Detecting Electricity Theft

Patrick GLAUNER
Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg
patrick.glauner@uni.lu

Introduction

Electrical power grids are the backbone of today's society. Losses during generation and distribution cause major problems, including financial losses to electricity providers and a decrease of stability and reliability. They can be classified into technical losses and non-technical losses. Technical losses are naturally occurring and mainly include losses to power dissipation in electrical components, such as in generators, transformers and transmission lines due to internal electrical resistance. They are possible to detect and control given a knowledge of the quantities of loads.
Non-technical losses (NTL) faced by electricity providers include, but are not limited to, electricity theft by rewiring or manipulating meters. Other types include faulty meters and errors in meter readings and billing. There are different estimates of the financial losses caused by NTLs and they can range up to 40% of the total electricity distributed in countries such as Brazil, India, Malaysia or Lebanon. They are also of relevance in developed countries, for example estimates of NTLs in the US range from USD 1-6 billion.

Fig. 1: That is what electricity theft looks[a].

In order to detect NTLs, inspections of customers are carried out, based on predictions whether there may be a NTL at a customer. The inspection results are then used in the learning of algorithms in order to improve predictions. However, carrying out inspections is expensive, as it requires physical presence of technicians. It is therefore important to make accurate predictions in order to reduce the number of false positives.
This project is in cooperation with CHOICE Technologies Holding Sàrl, which provides real data and domain expertise.

[a] http://extra.globo.com/incoming/13321838-a74-9d3/w448/Eletrotraficante-Rio-das-Pedras.jpg

Challenges

Detecting NTLs is challenging and includes the following factors:

- Wide range of possible causes of NTLs, such as different fraudulent types of customers.
- Imbalance of the data, meaning that there are significantly more regular customers than customers with NTLs.
- Inspection labels may be false-negative, because technicians got bribed or threatened.
- The inspection sample is biased and does not represent the population of all customers.
- Varying levels of NTLs in different cities/countries.
- Not only poor people steal, even the government does.

Related work

NTL detection can be treated as a special case of fraud detection. It highlights two approaches as key methods to detect fraudulent behavior in credit card fraud, computer intrusion and telecommunications fraud:

1. Expert systems that represent domain knowledge in order to make decisions typically using hand-crafted rules.
2. Data mining or machine learning techniques that employ statistics to learn patterns from sample data in order to make decisions for future unseen data.

Both approaches have their justification and neither is generally better or worse than the other one in artificial intelligence.
Most methods in the literature use supervised learning, please find a comprehensive discussion in our first paper.
It must be noted that most NTL detection methods are supervised. Anomaly detection – a superclass of NTL – is generally challenging to learn in a supervised manner.

NTL detection

The data used in this paper is from an electricity provider in Brazil. It consists of three parts: (i) 700K customer data, such as location, type, etc., (ii) 31M monthly consumption data from January 2011 to January 2015 such as consumption in kWh, date of meter reading and number of days between meter readings and (iii) 400K inspection data such as presence of fraud or irregularity, type of NTL and inspection notes.
Most inspections do not find NTLs, making the classes highly imbalanced. In order for the models to be applied to other regions or countries, they must be assessed on different NTL proportions. Each sample contains 100K inspection results.

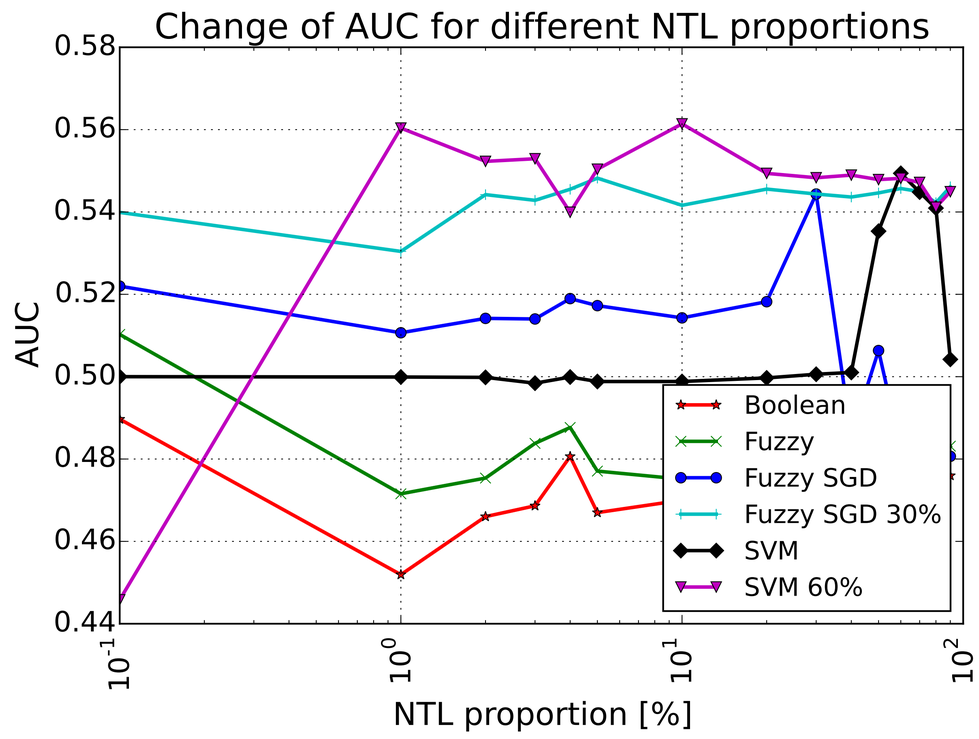

Fig. 2: Comparison of optimized classifiers tested on different NTL proportions.

Next steps

We are planning to evaluate unsupervised methods, in particular deep learning, in order to detect NTL more accurately by finding hidden correlations in the data. Furthermore, we are planning to use other features in our models, such as the location, latent features and hand-crafted features that put consumption patterns in relation to similar customers and the past. We are also planning to investigate cost-based optimization in order to maximize the total electricity recovered through inspections. Also, we are planning to make our implementations faster and more scalable using Apache Spark.

Conclusions

The initial Boolean and fuzzy models perform worse than random guessing and are therefore not suitable for real data, as they trigger too many inspections while not many of them will lead to NTL detection. Optimized fuzzy and SVM models trained on 30% and 60% NTL proportion, respectively, result in significantly greater AUC scores. However, both perform very differently, as the optimized fuzzy system is more conservative in NTL prediction. In contrast, the optimized SVM is more optimistic, leading also to a higher FPR. In general, neither can be named better than the other one, as picking the appropriate model from these two is subject to business decisions.

First paper

P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni and D. Duarte: Large-Scale Detection of Non-Technical Losses in Imbalanced Data Sets. arXiv:1602.08350. 2016. Submitted to The Seventh IEEE Conference on Innovative Smart Grid Technologies (ISGT 2016).

About

Patrick GLAUNER is a PhD student in machine learning at the University of Luxembourg in collaboration with CHOICE Technologies Holding, under the supervision of Dr. Radu STATE. He graduated as valedictorian in computer science from Karlsruhe University of Applied Sciences and obtained his MSc in machine learning from Imperial College London. He was a CERN Fellow, worked at SAP and is an alumnus of the German National Academic Foundation (Studienstiftung des deutschen Volkes). His current interests include artificial intelligence, machine learning, deep learning, anomaly detection and big data.