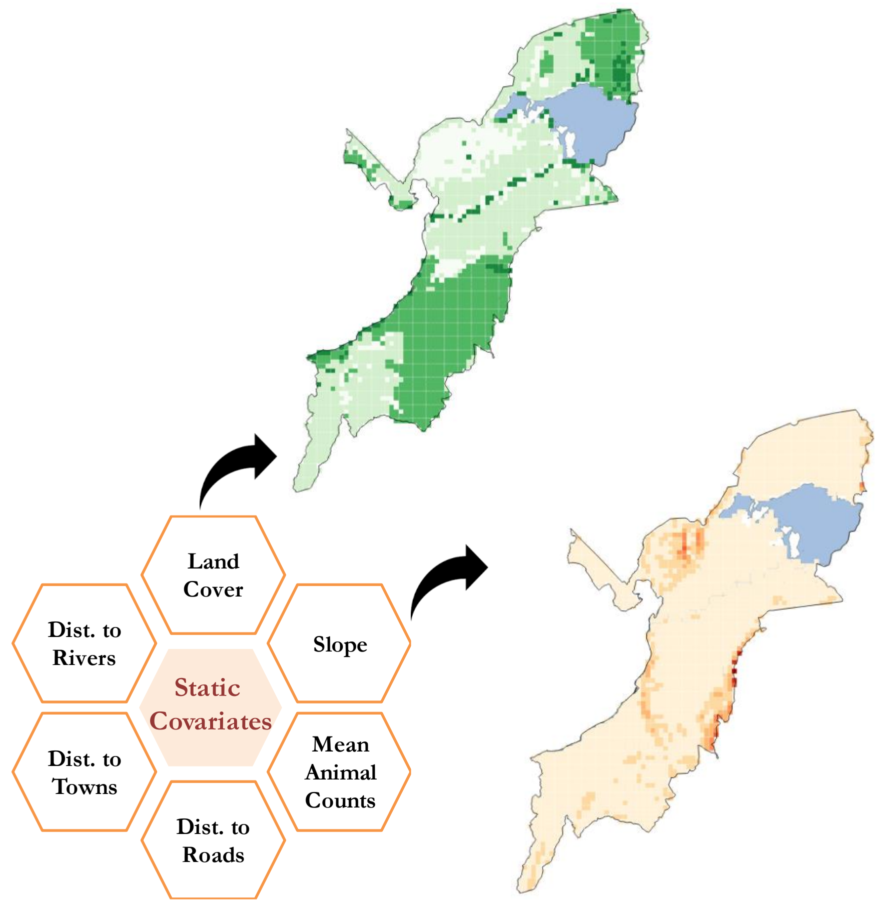
Rangers: Conduct patrols

- On foot or by ground vehicles
- From 2003-2017

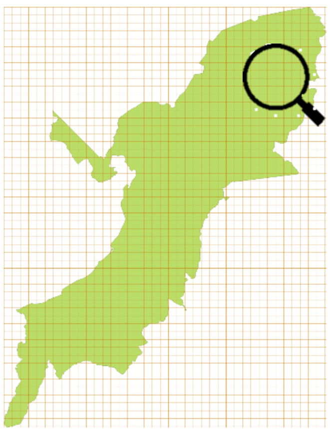
Collaborators: Wildlife Conservation Society, Uganda Wildlife Authority,
Rangers Pictures: Trip to Indonesia with World Wide Fund for Nature

Dataset Covariates: Queen Elizabeth Park



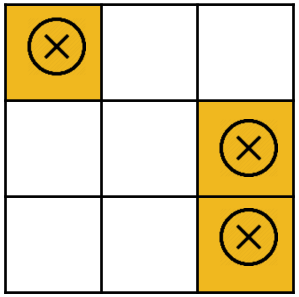


Challenges: Data Uncertainty

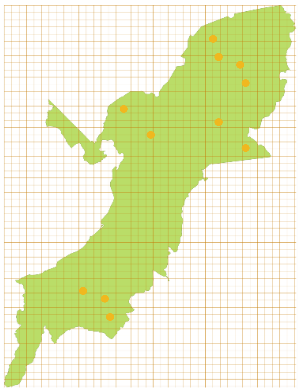


Attacked

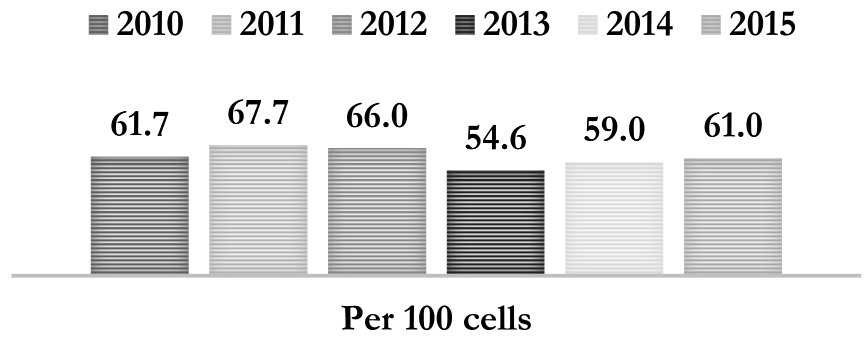
Not Attacked



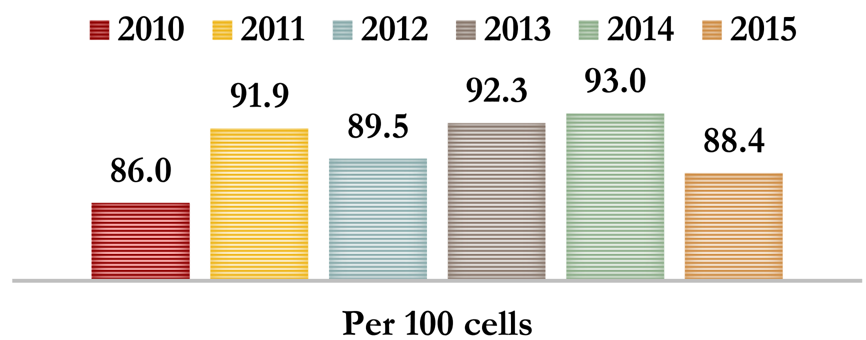
Challenges: Small Number of Recorded Attacks



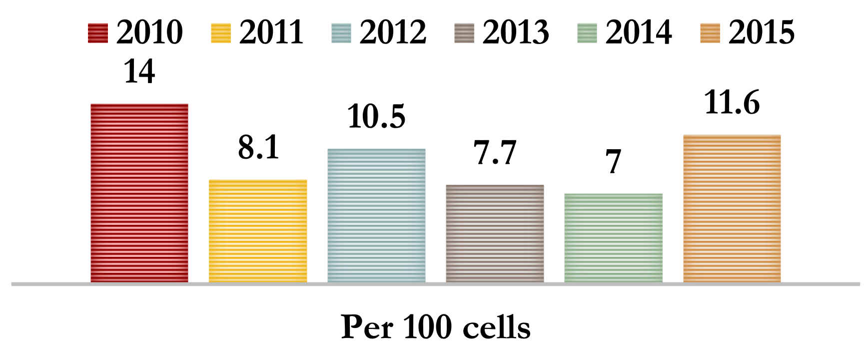
Patrolled Cells
(Year)



Not Attacked Patrolled Cells



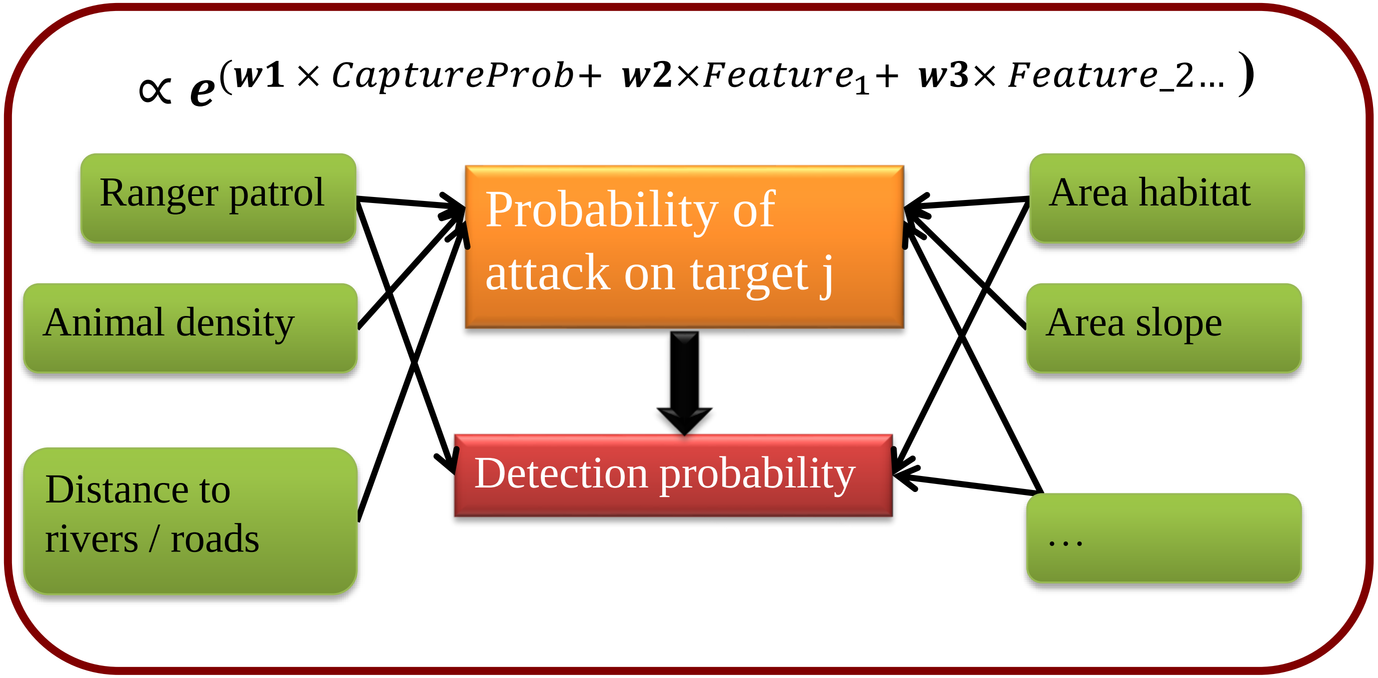
Attacked Patrolled Cells



Outline

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

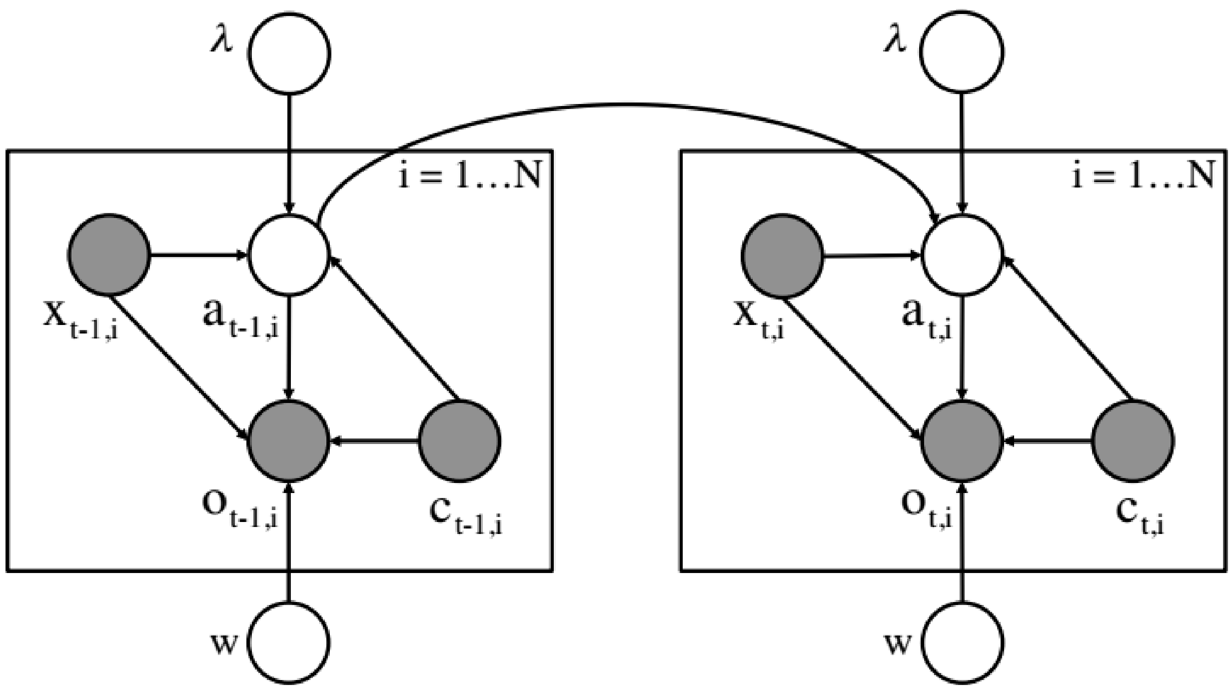
CAPTURE – Single Time Step



Nguyen et al. Capture: A new predictive anti-poaching tool for wildlife protection. In AAMAS, 2016

CAPTURE – Multiple Time Steps

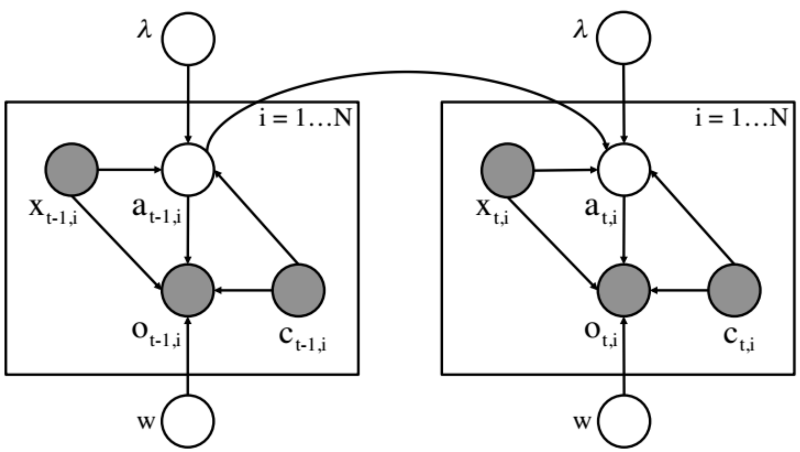
- Temporally-aware Dynamic Bayes Net



- Step t-1
- Step t

CAPTURE – Multiple Time Steps

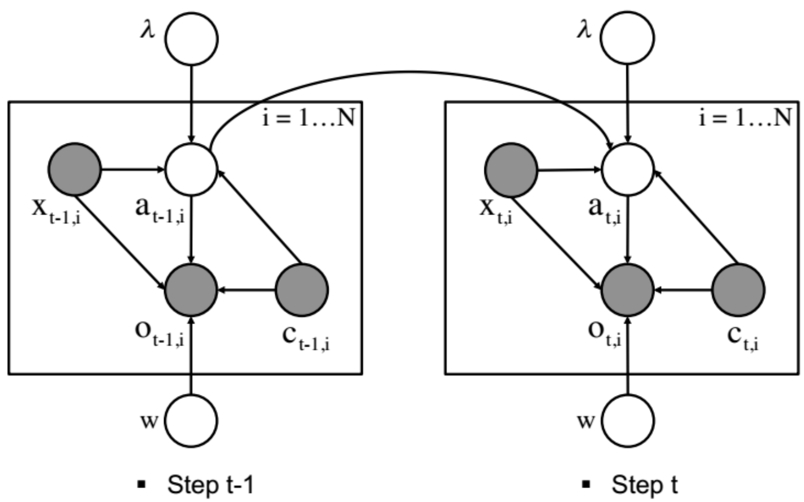
- Conditional probabilities
  - Logistic model
  - $p(a_{t,i} = 1 | a_{t-1,i}, c_{t,i}, x_{t,i}) = \frac{e^{\lambda^T [a_{t-1,i}, c_{t,i}, x_{t,i}, 1]}}{1 + e^{\lambda^T [a_{t-1,i}, c_{t,i}, x_{t,i}, 1]}}$
  - $p(o_{t,i} = 1 | a_{t,i} = 1, c_{t,i}, x_{t,i}) = c_{t,i} \times \frac{e^{w^T [x_{t,i}, 1]}}{1 + e^{w^T [x_{t,i}, 1]}}$



- Step t-1
- Step t

Learning in CAPTURE

- Learn/Train CAPTURE
  - Given a set of data
  - Find weights λ, w (Expectation Maximization Algorithm + Parameter Separation + Target Abstraction)



Inference with CAPTURE

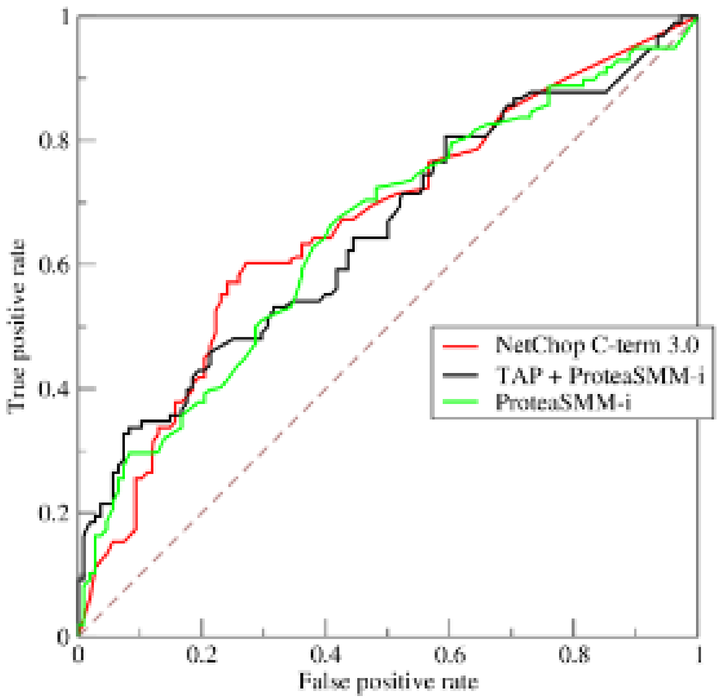
- Make inferences given trained CAPTURE
  - Infer the past
    - Input: Geospatial features, patrol coverage $c_{t,i}$, observations $o_{t,i}$, $t = 1 \dots T$
    - Output: Probability of poaching activity $a_{t,i}$, $t = 1 \dots T$
  - Predict the future
    - Input: Geospatial features, patrol coverage $c_{t,i}$, observations $o_{t,i}$, $t = 1 \dots T$; future patrol coverage $c_{T+1,i}$ (controlled by defender)
    - Output: Probability of poaching activity $a_{T+1,i}$ and probability of observing poaching activity $o_{T+1,i}$

Evaluate CAPTURE

- How to evaluate CAPTURE?
  - Data: Observations o only (no ground truth of a)
  - Evaluate "predict the future" task using historical data
  - Training/test sets
    - Training 1: Data in 2003–2006; Test 1: Data in 2007
    - …
    - Training 8: Data in 2010-2013; Test 8: Data in 2014

Evaluate CAPTURE

- What metrics can be used?
  - Accuracy / Recall / Precision / F1?
    - Need binary decision from probabilities
    - Set a threshold on probability
    - Value dependent on threshold
  - Receiver operating characteristic (ROC) curve
  - Area under the Curve (AUC)

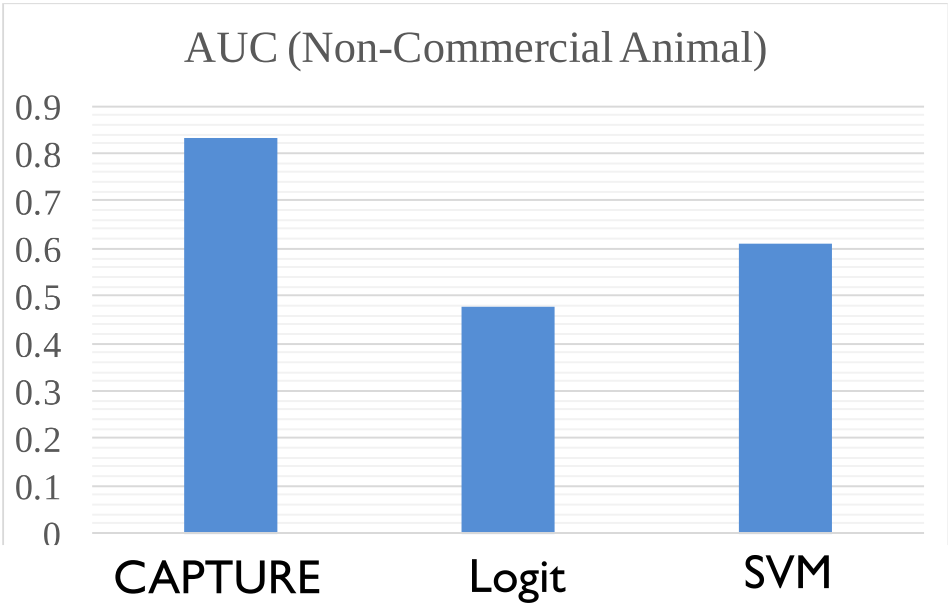


https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Quiz 1

- In ROC curve, the x-axis is false positive rate, and y-axis is true positive rate. Which point in the ROC space corresponds to a perfect classifier that makes correct predictions for all data points? Which point in the ROC space corresponds to a classifier that makes predictions based on the flip of a balanced coin?
- (0,1), (0.5,0.5)
- (0.5,0.5), (1,0)
- (1,0), (0.5,0.5)
- (0,1),(1,1)

Human Behavioral Model: CAPTURE



Dry Season (June-August 2008)

Nguyen et al. Capture: A new predictive anti-poaching tool for wildlife protection. In AAMAS, 2016

Limitations of CAPTURE

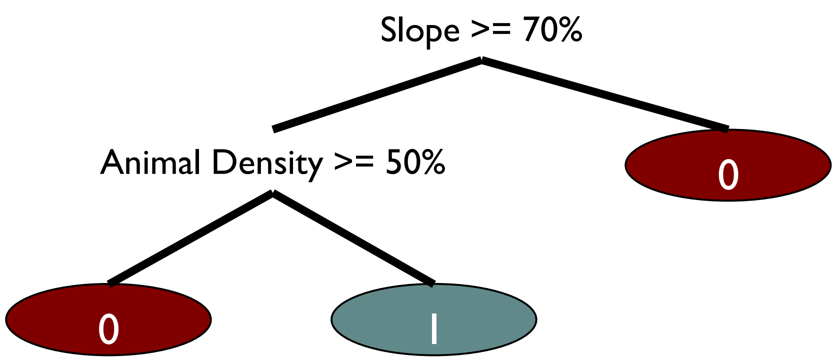
- Good at predicting observations o but not poaching activities a
- Difficulty to interpret
- Slow to run

Outline

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

Basic Decision Tree

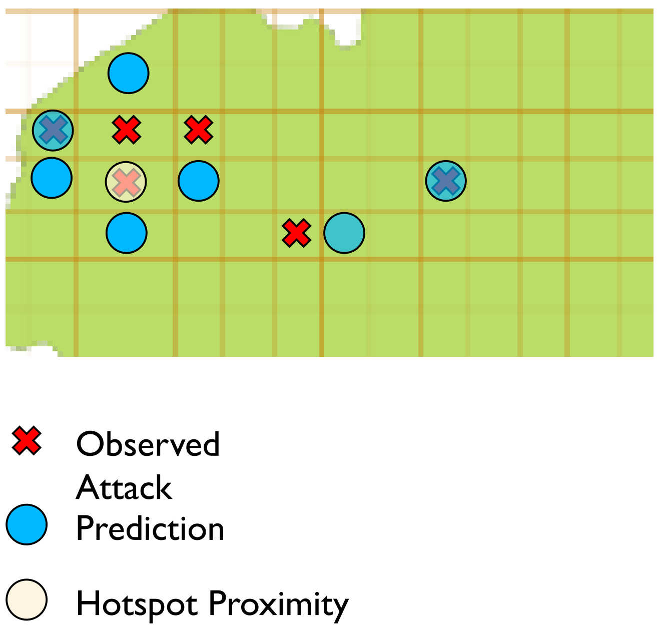
- Goal: Predict whether or not having poaching activity based on a set of input features
- Input features: Geospatial features + patrol coverage
- Label: Have poaching?



- Learn/Train the tree
  - Greedy decision tree learning
  - Greedily choose a feature and a threshold at a time
- Inference: traverse the tree

BoostIT: Spatially Aware Decision Tree

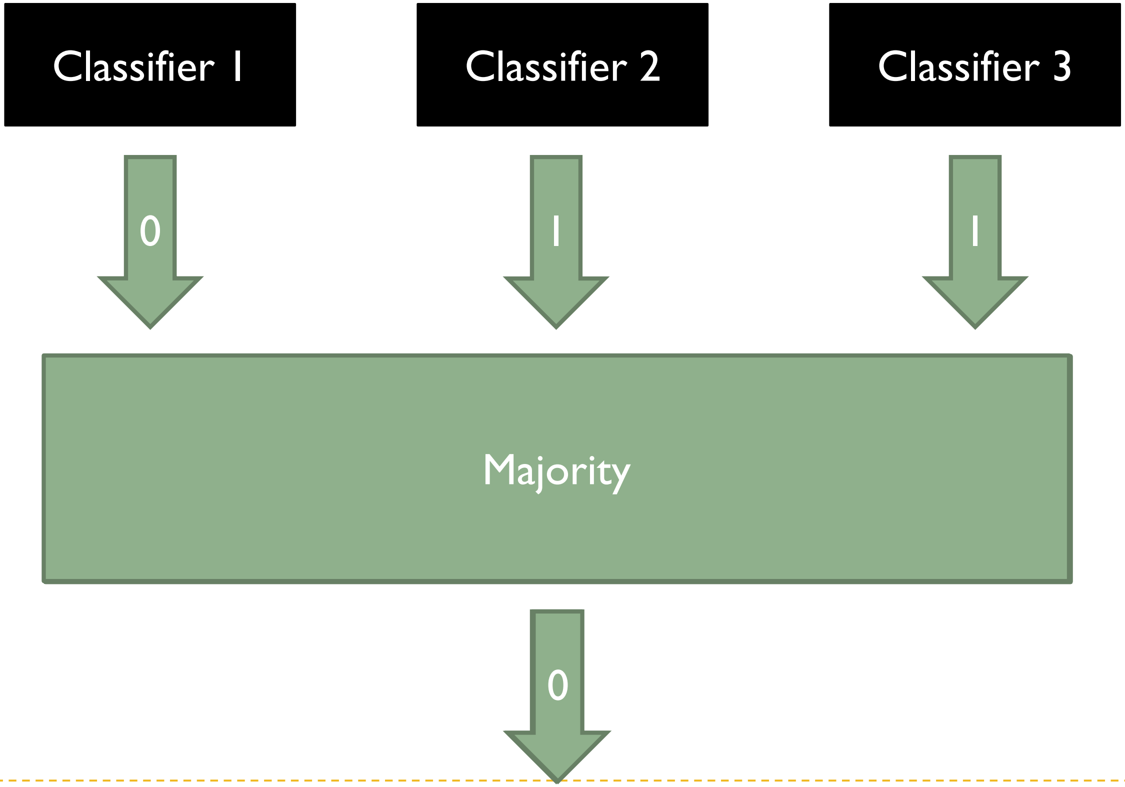
- Consider spatial correlations (hotspots)
  - Learn a decision tree
  - Compute predictions
  - Hotspot proximity computation
    - Feature = 1 if #positive neighbors $\geq \alpha$
  - Learn a new decision tree with hotspot proximity as a feature
  - Repeat until a stopping condition is reached



INTERCEPT: Build an Ensemble

- Set different stopping criteria for decision tree retraining
- Set different cost for false positive and false negative predictions

Is target **x** going to be attacked?



Evaluate INTERCEPT

- How to evaluate INTERCEPT?
  - Treat detections as labels for poaching activity but with uncertainty in negative label
  - Evaluate "predict the future" task using historical data

- Datasets
  - Trained: 2003-2014, Tested: 2015
  - Trained: 2003-2013, Tested: 2014

Evaluate INTERCEPT

- Metrics
  - INTERCEPT outputs binary label directly
  - Accuracy?
    - No. Class imbalance
  - Precision / Recall / F1?
    - Not enough. Does not consider the uncertainty in negative label
  - L&L score
    - Accounts for negative label uncertainty
    - Rewards recall heavily
    - Rewards selective models

$$L\&L(D, T_e) = \frac{recall^2}{\boxed{\Pr[f(T_e) = 1]}}$$

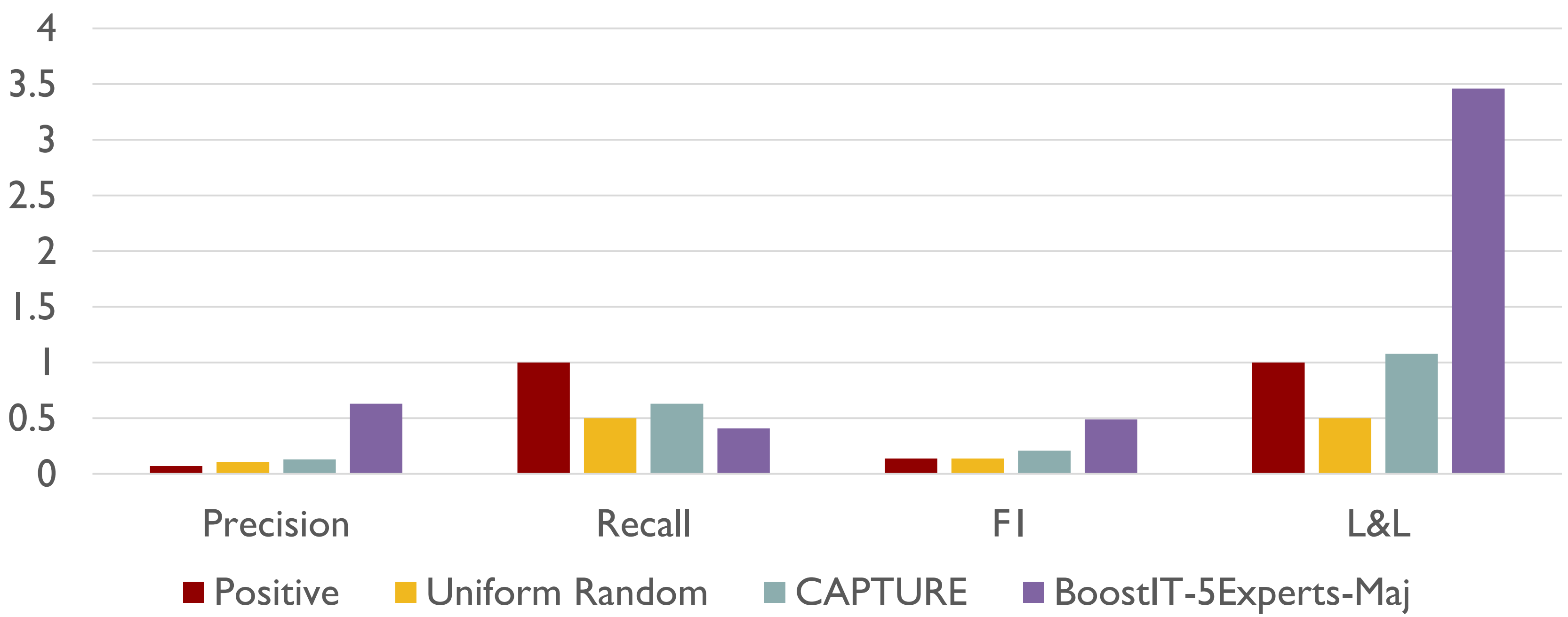
Probability of positive prediction

Quiz 2

- In the test set, 20% of the data points are actually positive. What is the L&L score of a perfect classifier? What is the L&L score of a classifier that predict every point to be positive?
- 1, 0
- 4, 0
- 5, 1
- 5, 0

Evaluate INTERCEPT

- Empirical Evaluation
  - 40 models w/ total of **192 model variations**
  - Best model: Decision tree ensemble with Standard decision tree + 2 BoostITs (α =2, 3) + 2 Decision Trees (FP cost = 0.6, 0.9)



Deploy INTERCEPT

- <u>Fast runtime</u> and <u>interpretability</u> led to its deployment

- Two 3x3 sq. km patrol areas
  - Infrequent patrols
  - Predicted hotspot
- Trespassing
  - 19 signs of litter, ashes, etc.
- Poached Animals
  - 1 poached elephant
- Snaring
  - 1 active snare
  - 1 cache of 10 antelope snares
  - <u>1 roll of elephant snares</u>

Historical Base Hit Rate	**Our Hit Rate**
Average: 0.73	<u>3</u>

Deploy INTERCEPT

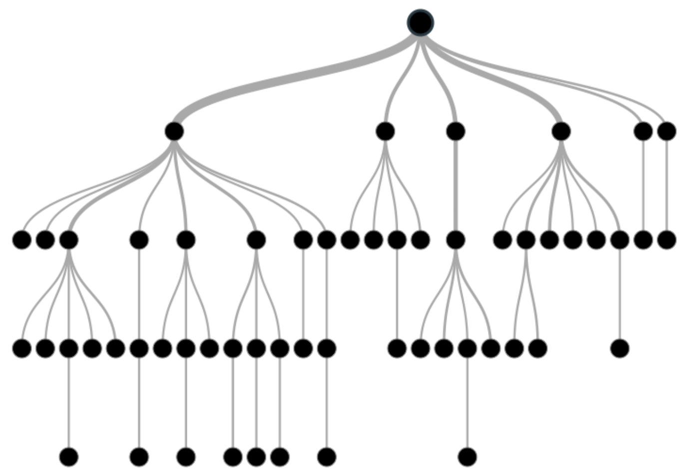
Outline

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

Classifier: Decision Tree

PROS

- High speed
- Learn global poachers behavior
- Learn nonlinearity in geo-spatial predictor



CONS

- No explicit temporal dimension
- No aspect for label uncertainty

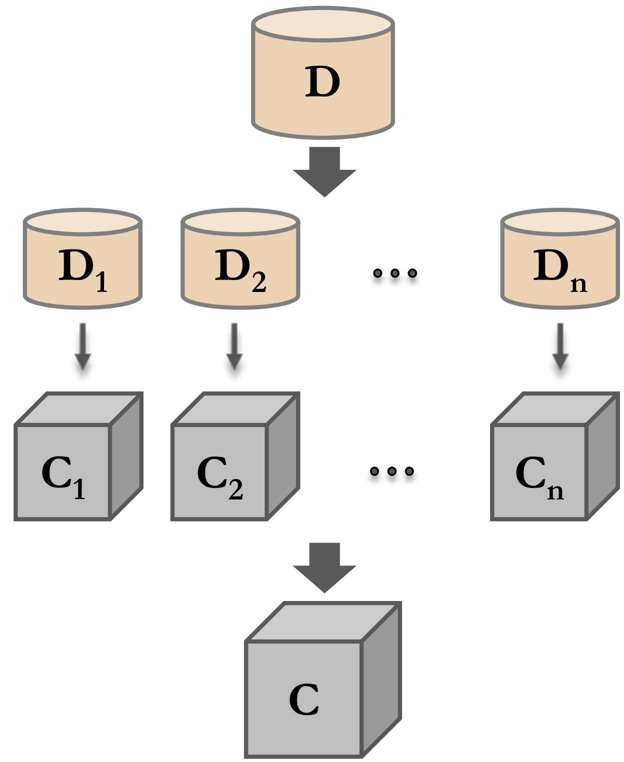
Bagging Ensemble: More Stable, Less Noisy due to Diversification

Original Training Data

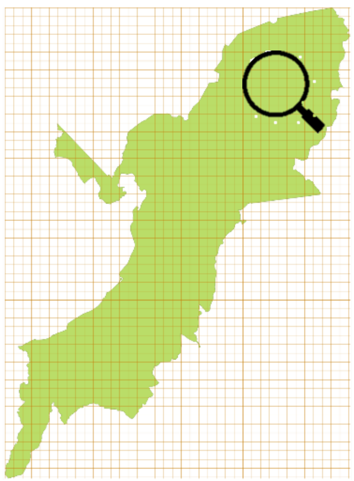
Create Multiple Datasets

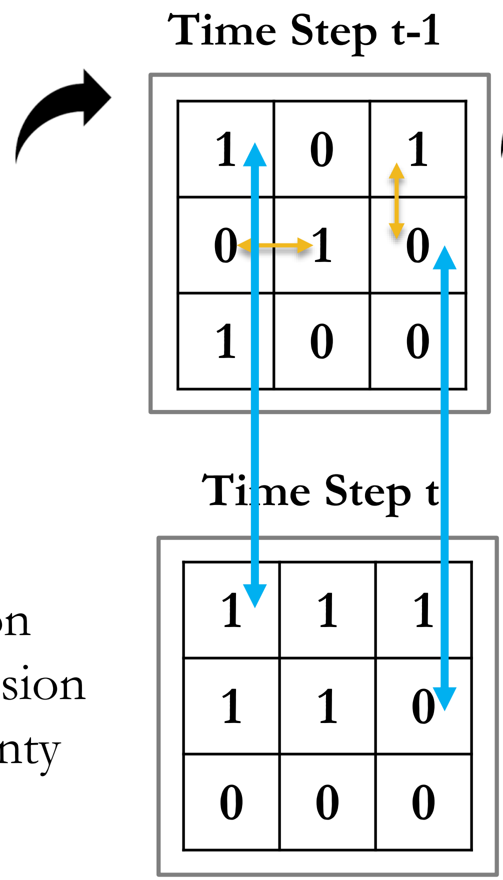
Build Multiple Classifier

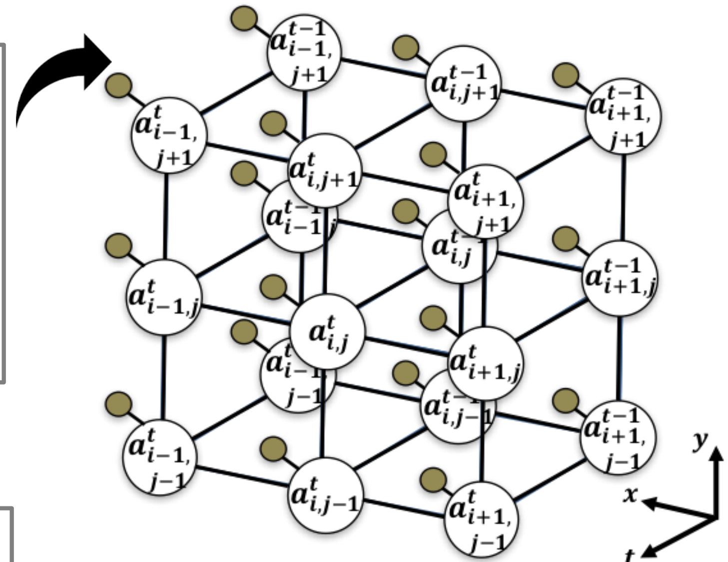
Combine Classifiers



Markov Random Field for Poaching







PROS

- Explicit spatial dimension
- Explicit temporal dimension
- Addresses label uncertainty

CONS

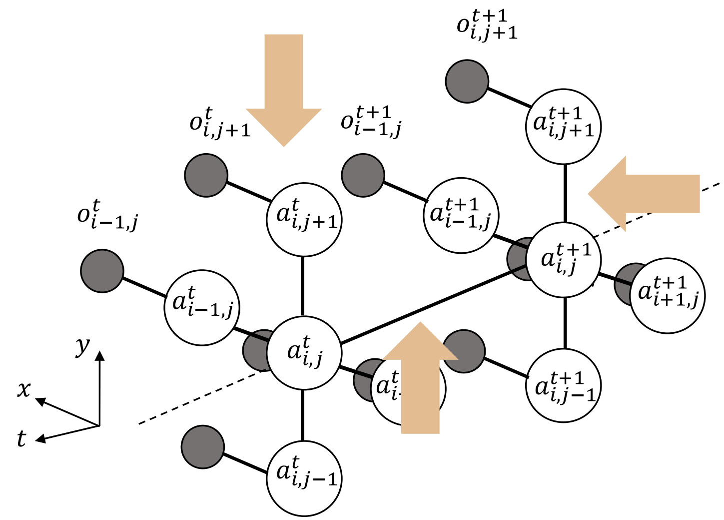
- Low speed
- Data greedy

Markov Random Field for Poaching

Observed Data Cliques

Spatial Cliques

Backward Temporal Cliques

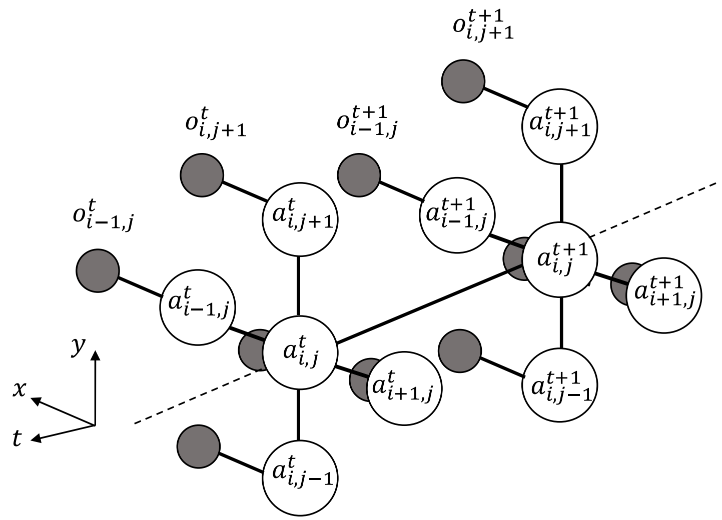


Markov Random Field for Poaching

Observed Data Cliques

$$\phi = \begin{bmatrix} P(o_i = 0|a_i = 0) & P(o_i = 0|a_i = 1) \\ P(o_i = 1|a_i = 0) & P(o_i = 1|a_i = 1) \end{bmatrix}$$

$$\phi = \begin{bmatrix} 1 & \frac{1}{1+e^{-\boldsymbol{\beta}.[c_i, p_i, 1]^T}} \\ 0 & \frac{e^{-\boldsymbol{\beta}.[c_i, p_i]^T}}{1+e^{-\boldsymbol{\beta}.[c_i, p_i, 1]^T}} \end{bmatrix}$$



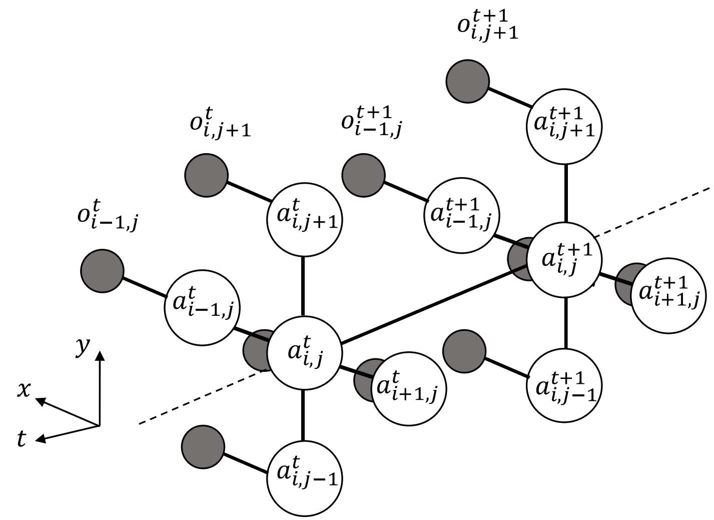
- Coverage, c_i
- Distance from patrol post, p_i

Markov Random Field for Poaching

Spatial Cliques

$$\psi = \begin{bmatrix} P(a_i = 0 | u_{N_i}^{t-1}) \\ P(a_i = 1 | u_{N_i}^{t-1}) \end{bmatrix}$$

$$\psi = \begin{bmatrix} \dfrac{1}{1 + e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}} \\ \dfrac{e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}}{1 + e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}} \end{bmatrix}$$



- Coverage, c_i
- Fraction of neighbors which are attacked, $u_{N_i}^{t-1}$
- All static features including distance from patrol posts, $\boldsymbol{X}$

Learn Parameters: EM

- **Goal:** $\boldsymbol{\theta}^* = \text{argmax } P(o|\theta)$
- **E-step,** $\theta = \{\alpha, \beta\}$**:**

$$Q(\theta|\theta^{(k)}) = \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{o},\theta^{(k)}}[logP(\boldsymbol{a},\boldsymbol{o}|\theta)]$$

$$= \sum_{a\epsilon\mathcal{A}} P\left(\boldsymbol{a}\,\middle|\boldsymbol{o},\theta^{(k)}\right).logP(\boldsymbol{a},\boldsymbol{o}|\theta)$$

- **M-step:**

$$\theta^{(k+1)} = \boldsymbol{argmax}_{\theta} Q(\theta|\theta^{(k)})$$

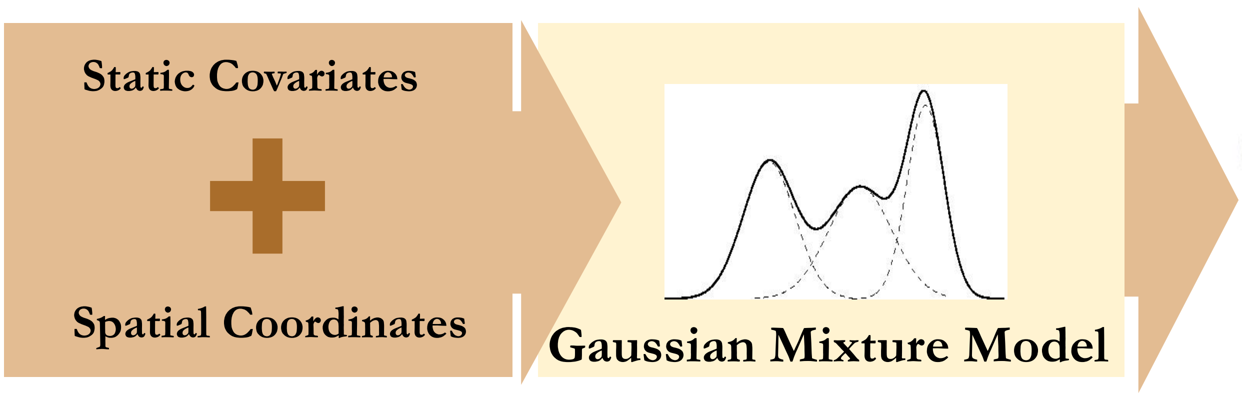
- **Update $\boldsymbol{\theta}$ until convergence:**

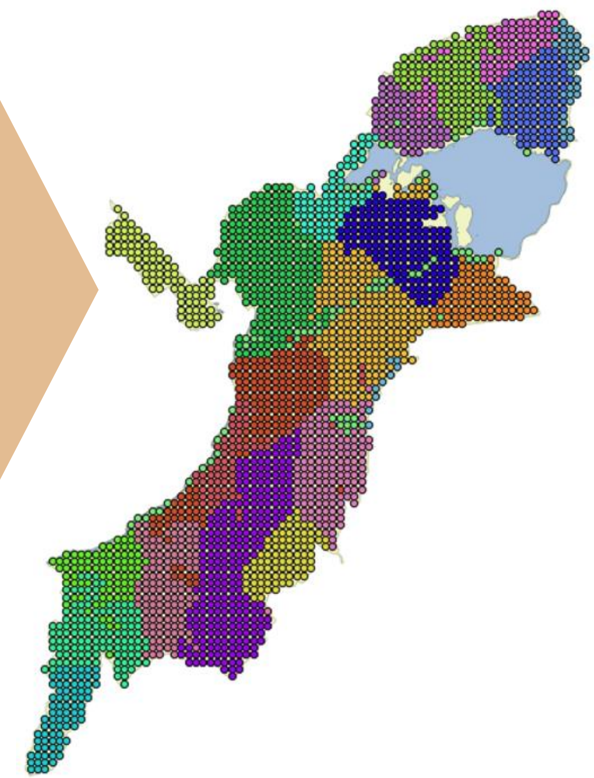
$$\theta^{(k)} \leftarrow \theta^{(k+1)}$$

Geo-clustering

Geo-clusters around patrol posts to learn:

- local poachers’ behavior
- Distinct parameters to expedite the local training of MRF

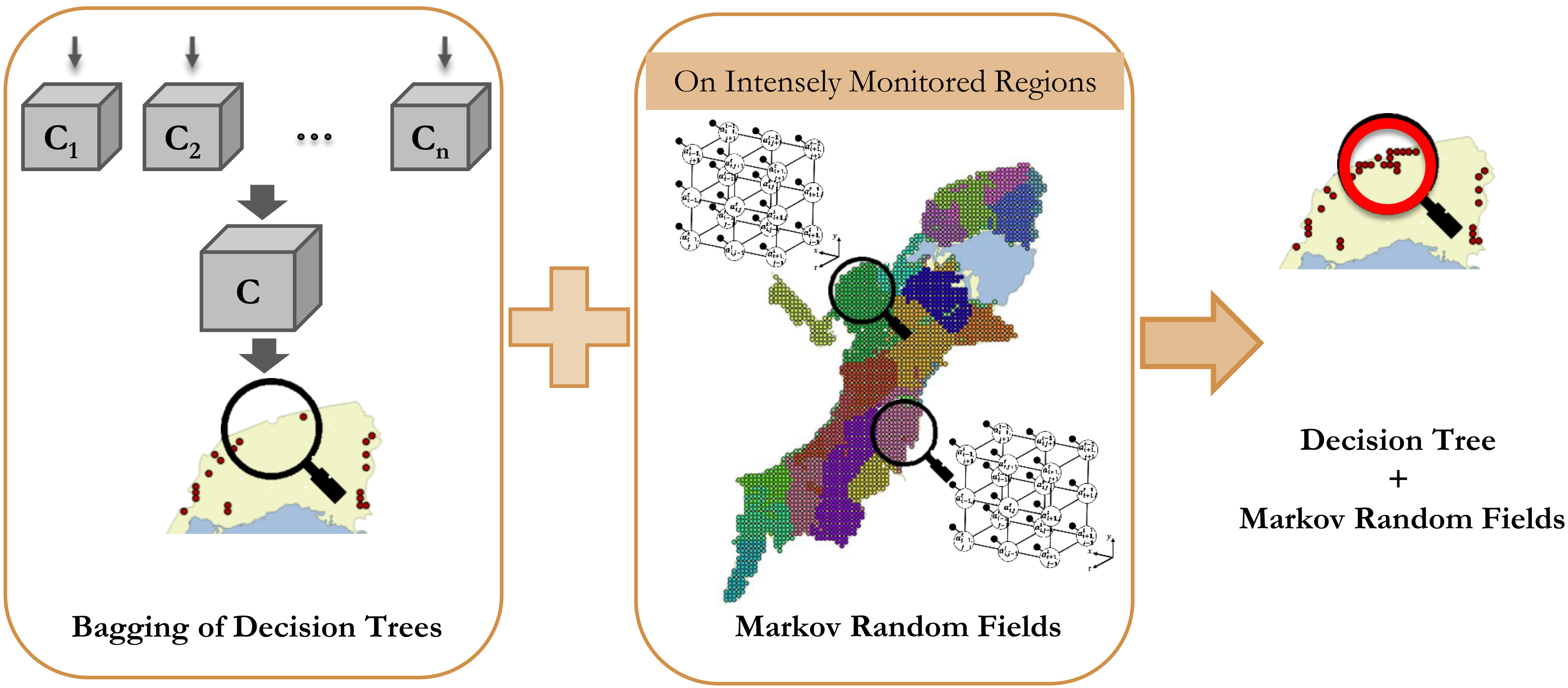




Hybridizing Bagging Model with Markov Random Fields

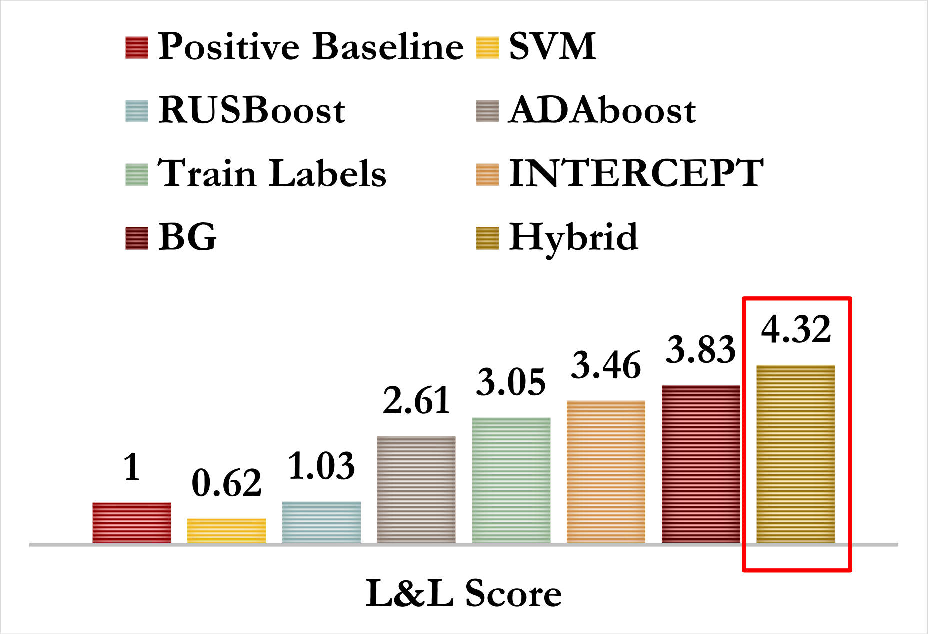
Boost by geo-clustered behaviorally inspired models:

- Improve the accuracy
- Learn local poachers' behavior; distinct parameters



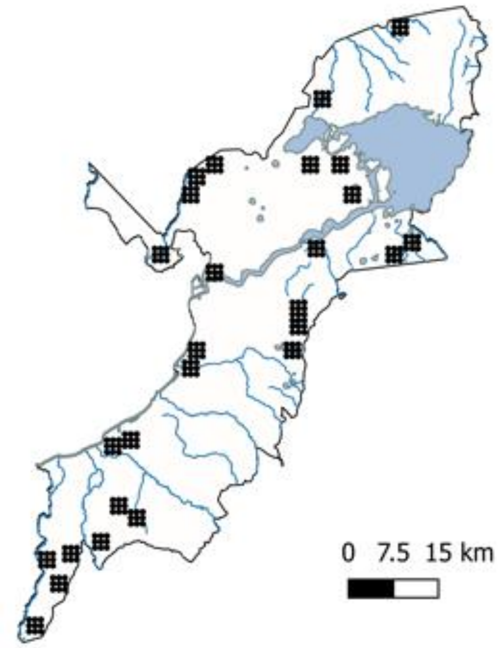
Empirical Evaluation

$$\text{L\&L Score} = \frac{Recall^2}{Probability\ of\ making\ a\ Positive\ Prediction}$$



Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each
- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas

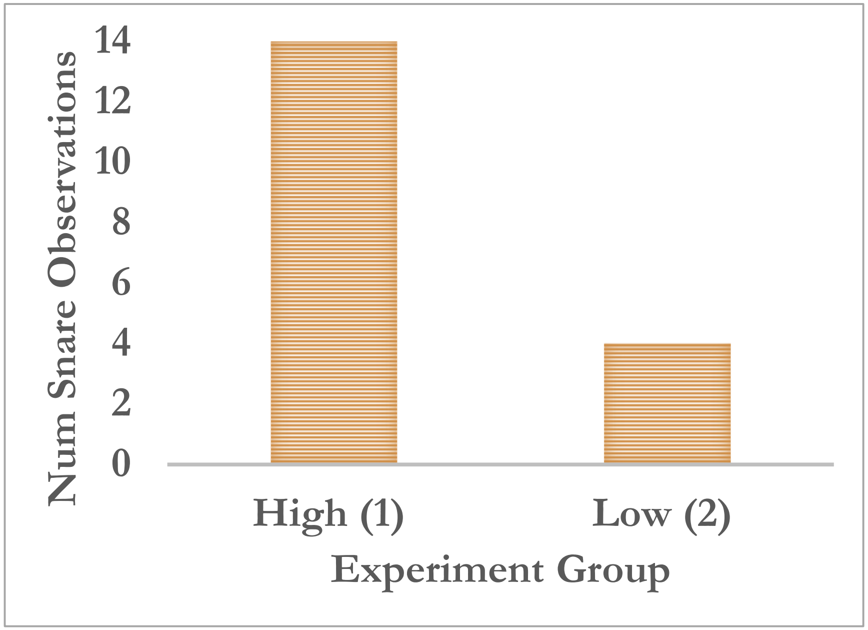


Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each

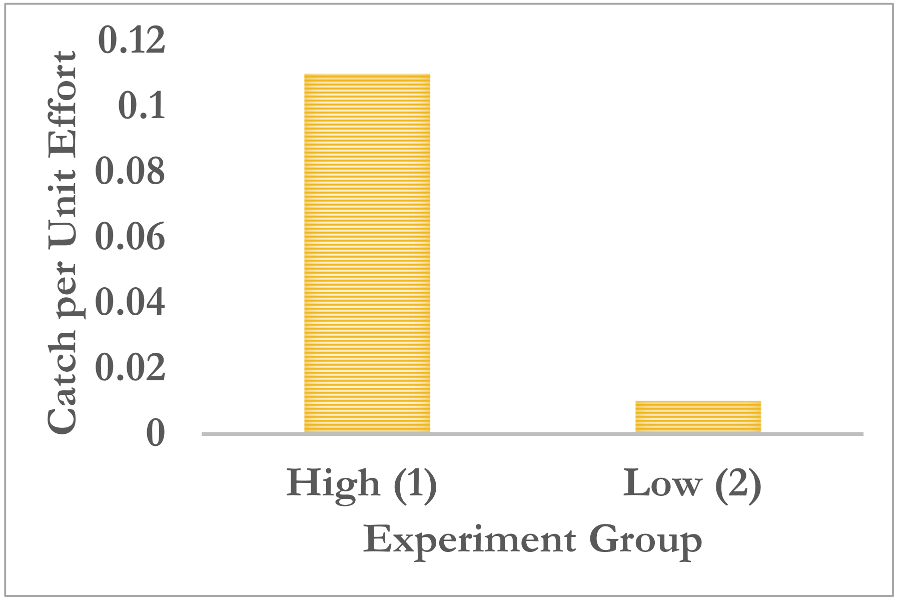
- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas

- 8 month, 452 km patrolled in total



Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each
- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas
- 8 month, 452 km patrolled in total
- Catch Per Unit Effort (CPUE)
  Unit Effort = km walked
  Historical CPUE: **0.03**



Real-world Deployment: 8 Months of Field Tests

- Statistical Significance

- Cohen's D

  Effect size: A standardized measure of the difference between two Means

$$d = \frac{experimental\ mean\ - control\ mean}{pooled\ standard\ deviation}$$

Interpretation*

0.2: Small

0.5: Medium (Visible to naked eye)

0.8: Large (Grossly perceptible)

High Group Mean (std)	Low Group Mean (std)	p-value	Cohen's d
0.12 (0.44)	0.01 (0.13)	$p<0.0001$	0.52

Outline

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

Urban Crime: Opportunistic Attack

In 2009
7,857,000 crime
$10,994,562,000

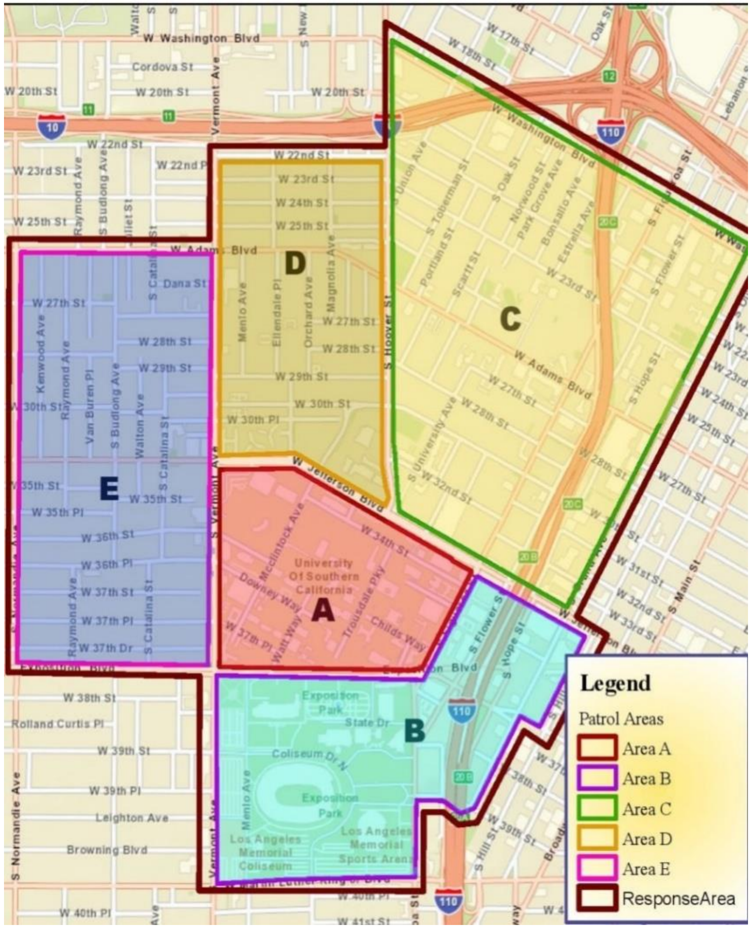
- Opportunistic adversaries (OA)
  - *Seek opportunities to commit attacks*
  - *Flexible in executing the plan*
    - Flexibility: Adapt plan with real time information

Predicting Opportunistic Crime

- Criminology based approach
  - General principles "crime predicts crime"
  - Have used many ML techniques – SVM, Regression, STL

- Ignores strategic interaction between defender and adversaries
  - Essential for planning patrols

- Opportunistic crime on the campus of University of Southern California (USC)
  - Department of Public Safety (DPS) allocates officers to 5 areas
  - Three patrol shifts per day
  - Criminals react *opportunistically*



Real-World Data

- Crime Report for 3 years

Area	CaseNbr	ccClass	DateOccured	TimeOccured
D	1200668	DISTURBANCE	02/16/12	9:00
C	1200669	CHILD	02/16/12	10:08
B	1200672	TRAFFIC	02/16/12	11:23
C	1200674	TRAFFIC	02/16/12	15:25
A	1200675	THEFT-PETTY	02/16/12	15:10
C	1200676	SERVICE	02/16/12	15:20
D	1200677	PROPERTY	02/16/12	18:30
C	1200679	DOMESTIC	02/16/12	17:30
A	1200680	THEFT-PETTY	02/16/12	19:15

Real-World Data

- Patrol Schedule for 3 years
  - Manually generated by domain experts

AREA	DAY		
A	**P3**	*1060*	Oosterhof
A	P23	1062	Hudson
B	P22	*1051*	Bouligny
C	P51	1187	Ramirrez
D	P30	*1067*	Guerra
E	P46	*1061*	Harris

Real-World Data

- Count number of crimes/officers in each shift in each area

Shift	A	B	C	D	E
1	1	1	2	2	2
2	1	1	1	2	1
3	2	1	1	3	1

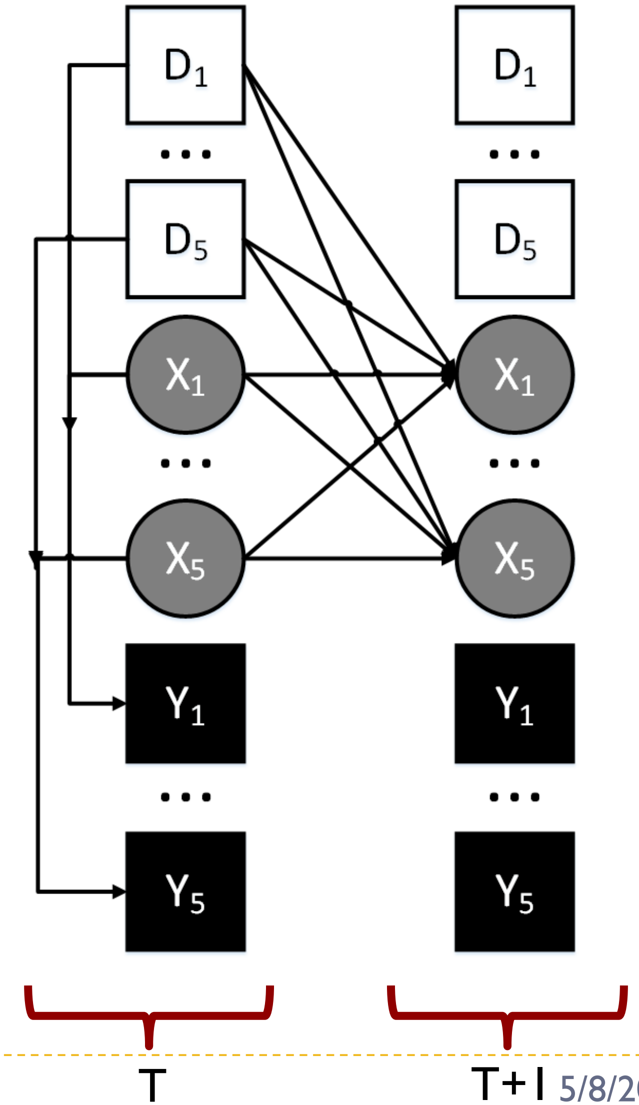
Table 1: Crime data for 3 shifts.

Shift	A	B	C	D	E
1	2	1	1	1	1
2	1	1	2	2	2
3	2	1	1	3	1

Table 2: Patrol data for 3 shifts.

Dynamic Bayesian Network Model

- DBN captures interaction between officers and criminals
  - D: Number of defenders (known)
  - X: Number of criminals (hidden)
  - Y: Number of crimes (known)
  - T: Step = Shift

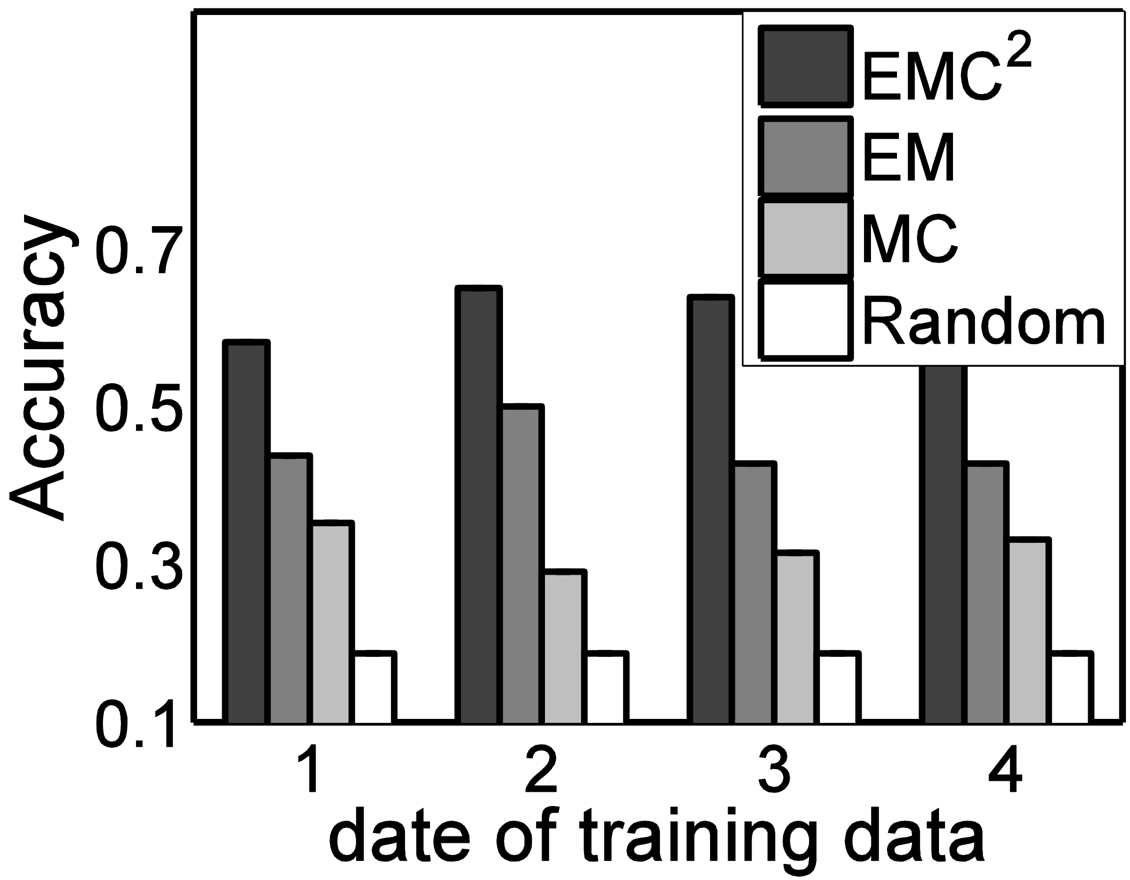


Dynamic Bayesian Network Model

- Learn/Train the model
  - Directly apply Expectation Maximization does not work:
    - Huge transition matrix and output matrix
    - Over-fitting
    - Exponential Runtime
  - EMC^2: Improve EM for this specific problem
    - Factorize output matrix
    - Pairwise transition matrix
    - Distributive law

Evaluate the Model

- Metric: Accuracy



Summary

- Predicting Poaching Activity
  - Model 1: CAPTURE
    - Dynamic Bayesian Network
  - Model 2: INTERCEPT
    - Spatially Aware Decision Tree Ensemble
  - Model 3: Hybrid Model
    - Gaussian Mixture Model × Decision Tree with Bagging + Markov Random Fields
- Predicting Urban Crime
  - Dynamic Bayesian Network

Quiz 3

- For the problem of identifying fraudulent firms, which statements of the following are true?
  - The dataset is unbalanced
  - The dataset only has positive labels
  - Decision tree-based approach can be a good fit
  - The dataset has entries with the same features but different labels