World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:11, No:3, 2017

Ensemble Methods in Machine Learning: An Algorithmic Approach to Derive Distinctive Behaviors of Criminal Activity Applied to the Poaching Domain

Authors : Zachary Blanks, Solomon Sonya

Abstract : Poaching presents a serious threat to endangered animal species, environment conservations, and human life. Additionally, some poaching activity has even been linked to supplying funds to support terrorist networks elsewhere around the world. Consequently, agencies dedicated to protecting wildlife habitats have a near intractable task of adequately patrolling an entire area (spanning several thousand kilometers) given limited resources, funds, and personnel at their disposal. Thus, agencies need predictive tools that are both high-performing and easily implementable by the user to help in learning how the significant features (e.g. animal population densities, topography, behavior patterns of the criminals within the area, etc) interact with each other in hopes of abating poaching. This research develops a classification model using machine learning algorithms to aid in forecasting future attacks that is both easy to train and performs well when compared to other models. In this research, we demonstrate how data imputation methods (specifically predictive mean matching, gradient boosting, and random forest multiple imputation) can be applied to analyze data and create significant predictions across a varied data set. Specifically, we apply these methods to improve the accuracy of adopted prediction models (Logistic Regression, Support Vector Machine, etc). Finally, we assess the performance of the model and the accuracy of our data imputation methods by learning on a real-world data set constituting four years of imputed data and testing on one year of non-imputed data. This paper provides three main contributions. First, we extend work done by the Teamcore and CREATE (Center for Risk and Economic Analysis of Terrorism Events) research group at the University of Southern California (USC) working in conjunction with the Department of Homeland Security to apply game theory and machine learning algorithms to develop more efficient ways of reducing poaching. This research introduces ensemble methods (Random Forests and Stochastic Gradient Boosting) and applies it to real-world poaching data gathered from the Ugandan rain forest park rangers. Next, we consider the effect of data imputation on both the performance of various algorithms and the general accuracy of the method itself when applied to a dependent variable where a large number of observations are missing. Third, we provide an alternate approach to predict the probability of observing poaching both by season and by month. The results from this research are very promising. We conclude that by using Stochastic Gradient Boosting to predict observations for non-commercial poaching by season, we are able to produce statistically equivalent results while being orders of magnitude faster in computation time and complexity. Additionally, when predicting potential poaching incidents by individual month vice entire seasons, boosting techniques produce a mean area under the curve increase of approximately 3% relative to previous prediction schedules by entire seasons.

Keywords : ensemble methods, imputation, machine learning, random forests, statistical analysis, stochastic gradient boosting, wildlife protection
Conference Title : ICCSS 2017 : International Conference on Computational and Statistical Sciences
Conference Location : Paris, France
Conference Dates : March 29-30, 2017