Machine Learning Engineer Nanodegree

Capstone Proposal

Xueqiu Feng

3.5.2018

Using Machine Learning to Detect Credit Card Fraud

Domain Background

Wikipedia defines fraud as: *deliberate deception to secure unfair or unlawful gain, or to deprive a victim of a legal right*. Fraud or deliberate deception is a skill that every species has already mastered to perfection through evolution, which is especially true for human being. In the modern era, with the development of new technologies, new forms of fraud have also been invented. Although fraud appears to be rarely happening, it can result in an huge amount of loss. Therefore, fraud detection is a focal point for insurance, financial, retail and tele-communication sectors. It also has drawed attention from the academical research *(Pozzolo et al. 2015)*, where machine learning algorithms are applied to tackle this problem.

Problem Statement

Credit card is almost a necessity to survive in the modern society. However, it associates with risk exposure because of physical or informational theft. Whereas a stolen card can be reported and frozen immediately to prevent unauthorised transactions, a compromised account can be abused without notice untill receiving bill statement, where before the frudulent use the account information can be hold for an arbitrary time, making the source even harder to trace. Because of the huge amount of transactions, it is nearly impossible to check them one by one manuelly, which will also inevitablly result in inacceptable delay of transactions. Therefore it would be very meaningful to build a system which can automatically detect dubious transactions and freeze it for further inspection.

Datasets and Inputs

The dataset I plan to use is the kaggle dataset Credit Card Fraud Detection (https://www.kaggle.com/mlg-ulb/creditcardfraud), whose official description is as followed:

The datasets contains transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features $V_1, V_2, \ldots V_{28}$ are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed

between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-senstive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.

The dataset has been collected and analysed during a research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be (http://mlg.ulb.ac.be)) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection. More details on current and past projects on related topics are available on http://mlg.ulb.ac.be/BruFence (http://mlg.ulb.ac.be/BruFence) and http://mlg.ulb.ac.be/ARTML (http://mlg.ulb.ac.be/ARTML)

Solution Statement

The creditcard fraud detection problem can be easily formulated as a supervised machine learning problem, with the binary labels *fraud* or *no fraud*, i.e. binary classification. Therefore almost all the possible supervised machine learning algorithms can be applied. Where the AUC can be very relevant in terms of determining the quality of the classifier because the dataset is strongly imbalanced.

Neural network will not be considered because of the long training time and the difficulty of selecting the best architecture. Moreover, my goal is to use the conventional machine learning algorithms to come as close as possible to be competitive against neural network, which will also be my benchmark model. At the end the classifiers will be compared such that the optimal one, depending on my own definition of optimality, will be chosen to tackle this particular problem.

Benchmark Model

As mentioned before, I plan to benchmark my model against a neural network, which is trained by Currie32 on Kaggle (https://www.kaggle.com/currie32/predicting-fraud-with-tensorflow).

Evaluation Metrics

Because this dataset is highly imbalanced, F_{beta} score will be used in order to put more emphasis on recall over precision. Moreover, area under the ROC is also of relevance to tackle the problem of imbalanced dataset.

Project Design

Similar to the previous projects, the first step of this project would be data exploration. I plan to use some simple decriptive statistics to obtain an intuition about the dataset. After that, the dataset will be pre-processed with missing value deletion and rescaling. Then data visualization techniques will be used to find out which features are relevant or irrelevant in predicting fraud. Moreover, data visualization would also serve the purpose of determining whether clustering algorithms would be suitable or not.

In the second step, different classifiers will be trained to identify fraud. Then they will be compared using the evaluation metrics mentioned above.

The last step would be benchmarking my best model against the neural network model by Currie32 on Kaggle (https://www.kaggle.com/currie32/predicting-fraud-with-tensorflow).

Bibliography

Pozzolo, A. D., O. Caelen, R. A. Johnson, and G. Bontempi. 2015. “Calibrating Probability with Undersampling for Unbalanced Classification.” In *2015 IEEE Symposium Series on Computational Intelligence*, 159–66. doi:10.1109/SSCI.2015.33.