Project Proposal

Machine Learning for Public Policy

Cheng Yee Lim, Rodrigo Valdez-Ortiz, J.Adrián Sánchez

April, 2017

Project goals

Problem definition

Importance

Impact

Who cares

The public, the police department

Who will take action based on your work

The Charlotte-Mecklenburg Police Department (CMPD) will apply the machine learning algorithms we have devised to improve on their existing Early Intervention System. By detecting police officers at risk of adverse events prior to them committing the action, the CMPD can direct these officers to retraining programmes to extract them from high-stress environments, remind them the ethos of the CMPD and improve on their capacity to deal with such situations.

What are the policy goals you care about?

Leveraging on a more data-driven approach in the Early Intervention System allows the department to better identify police officers at risk, increase internal accountability and reduce inappropriate uses of force.

A tangible policy goal we would like achieve through this project is a safer public-facing CMPD.

By reducing the number of adverse actions committed by the CMPD, we hope to ease the existing tensions and foster stronger trust between the community and the CMPD. A tangible policy goals that we would like to achieve through this project is a reduction of instances of unjustified deadly shootings or instances of racial profiling.

Data

Analysis

We are looking to identify a risk ranking score that determines the probability that police officers will engage in adverse event/misconduct behavior[1]. As explained before, our predictors (features) will be **** from a

[1]We refer to adverse events or misconduct behavior for the following criteria: (1) officer's improper use of force, (2) citizen or officer injury or accident and (3) any sustained serious complaint from a citizen or colleague.

time-period that will predict a risk ranking score which could be used, at different thresholds, to obtain a dummy output variable $0, 1$ for a top-N list, that indicate if an officer is likely to misbehave in a period of time.

As the previous section describes, our label (dependent variable) will be based on the

For this endeavor, we will look to implement three different binary classification models: Logistic Regression, Support Vector Machine (SVM's) and Random Forests. We expect to estimate the probability assigned to class C_i from the three proposed models with distinct parameters settings.

What actions will this enable or improve?

The objective of police departments is to train and retain the very highest quality police force. As budget constraints exists, police officers need to train and mentor those that needed the most. In this sense, the objective is to better identify police officers that are riskier to misconduct. By the provision of a score (probability) to each police officer, we could rank police officers by the risk of adverse behavior controlling for precision-recall tradeoffs that can give more control and flexibility to police departments to make interventions such as training, counselling or mentorship that can improve their quality services to the society. As there are limited resources to intervene all the police force, the ranking allows to the police department to adjust the risk threshold depending on their resources and their capacity to intervene police officers.

Evaluation

We will train our data for a fixed period of time, 2013-2014, to test in 2015. As we expect that our class of interest is the minority class, accuracy and performance on the majority class are not the right metrics to optimize. For this reason, for assessing the classification performance, we will use precision (true negative rate), the proportion of actual positives among the predicted positives.[2]. We will evaluate the process with a 3-fold cross validation in order to compare the models.

Policy recommendations:

What kind of recommendations do you hope to give to policymakers based on this analysis/project. ## How will you validate whether what you are proposing will have the desired impact?

[2]Flach, Peter. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, 2012