Capstone 1 Project Proposal

11/20/2017

Ginny Jie Zhu

Problem description

Credit card fraud detection

With ecommerce and online transactions increasing each day, credit card fraud has become a real concern for consumers as well as financial companies. According to Forbes news(1), card fraud losses actually hit $21.84 billion worldwide in 2015, which was truly alarming.

So the question I am going solve here is: based on multiple features of transactions dataset, we are to predict what are the potential frauds and what are not.

Clients

The real-world clients could be those financial companies like Chase, Capital One, American Express, Discover etc.

For the company themselves, the early prediction of potential frauds will be able to prevent unnecessary financial losses by allowing them to take appropriate actions.

For the customers, the detection features could be built into credit card apps, so that when a potential fraud is detected, the customers could be alerted of the situation by texts or calls, and therefore they could respond and communicate with card companies to solve the issues together.

After all, who doesn't like a reliable model/method that helps guard your bank accounts and prevent unnecessary financial losses?

Approach

Very generally, I'm going to experiment with several classification algorithms, the baseline classifiers could be naive bayes, random forest, SVM, and literature also suggested frequent itemset mining(2). With tuning techniques built on these base classifiers, or

combining several classifiers together, I'm planning on finding out one algorithm with best prediction performance.

Data:

https://www.kaggle.com/dalpozz/creditcardfraud

It's in the format of .csv tabular data. The data is unbalanced, containing numerical values out of PCA transformation result.

In total, it has 28 principal components, with additional columns that have the following information:

Time	the seconds elapsed between each transaction and the first transaction in the dataset.	Numeric
Amount	Transaction amount	Numeric
Class	The actual classification classes	Numeric

Final deliverables:

The final deliverables will be a project paper, a slide/prezi presentation deck and relevant code in an organized github repository.

References:

(1) https://www.forbes.com/sites/rogeraitken/2016/10/26/us-card-fraud-losses-could-exceed-12bn-by-2020/#53aa8baed243

(2) K.R., Seeja & Zareapoor, Masoumeh. (2014). FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining. TheScientificWorldJournal. 2014. 252797. 10.1155/2014/252797.