CS 238 Project Proposal

Group 88
Kaushal Alate, Nicholas Lai, Kadar Qian
Due: 10/05/18

Proposal:

For our project, we would like to solve the problem of detecting whether a mobile payment transaction is legal or fraudulent.

First, the decision to be made by the model is a classification of a mobile payment transaction as legal or fraudulent. We would try to model this problem based on a Bayesian network and perhaps a POMDP if we find some temporal approach in the weeks to come.

Second, there exists uncertainty in classifying these transactions as valid or invalid. It is often difficult to determine whether a transaction was actually executed by the user or if it was fraudulent. These sources of uncertainty arise because fraudulent transactions may individually look like valid transactions. However, by training a model on a set of multiple transactions using features such as the transaction type, the amount taken out, and the amounts before and after, we may be able to predict with more confidence whether a transaction is legal or fraudulent. For more information about the information available for each individual transaction in the data, please find a link to the training dataset we would like to use:
https://www.kaggle.com/ntnu-testimon/paysim1/home.

Third, we will talk further about the objectives of doing such a project: As mobile payments become increasingly popular, it is fundamental to protect people's financial safety. We believe that if this model works, it can be applied to a whole emerging industry, helping users safely and quickly execute mobile transactions.

Fourth, we need a way to measure how well our model detects fraudulent activity. We plan to split the data into training data that our model can use to learn its parameters and testing data that we can use to measure predictive accuracy. We look forward to learning more about Bayesian networks and POMDPs in class so that we can refine and solidify our plan to measure our model's success at identifying fraudulent transactions.

Mobile payments security (and payments security in general) is an important issue but privacy concerns mean that data is not publicly available. The dataset we are using was synthetically produced by a simulator named PaySim and was produced using a month's worth of real transaction data from a mobile payments company in an African country. It "resembles the normal operation of transactions and injects malicious behaviour to later evaluate the performance of fraud detection methods", as the dataset's "Overview" page notes.

We are excited to begin applying techniques from CS 238 to investigate this important area of security and would appreciate any feedback on our approach!