Carnegie Mellon University Africa
Certificate I: Understanding AI and Machine Learning

Course AIML02: AI and Machine Learning in Africa

Module 2: Application Case Studies
Lecture 6: Conservation

Welcome to Lecture 4 of Module 2. We continue to look at applications of AI and machine learning in an African context, focusing here on a case study in conservation.

For this, we will refer to an article by Xu et al., published in 2019, with the title

"Ahead of Poachers: Illegal Wildlife Poaching Prediction and Patrol Planning Under Uncertainty with Field Test Evaluations"

The article is quite long so, for the purposes of this case study, we will focus only the sections on predicting poaching attacks, while recognizing that these predictions can then be used to plan effective patrols for poachers.

As well as being a long article, it also contains an extensive amount of technical detail. Our goal here, as with other case studies, is to outline the approach, previewing the techniques that will be covered in later courses as a way of emphasizing the relevance of AI and machine learning in Africa, and motivating the subsequent study of these techniques. Consequently, we won't go into the detailed operation of these techniques here, but, as with the other case studies, we will provide outline explanations of these techniques where appropriate to help familiarize you with them and prepare you to study them later on.

In this lecture, we highlight the negative impact of poaching, and the challenges of applying machine learning in conservation and anti-poaching efforts.

To provide some context for the approach taken in the target article, we look briefly at some of the other approaches that have been used in the past.

We then walk through the approach used in the target article for predictive modelling with uncertain data. This forms the core of the lecture.

We will finish up by summarizing what we have covered and identifying the articles that you should read to consolidate what you have learned.

As before, we will encounter several AI and machine learning techniques which you have not yet studied in detail, some of which we introduced briefly in the first course, AIML01, and others that we introduced earlier in this course. As we do so, we will flag where they were introduced and provide a little more detail if appropriate. However, as we already said, we treat them mainly as a preview of material to follow in later courses where these techniques are covered in greater depth.

We have three learning objectives, so that, after studying the material covered in this lecture, you should be able to do the following.

1. Identify the impact of poaching and explain the importance of wildlife conservation.

2. Identify the challenges of modelling the behavior of poachers.

3. Explain how machine learning can be used to predict poaching activity and plan patrols in protected parks.

4. Explain the importance of addressing prediction uncertainty in machine learning models.

5. Identify different machine learning techniques that can be deployed to minimize prediction uncertainty and outline the key principles of each one.

Slide 1 Welcome to Lecture 6 of Module 2.

We continue to look at applications of AI and machine learning in an African context, focusing here on a case study in conservation.

For this, we will refer to an article by Xu et al., published in 2019, with the title

"Ahead of Poachers: Illegal Wildlife Poaching Prediction and Patrol Planning Under Uncertainty with Field Test Evaluations"

Slide 2 Illegal wildlife poaching threatens biodiversity, ecological balance, and ecotourism.

Many species are being poached to near-extinction:

Elephants for their ivory tusks
Rhino for horn
and tigers for their skin

Many other animals, such as wild pigs and apes, are hunted for their meat.

Slide 3 The article by Xu et al. opens by highlighting the potential for the use in artificial intelligence to combat poaching.

"Artificial intelligence frameworks can significantly advance wildlife protection efforts by learning from past poaching activity to prescribe actionable recommendations to park managers."

Slide 4 Rangers require deep knowledge of the behavior of poachers if they are to be successful in assessing risk of poaching attacks and planning patrols.

Slide 5 Learning the behavior of poachers is a challenging machine learning problem because:

- Wildlife crime datasets are very imbalanced with up to 99.6% negative labels, indicating an absence of poaching activity, and just 0.4% positive labels, indicating poaching activity.

- The challenge of having a majority of negative labels is made worse because the negative labels are often unreliable due to the difficulty of detecting well-hidden traps

- Furthermore, historical poaching observations are not collected thoroughly and uniformly. This results in a biased dataset.

- And, finally, poaching patterns vary with region.

Slide 6 AI & machine leaning often adopts what is known as a data-to-deployment pipeline

Data is input to a predictive machine learning algorithm. This generates predictions.

These predictions are input to a prescriptive process that makes decisions.

These decisions are then acted upon.

The Protection Assistant for Wildlife Security, also known as PAWS, has used this approach in the past.

Slide 7 The approach taken by Xu et al. (2019) in the target article takes a similar approach, but with tighter integration of the components in the pipeline.

It involves a two-stage approach.

The first stage uses machine learning to develop a predictive model of relative poaching risk, using data from past patrol and information on the protected area.

The second stage uses these predictions with a game theoretic model to determine future poaching activities and recommend patrol routes to rangers.

We focus on the first stage in this case study.

Slide 8 The ultimate goal of conservation is to save the lives of as many animals as possible

There are three elements to the strategy adopted in the target paper to achieve this goal.

Slide 9 First, it means maximizing the number of snares removed.

To do this, we need to predict where these snares are likely to be.

Slide 10 Second, to achieve this – and this is the key to the entire approach – it means quantifying the uncertainty of the predictions of poaching risk.

Slide 11 Recall this diagram from the first lecture in Module 1 of AIML01 Artificial Intelligence – Past, Present, and Future, showing the three primary areas of AI:

Machine learning, including deep learning;

Knowledge-based systems, which emulate the problem-solving and reasoning skills of a human expert, and

Probabilistic Bayesian learning, which deal with uncertainty and, using probabilistic inference, draw conclusions that are the most likely to be correct when solving problem.

The target article uses techniques from the third area. Specifically, it uses Gaussian processes.

We met Gaussian processes already in Lecture 3 of this module in the case study on agriculture, specifically concerning the generation of a spatial map of disease, given sparse samples. The approach here is similar. We will meet Gaussian processes again in Module 3, Lecture 4, and we will provide a short tutorial introduction to them later in this lecture

Slide 12 The predictions, and their associated uncertainty, allows an optimal patrol strategy to be determined.

The target article provided details on how to do create such a strategy but we will omit these from this case study.

Slide 13 Large-scale field tests were carried out in three conservation areas, two in Uganda,

Murchison Falls National Park and Queen Elizabeth National Park,

and one in Cambodia,

Srepok Wildlife Sanctuary.

These protected areas cover cover 5000 sq. km, 2500 sq. km and 4300 sq. km, respectively.

We will focus on the parks in Uganda in the remainder of this lecture.
These parks are critically important for ecotourism and conservation, and provide habitat for elephants, giraffes, hippos, and lions.

Park rangers combat poaching by patrolling the parks, using GPS trackers to record their observations, while confiscating animal traps, rescuing live animals caught in snares, and arresting any poachers when possible.

Slide 14 The dataset used to model poaching activity comprises static geospatial features and dynamic poaching & patrol data.

The static geospatial data include

terrain features, such as elevation maps, rivers, and forest cover;

landscape features, such as roads, park boundaries, local villages, and patrol posts

and ecological features such as animal density.

Slide 15 The dynamic observed poaching data & historical patrol data include the GPS location, date, time of each observation, patrol leader, and method of transport,

along with the observations: animals or humans spotted; indications of illegal or poaching activity such as campsites, cut trees, firearms, bullet cartridges, snares, or slain animals.

Each observation is classified as either a poaching or a non-poaching activities.

Slide 16 Finally, a discrete map of the protected area is constructed.
Each cell in the map is 1 kilometer x 1 kilometer,

each cell comprising a sequence of *T* three-month time intervals,

recording *k* features, both static geospatial features and one time-variant feature identifying the amount of patrol effort in the previous time step.

Slide 17 In summary, the dataset D comprises a data matrix X and an observation vector of labels.

Slide 18 The data matrix X has T 3-month time intervals times N locations (or cells), and a feature vector of k features for each one.

One of these features is the time-variant patrol effort *c.*

We will see why this is important in a moment.

Slide 19 The dataset also comprises an observation vector y that labels each cell n during time t as either 1 or 0, indicating "poaching activity" or "no poaching activity", respectively.

Slide 20 We come now to the key issue in this case study: the uncertainty of the data

The success with which rangers detect poaching activity in a given 1 km x 1 km cell depends on the amount of effort they exert in patrolling that cell; equivalently how far they extend their search.

Positive instances, that is, cells at a given time that are labelled poaching, are reliable, irrespective of the amount of patrol effort.

Why? If rangers find a snare in a cell, poaching occurred with certainty.

However, negative instances have different levels of uncertainty that depend on the patrol effort c_t,n exerted in cell n during time t.

Slide 21 The percentage of poaching activity detected within each cell increases proportionally with the patrol effort exerted within that cell

To quantify the uncertainty of negative instances, we can apply a threshold *theta* of patrol effort

A patrol effort $c_{t,n} \geq$ *theta* is more reliable, that is, less uncertain

A patrol effort $c_{t,n} <$ *theta is* less reliable, that is, more uncertain

Slide 22 The predictive model is based on an existing approach: the imperfect observation aWare Ensemble (iWare-E)

This approach uses a bagging ensemble of weak classifiers (either decision trees or support vector machines) as weak learners to create a strong classifier.

That's a heavily-loaded sentence so let's unpack it a little.

Slide 23 We already met support vector machine classifiers on several occasions: in Module 2, Lecture 3 Statistical Machine Learning in AIML01 and in Module 2, Lecture 1 Healthcare of this course.

Slide 24 Decision trees are an alternative form of machine learning classifier that works by testing feature values against successive decision criteria.

Each branch represents an outcome of the test, and each leaf node (or terminal node) determines the designated class.

Slide 25 Next, the core idea of ensemble learning is to combine the results of several weak learners, or classifies, to generate a strong learner or classifier.

Bagging is one way of generating this ensemble classifier.

In this case, one generates randomly selected sub-sets of the training set and uses each one to train an individual (weak) classifier.

Slide 26 The results of these individual classifiers, here shown as decision trees, are then aggregated: the class receiving the most votes is selected as the true class.

Slide 27 Returning to the iWare-E model, three weak learners are shown.

Each is trained on a different subset of the dataset, keeping all positive instances (or samples), and removing less reliable or uncertain negative instances (or samples). The negative samples that are removed depend on the values of the thresholds theta_i for the patrol effort $c_{t,n}$

In this diagram, the subsets of the training set that are less than the value of the thresholds, and are thus removed from the training set, are shown in white. Those shown by the shaded bars are included.

Thus, each weak classifier is trained on a different dataset.

Slide 28 The approach in the target paper introduces three improvements to iWare-E.

The first improvement allows optimal classifier weights to be computed, allowing the classifiers to be trained using the entire dataset rather than a subset.

This ensures that classifiers trained on data with high patrol effort can predict on the entire data and not just on data with equal or higher patrol effort as is the case with the conventional iWare-E model.

The second improvement is the manner in which the patrol effort thresholds theta_i are selected. Here, they are selected on the basis of patrol effort percentiles and characteristics of the data instead of using equally distanced values as is the case with the original iWare-E model.

This makes sure that imbalance in the dataset is accounted for.

The final improvement is to explicitly account for prediction uncertainty by using Gaussian process classifiers as the weak learners.

We met Gaussian processes already in Lecture 3 of this module in the case study on agriculture and we will briefly meet them again in Module 3, Lecture 4.

The explicit inclusion on uncertainty plays an important part in planning patrol routes.

Slide 29 [This is an optional slide]

Let's take this a little further, just to give you a flavor of the underlying technique. Again, it's important to be aware that this is a preview, and you don't need to understand this, or any other technique, in detail. The goal in these case studies is to motivate the use of AI and machine learning in applications in Africa and to provide a window onto the AI techniques that are commonly used.

A Gaussian process is a probability distribution over possible **functions** that fit a set of data points, as opposed to the more usual case of a probability distribution over possible outcomes or events.

This allows us to determine a set of **functions** that fits the data

by sampling the multi-dimensional Gaussian distribution to determine candidate function,

use the mean of these to provide reliable predictions,

and get an indication of the uncertainty (or confidence) of these predictions.

The uncertainty will be low close to the training data and higher further away, exactly as you would expect, and as we saw in the spatial map showing the incidence of disease in Module 2, Lecture 3.

Note that this is a non-parametric approach to machine learning. This simply means that we need to keep the training data which we use when making the predictions. This has an impact on the time complexity and space complexity of Gaussian process algorithms and associated software.

In contrast, parametric approaches to machine learning use the training data to determine the parameters that characterize the process we are modelling, and once these parameters have been determined, the training data can be discarded.

Slide 30 To evaluate the predictive models, the authors of the target article studied predictive performance across different datasets with three different weak learners

Support Vector Machines
Decision Trees
Gaussian Processes

both with and without the iWare-E model.

Slide 31 We show here the results for the two parks in Uganda.

The results show the prediction in a given year after training on the previous three years.

The iWare-E model increases Area Under Curve by 0.1 on average when compared to bagging weak learner.

The Gaussian Process iWare-E model, GPB, in the last column outperforms the other models in 5 of the 8 tests.

The results are even better when the Cambodia park tests are considered.

Slide 32 The heatmaps in red at the top show the predicted probabilities of detecting poaching activity in the Murchison Falls National Park based on the Gaussian Process iWare-E model.

The uncertainty associated with these predictions is shown in green heatmaps underneath.

The three pairs of maps show how the predictions and the uncertainty changes as the patrol effort increases from 0.0 km to 2.0 km.

Inspection of these heatmaps reveals important insights.

The predicted probability of detection generally increases as patrol effort increases. However, the increase is not uniform.

Some cells have near-zero probability despite increased patrol effort, indicating that there is almost no likelihood of attack in those cells.

Slide 33 Regions where the predicted probability increases slowly indicate that the likelihood of detection will not become very strong. Consequently, it would be wasteful to allocate excessive patrolling resources in those areas.

Several areas with historically high levels of patrol effort, such as the northwest section, are predicted to have low risk of detecting poaching activity. This means that rangers should instead focus on patrolling elsewhere.

Slide 34 Note that that uncertainty increases as we make predictions with high levels of patrol effort. This is because there is less historical data with higher levels of patrol effort.

To summarize:

1. Ecosystems are put in peril and endangered species are driven to extinction by illegal wildlife hunting and the limited resources to enforce wildlife conservation.

2. PAWS is a machine learning pipeline created as a data-driven strategy to identify regions in protected areas that are at high risk of poaching and compute the best patrol routes to prevent poaching.

3. The use of Gaussian processes as weak learners improves the existing iWare-E model in PAWS by modelling the uncertainty of the predictions of detecting poaching attacks.

Here is the article on which this lecture is based.

Xu, L., Gholami, S., McCarthy, S., Dilkina, B., Plumptre, A., Tambe, M., . . . Enyel, E. (2020). Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations (short version). In 2020 IEEE 36th international conference on data engineering (ICDE). IEEE. https://arxiv.org/abs/1711.06323

Here are the references cited.

Yang, R., Ford, B., Tambe, M., & Lemieux, A. (2014). Adaptive resource allocation for wildlife protection against illegal poachers. In Proceedings of the 2014 international conference on autonomous agents and multi-agent systems (pp. 453–460). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems. https://dl.acm.org/doi/10.5555/2615731.2615805