Overview

Fraud detection demonstrates problems like stolen credit cards, telemarketing / internet scams, exaggerated insurance losses etc. The key for building a successful fraud detection system are hybrid models and active collaborations between domain experts from business practices, law and data scientists.

Main difficulties

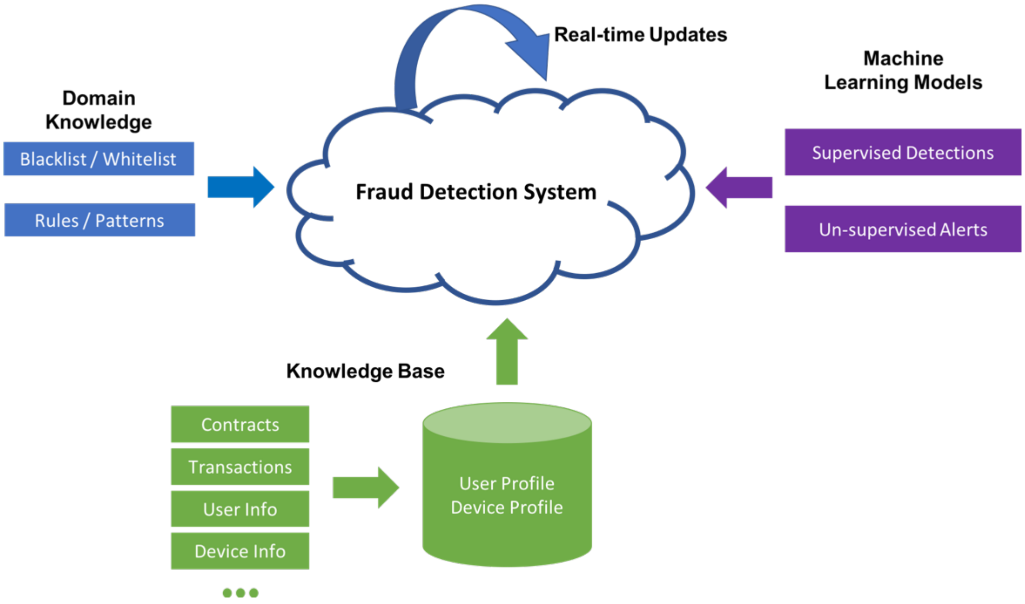
1. Criminals are extremely devious. New tactics and unknown schemes are emerging all the time. Therefore, historical records have weak relationship to new data.
2. Fraudulent patterns are too complex to be explicitly described by rules or conventional mathematical models.
3. Datasets are highly biased. Fraudsters are hidden behind large amount of normal users. Detection models are easily over punishing normal transactions or overlooking fraudulent transactions.

Solutions

A hybrid of domian knowledge based models and machine learning models is mandantory to cope with the complexity in fraud detection.

Knowledge Base

The success of fraud detection is based on high quality, various sources and real-time knowledge base. 360° user profiles and device profiles can be built by collecting data from contracts, financial transactions and network communications etc. Profiles are linked by algorithms to build a knowledge graph, which enables data mining techniques such as user group clustering and anomaly detection at a later stage.



Domain knowledge

Domain knowledge from business practices and law experts plays a central role in fraud detection. They need to analyse fraud patterns and help data scientists model them using formal data structures. Two of the most effective and popular methods are users / devices blacklist and rule-based detection system.

When defined elaborately, these tools require minimal engineering and dataset to implement and detect fraud behaviors accurately. Domain experts' inputs are also crucial to machine learning models. The downsides of these methods are that only a small portion of fraud behaviors can be detected in this manner and the maintanance of these tools is expensive.

Machine Learning Models

Machine learning techniques have negative and positive strategies. They are used to cope with schemes which can not be explicilitly defined by domain experts and unseened patterns.

Negative strategy use historical data and supervised learning models such as SVMs and Neural Networks to learn hidden dynamics behind extremely complex fraud behaviors. Based on historical behaviors, those trained models can be used to detect similar patterns.

Positive strategy use un-supervised learning models such as Probablistic Graphical Models (PGMs) and Neural Networks to discover unseened fraud tactics and raise alerts before actual fraud activity happens.

Machine learning models outperform other models in large-scale real-time scenarios and can be adapted to a much wider range of problems. The downsides are that they require large scale dataset to develop and the implementation and training of them are expensive.