The Workshops of the Thirty-Second
AAAI Conference on Artificial Intelligence

Evaluation of Predictive Models for Wildlife Poaching Activity through Controlled Field Test in Uganda

**Shahrzad Gholami,[1] Benjamin Ford,[1] Debarun Kar,[1] Fei Fang,[2] Milind Tambe,[1]
Andrew Plumptre,[3] Margaret Driciru,[4] Fred Wanyama,[4] Aggrey Rwetsiba,[4]
Mustapha Nsubaga,[3] Joshua Mabonga[3]**
[1]University of Southern California,{sgholami, benjamif, dkar, tambe}@usc.edu,
[2]Carnegie Mellon University,feifang@cmu.edu,
[3]Wildlife Conservation Society,{aplumptre, mnsubuga, jmabonga}@wcs.org,
[4]Uganda Wildlife Authority, {margaret.driciru, fred.wanyama, aggrey.rwetsiba}@ugandawildlife.org

Abstract

Worldwide, conservation agencies employ rangers to protect conservation areas from poachers. However, agencies lack the manpower to have rangers effectively patrol these vast areas frequently. While past work modeled poachers behavior so as to aid rangers in planning future patrols, those models predictions were not validated by extensive field tests. We conducted two rounds of field tests in Ugandas Queen Elizabeth Protected Area to evaluate our proposed spatio-temporal model that predicts poaching threat levels. In the first round, a one-month field test was conducted to test the predictive power of the model and in the second round an eight-month test was conducted to evaluate the selectiveness power of the model. To our knowledge, this is the first time that a predictive model is evaluated through such an extensive field test in this domain. These field tests will be extended to another park in Uganda, Murchison Fall Protected Area. Once such models are evaluated in the field, they can be used to generate efficient and feasible patrol routes for the park rangers.

Introduction

Wildlife poaching continues to be a global problem as key species are hunted toward extinction. For example, the latest African census showed a 30% decline in elephant populations between 2007 and 2014 (Chase et al. 2016). Wildlife conservation areas have been established to protect these species from poachers, and these areas are protected by park rangers. These areas are vast, and rangers do not have sufficient resources to patrol everywhere intensively.

At many sites now, rangers patrol and collect data related to snares they confiscate, poachers they arrest, and other observations. Given rangers' resource constraints, patrol managers could benefit from tools that analyze these data and provide future poaching predictions. However, this domain presents unique challenges. First, this domain's real-world data are few, extremely noisy, and incomplete. To illustrate, one of rangers' primary patrol goals is to find wire snares, which are deployed by poachers to catch animals. However, these snares are usually well-hidden (e.g., in dense grass), and thus rangers may not find these snares and (incorrectly) label an area as not having any snares. Second, poaching activity changes over time, and machine learning models must account for this temporal component. Third, because poaching happens in the real world, there are mutual spatial and neighborhood effects that influence poaching activity. Finally, while field tests are crucial in determining a model's efficacy in the world, the difficulties involved in organizing and executing field tests often precludes them.

In this paper, we summarize our efforts for conducting field tests in Ugandas Queen Elizabeth Protected Area (QEPA) which were reported in details in (Kar et al. 2017) and (Gholami et al. 2017). Also, we discuss our attempts for extending of such field tests to other protected areas in Uganda.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Related Works

(Nguyen et al. 2016) introduced a two-layered temporal Bayesian Network predictive model (CAPTURE) that was also evaluated on real-world data from QEPA. CAPTURE, however, assumes one global set of parameters for all of QEPA which ignores local differences in poachers' behavior. While CAPTURE includes temporal elements, it does not include spatial components and thus cannot capture neighborhood specific phenomena. In contrast to CAPTURE, (Kar et al. 2017) presented a behavior model, INTERCEPT, based on an ensemble of decision trees and was demonstrated to outperform CAPTURE. While that model accounted for spatial correlations, it did not include a temporal component. In contrast to these predictive models, our latest model addresses both spatial and temporal components.

In game theory literature, learning adversary models has been mostly done based on simulated games where data is collected by human subject experiments in the laboratory (Gholami et al. 2016) rather than real world poachers. It is vital to validate predictive models in the real world, (Critchlow et al. 2016) conducted a controlled experiment where their goal, by selecting three areas for rangers to patrol, was to maximize the number of observations sighted per kilometer walked by the rangers. Their test successfully demonstrated a significant increase in illegal activity detection, but they did not provide comparable evaluation metrics for their predictive model. Also, our second field test was much larger in scale, involving 27 patrol posts compared to their 9 posts.

Wildlife Crime Dataset

This study's wildlife crime dataset is from two wildlife conservation parks in Uganda. There are several patrol posts situated across the parks from which Uganda Wildlife Authority rangers conduct patrols to apprehend poachers, remove any snares or traps, monitor wildlife, and record signs of illegal activity. Along with the amount of patrolling effort in each area, both datasets contain 14 years (2003-2016) of the type, location, and date of wildlife crime activities.

Rangers lack the manpower to patrol everywhere all the time, and thus illegal activity may be undetected in unpatrolled areas. Patrolling is an imperfect process, and there is considerable uncertainty in the dataset's negative data points (i.e., areas being labeled as having no illegal activity); rangers may patrol an area and label it as having no snares when, in fact, a snare was well-hidden and undetected. These factors contribute to the dataset's already large class imbalance. It is thus necessary to consider models that estimate hidden variables (e.g., whether an area has been attacked). We divide the parks into 1 square kilometer grid cells, and we refer to these cells as targets. Each target is associated with several static geospatial features such as terrain (e.g., slope), distance values (e.g., distance to border), and animal density. Each target is also associated with dynamic features such as how often an area has been patrolled (i.e., coverage) and observed illegal activities (e.g., snares).

Hybrid of Markov Random Fields and Bagging Ensemble

(Kar et al. 2017) proposed INTERCEPT, which is a model based on an ensemble of decision trees. This model was shown to predict poachers' attacks more effectively compared to the previous models. The first round of experiments was conducted based on this technique. To improve the predictive accuracy even more, the next generation of the predictive models which was a hybrid of Markov Random Fields and Bagging ensemble was developed in (Gholami et al. 2017).

Since the amount and regularity of data collected by rangers varies across regions of QEPA, predictive models perform differently in different regions. As such, the latter paper proposed using different models to predict over different regions; first, a Bagging ensemble model is used, and then predictions were improved in some regions using the spatio-temporal model with graphical modeling approaches, i.e, Markov Random Fields (MRF). A Bagging ensemble model or **B**ootstrap **agg**regation technique, called Bagging, is a type of ensemble learning which bags some weak learners, such as decision trees, on a dataset by generating many bootstrap duplicates of the dataset and learning decision trees on them. Each of the bootstrap duplicates are obtained by randomly choosing M observations out of M with replacement, where M denotes the training dataset size. Finally, the predicted response of the ensemble is computed by taking an average over predictions from its individual decision trees.

Capturing temporal trends requires a sufficient amount of data to be collected regularly across time steps for each target. Due to the large amount of missing inspections and uncertainty in the collected data, more complex models like MRF with hidden layer for latent poaching activity focuses on learning poaching activity only over regions that have been continually monitored in the past. The spatio-temporal model is designed to account for temporal and spatial trends in poaching activities. However, since learning those trends and capturing spatial effects are impacted by the variance in local poachers' behaviors, a geo-clustered model was examined which consists of multiple sets of local parameters throughout QEPA with spatial effects.

For geo-clustered models, for targets in the continually monitored subset (where temporally-aware models can be used practically), the MRF model's performance varied widely across geo-clusters according to the experiments. Thus, for each geo-cluster, if the average Catch Per Unit Effort (CPUE), is relatively large, the MRF model is used. In Conservation Biology, CPUE is an indirect measure of poaching activity abundance. A larger average CPUE for each cluster corresponds to more frequent poaching activity and thus more data for that cluster. Consequently, using more complex spatio-temporal models in those clusters becomes more reasonable. To compute CPUE, effort corresponds to the amount of coverage (i.e., 1 unit = 1 km walked) in a given target, and catch corresponds to the number of observations. Hence, boosting is done selectively according to the average CPUE value; some clusters may not be boosted by MRF, and so only Bagging ensemble model is used for making predictions on them. Experiments on historical data show that selecting 15% of the geo-clusters with the highest average CPUE results in the best performance for the entire hybrid model (Gholami et al. 2017).

Experiment A: One-month Field Test

In the first round of the experiments, INTERCPET (or in short INT) was tested to evaluate the predictive ability of the model. After development and evaluation of the model on historical data, it was deployed to the field. Based on the predictions, two patrol areas were chosen for QEPA rangers to patrol for one month. These areas were selected (approximately 9 square km each) such that they were (1) predicted to have multiple attacks and (2) previously infrequently patrolled as rangers did not previously consider these as important as other areas (and thus are good areas to test the model predictions). After providing the rangers with GPS coordinates of particular points in these areas, they patrolled these areas on foot and utilized their expert knowledge to determine where exactly in these areas they were most likely to find snares and other signs of illegal human activity (e.g., salt licks, watering holes). On each patrol, in addition to their other duties, rangers recorded their observations of animal sightings (i.e., 21 animals were sighted in one month) and illegal human activity. The key findings are demonstrated in Tables 1 and 2 and a selection of photos

Week#	Illegal Activity	Count
2	Trespassing	19
3	Active Snares	1
	Plant Harvesting	1
4	Poached Elephants	1
	Elephant Snare Roll	1
	Antelope Snares	10
	Fish Roasting Racks	2

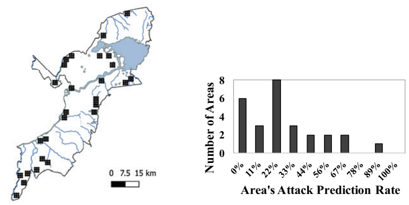
Table 1: Real World Patrol Results: Illegal Activity

(a) (b)

Figure 1: Elephant snare roll found by rangers directed by INT Photo credit: Uganda Wildlife Authority ranger

Crime Type	INT	Average	Percentile
AnimalCom	**1**	0.16	89%
AnimalNoncom	**3**	0.73	91%
Fishing	**1**	0.73	79%
PlantNoncom	**1**	0.46	76%
Trespassing	**19**	0.20	100%
Total	**25**	2.28	

Table 2: Base Rate Comparison: Hits per Month



(a) Patrolled areas (b) Prediction rates

Figure 2: Patrol Area Statistics

taken by park rangers are shown in Figures 1(a) and 1(b). The most noteworthy findings of these patrols are those related to elephant poaching; rangers, unfortunately, found one poached elephant with its tusks removed. However, this result demonstrates that poachers find this area, predicted by the model, attractive for poaching but they had missed to visit them thoroughly in the past. On a more positive note, the model's predictions led rangers to find many snares before they caught any animals: one large roll of elephant snares, one active wire snare, and one cache of ten antelope snares. In fact, the machine learning model predictions assisted rangers' efforts in potentially saving the lives of multiple animals including elephants.

In addition to wildlife signs, which represent areas of interest to poachers, the findings of trespassing (e.g., litter, ashes) are significant as these represent areas of the park where humans were able to enter illegally and leave without being detected; if we can continue to patrol areas where poachers are visiting, rangers will eventually encounter the poachers themselves.

So as to provide additional context for these results, set of base rates are presented in Table 2. These base rates, computed in and around the proposed patrol areas, correspond to the average number of observed crimes per month from 2003-2015. Animal commercial (AnimalCom) crimes correspond to elephant, buffalo, and hippopotamus poaching; animal noncommercial (AnimalNoncom) corresponds to all other poaching and poaching via snares; and plant noncommercial (PlantNoncom) corresponds to illegal harvesting of non-timber forest products (e.g., honey). The percentile rank corresponds to the number of months where the deployed patrols recorded more observations than in the historical data. For animal noncommercial crime, there was an average of 0.73 attacks observed monthly; for the deployed patrols, there were 3 separate observations (such as a roll of elephant snares), and in 91% of the months from 2003-2015, 2 or fewer observations were recorded.

Experiment B: Eight-month Field Test

As the efforts for improving model performance continued, it was important to extend the field test to evaluate the predictive power and selectiveness of the new models which were shown to be outperforming the previous ones. Experiment A was conducted in collaboration with the Wildlife Conservation Society (WCS) and the Uganda Wildlife Authority (UWA) and showed promising improvements over previous patrolling regimes. Due to the difficulty of organizing such a field test, its implications were limited: only two 9-sq km areas (18 sq km) of QEPA were patrolled by rangers over a month. Because of its success, however, WCS and UWA graciously agreed to a larger scale, controlled experiment: also in 9 sq km areas, but rangers patrolled 27 of these areas (243 sq km, spread across QEPA) over eight months; this is the largest to-date field test of ML-based predictive models in this domain. We show the areas in Figure 2(a). Note that rangers patrolled these areas in addition to other areas of QEPA as part of their normal duties.

This experiment's goal was to determine the selectiveness of the hybrid model's snare attack predictions mentioned earlier in this paper: does the model correctly predict both where there are and are not snare attacks? We define attack prediction rate as the proportion of targets (a 1 km by 1 km cell) in a patrol area (3 by 3 cells) that are predicted to be attacked. We considered two experiment groups that corresponded to the hybrid model's attack prediction rates from November 2016 - March 2017: High (group 1) and Low (group 2). Areas that had an attack prediction rate of 50% or greater were considered to be in a high area (group 1); areas with less than a 50% rate were in group 2. For example, if the model predicted five out of nine targets to be attacked in an area, that area was in group 1. Due to the importance of QEPA for elephant conservation, we do not show which areas belong to which experiment group in Figure 2(a) so that we do not provide data to ivory poachers. To start, we exhaustively generated all patrol areas such that (1)

Table 3: Patrol Area Group Memberships

Group	All Patrol Areas	Final Patrol Areas
High	50 (9%)	5 (19%)
Low	494 (91%)	22 (81%)

Table 4: Field Test Results: Observations

Group	Counts (%)	Mean (std)	Effort	CPUE
High	15 (79%)	3 (5.20)	130	0.12
Low	4 (21%)	0.18 (0.50)	322	0.01

each patrol area was 3x3 sq km, (2) no point in the patrol area was more than 5 km away from the nearest ranger patrol post, and (3) no patrol area was patrolled too frequently or infrequently in past years (to ensure that the training data associated with all areas was of similar quality); in all, 544 areas were generated across QEPA. Then, using the model's attack predictions, each area was assigned to an experiment group. Because we were not able to test all 544 areas, we selected a subset such that no two areas overlapped with each other and no more than two areas were selected for each patrol post (due to manpower constraints). In total, 5 areas in group 1 and 22 areas in group 2 were chosen. Note that this composition arose due to the preponderance of group 2 areas (see Table 3).

We provide a breakdown of the areas' exact attack prediction rates in Figure 2(b); areas with rates below 56% (5/9) were in group 2, and for example, there were 8 areas in group 2 with a rate of 22% (2/9). Finally, when we provided patrols to the rangers, experiment group memberships were hidden to prevent effects where knowledge of predicted poaching activity would influence their patrolling patterns and detection rates. The field test data we received was in the same format as the historical data. However, because rangers needed to physically walk to these patrol areas, we received additional data that we have omitted from this analysis; observations made outside of a designated patrol area were not counted. Because we only predicted where snaring activity would occur, we have also omitted other observation types made during the experiment (e.g., illegal cattle grazing). We present results from this eight-month field test in Table 4. To provide additional context for these results, QEPA's park-wide historical CPUE is also computed (from November 2015 to March 2016): 0.04.

Areas with a high attack prediction rate had significantly more snare sightings than areas with low attack prediction rates (15 vs 4). This is despite there being far fewer group 1 areas than group 2 areas (5 vs 22); on average, group 1 areas had 3 snare observations whereas group 2 areas had 0.18 observations. It is worth noting the large standard deviation for the mean observation counts; the standard deviation of 5.2, for the mean of 3, signifies that not all areas had snare observations. Indeed, two out of five areas in group 1 had snare observations. However, this also applies to group 2's areas: only 3 out of 22 areas had snare observations.

Catch per Unit Effort (CPUE) results are presented in Table 4. When accounting for differences in areas' effort, group 1 areas had a CPUE that was over ten times that of group 2 areas. Moreover, when compared to QEPA's park-wide historical CPUE of 0.04, it is clear that the new hybrid model successfully differentiated between areas of high and low snaring activity. The results of this large-scale field test, the first of its kind for Machine Learning models in this domain, demonstrated that the Machine Learning model's superior predictive performance in the laboratory extends to the real world.

Experiment C: Ongoing Field Test

To evaluate the capability of the models to be generalized to other conservation areas, we aim to test it in Murchison Fall park in Uganda. Basically, two datasets (i.e., Queen Elizabeth and Murchison Fall) have similar set of covariates and features, however, the amount of poaching activities detected in them are different. So we are extending the field tests in collaboration with Wildlife Conservation Society and the Uganda Wildlife Authority. While the target regions are selected from three groups of low, medium and high in terms of likelihood of poaching activity occurrence, they are transferred to the park rangers without group labels. The field tests are planned to begin in late November, 2017.

Acknowledgments: This research was supported by MURI grant W911NF-11-1-0332, NSF grant with Cornell University 72954-10598. We are grateful to the Wildlife Conservation Society and the Uganda Wildlife Authority for supporting data collection in QEPA.

References

Chase, M. J.; Schlossberg, S.; Griffin, C. R.; Bouché, P. J.; Djene, S. W.; Elkan, P. W.; Ferreira, S.; Grossman, F.; Kohi, E. M.; Landen, K.; et al. 2016. Continent-wide survey reveals massive decline in african savannah elephants. *PeerJ* 4:e2354.

Critchlow, R.; Plumptre, A. J.; Alidria, B.; Nsubuga, M.; Driciru, M.; Rwetsiba, A.; Wanyama, F.; and Beale, C. M. 2016. Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected monitoring data. *Conservation Letters*.

Gholami, S.; Wilder, B.; Brown, M.; Thomas, D.; Sintov, N.; and Tambe, M. 2016. Divide to defend: Collusive security games. In *GameSec*, 272–293. Springer.

Gholami, S.; Ford, B.; Fang, F.; Plumptre, A.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Nsubaga, M.; and Mabonga, J. 2017. Taking it for a test drive: a hybrid spatio-temporal model for wildlife poaching prediction evaluated through a controlled field test. In *Proceedings of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases, ECML PKDD*.

Kar, D.; Ford, B.; Gholami, S.; Fang, F.; Plumptre, A.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Nsubaga, M.; et al. 2017. Cloudy with a chance of poaching: Adversary behavior modeling and forecasting with real-world poaching data. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 159–167. AAMAS.

Nguyen, T. H.; Sinha, A.; Gholami, S.; Plumptre, A.; Joppa, L.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Critchlow, R.; et al. 2016. Capture: A new predictive anti-poaching tool for wildlife protection. 767–775. AAMAS.