Specific theme: Detecting payment fraud with artificial intelligence

Filip Caron

The cyberthreats faced by the payment system and network users have never been greater. Cyberadversaries are adopting increasingly sophisticated techniques to stealthily access users' critical assets. Forensic analyses of recent cyberincidents have uncovered highly covert malware that could bypass advanced controls like two-factor authentication. Similar threat evolutions have been reported for the retail payment market. Febelfin, the association representing the Belgian financial sector, recently reported an important increase in retail payment (e-banking) fraud as a direct result of cyberattacks. The silver linings to this evolution are an increased cybersecurity awareness and the launch of innovative artificial intelligence-based fraud detection systems.

Artificial intelligence is gaining significant traction under the impulse of converging trends. First, data is proliferating at an extraordinary rate as a direct result of increased data generation, new storage paradigms like Apache Hadoop and accessible cloud storage. Secondly, computing power has exponentially increased after decades of improvements in line with Moore's Law[1] and recent hardware innovations geared towards AI specific improvements like Spark and Googles Tensor Processing Units. Thirdly, while developing full general AI remains an objective for the far future, recent algorithmic advances have resulted in remarkable improvements in areas like image recognition.

Payment fraud detection has been identified as a promising use case for AI. Vast amounts of historical data can be analysed to identify suspicious behavioural patterns. Compared to more traditional methods, AI-based systems are expected to develop more accurate and complex criteria to determine whether a payment is likely to be fraudulent. Additionally, given the increasing demands for fast or even immediate payment processing, screening should be completed in a fraction of a second. Fraudulent transactions are frequently reported by victims to the payment system, which enables the fraud detection algorithms to identify evolving tactics.

The recognition of AI-based systems' potential to detect and ultimately prevent payment fraud has been driving the research agenda of several payment system and network operators. Examples in the wholesale and retail payment segments are provided in the box below. The aim of this article is to provide its reader with key insights into the AI concepts under research, the opportunities and challenges of applying AI in payment fraud detection as well as the relation with recent policy frameworks and strategies.

Artificial intelligence: From concept to specific techniques

Artificial intelligence (AI) focuses on simulating human-like cognitive functions, such as perceiving and correctly interpreting data, flexibly adapting and learning, as well as problem solving. Predictions and recommendations made

1 According to Gordon Moore's Law, the number of transistors on a computer chip doubles every two years. This historical trend has been observed between 1975 and 2012.

by AI-based systems are typically the result of detecting patterns in vast amounts of data, rather than executing explicit specified instructions in programme code. AI algorithms are not static but adapt and improve in response to new data.

Three objectives of data analytics are typically being distinguished, namely, in an increasing order of complexity, descriptive, predictive and prescriptive analytics. Descriptive analytics are used to describe what happens and are heavily used in the financial industry. The most promising opportunities in AI are offered by predictive analytics, that foretell what will happen in the future and prescriptive analytics, that recommend a course of actions undertaken to achieve specific objectives.

The extensive set of AI-techniques that have been proposed over the course of seven decades[1] are typically classified in three broad categories: supervised learning, non-supervised learning and reinforcement learning.

Supervised learning

A supervised learning algorithm infers a relation between the set of input variables and the output variable, based on known input-output pairs. For example, AI-based fraud detection looks at how inputs like payment value, currency and timing could be used to predict whether a payment is fraudulent or not.

The major precondition for supervised learning is that it is possible to provide a training set with known input-output pairs, which for certain payment systems and networks is available, as fraudulent payments are frequently reported thanks to incentives such as potential refunds or avoidance of further losses. The reported payments are labelled as fraudulent, while the other ones are typically considered non-fraudulent.

Once the deduced relation is considered sufficiently accurate (predetermined by the data scientist tasked with training the algorithm), the supervised learning algorithm can be applied to new data sets to determine whether they contain fraudulent payments.

More popular techniques include regression analyses, decision trees, Bayesian belief networks and simple neural networks. The Chart below provides an overview of the key constructs for the different techniques.

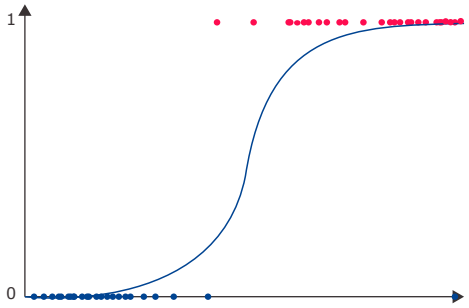
- *Linear and logistic regression analyses* are standard statistical methods that result in a mathematical expression that explicitly specifies the impact of the different input types on the output data. Regression analyses are simple to implement, but tend to be outperformed by the other techniques in this section;
- *Decision trees* provide highly interpretable classification models composed of a structure with decision nodes based on data feature values (e.g. timing) that split into branches (e.g. within versus outside normal operating hours) until a final decision output is made. In a simplistic example, the next decision node could centre on the data feature "beneficiary country", which will trigger a final decision output "fraudulent payment" if the country is on a high-country risk list;
- *Bayesian belief networks* represent the directed causal relations between events. In the context of fraudulent payments, a Bayesian belief network could include relations specifying the probability that a payment from an originator in Country X is in USD, the probability that a payment originated in Country X is fraudulent, the probability that a payment in USD is fraudulent, etc. Bayes theorem allows to calculate the probability of an event based on knowledge of other events. Bayesian belief networks allow for high computational efficiency, but require a strong understanding of typical and fraudulent payment behaviour;
- *Artificial neural networks* are algorithms that are vaguely inspired by the human brain. These networks rely on – anywhere from a few dozen to millions of – artificial neurons that process input data and influence other artificial neurons. Artificial neurons are grouped in different layers and are connected with neurons from adjacent layers. These connections between artificial neurons are weighted, with a higher weight resulting in greater influence. Artificial neural networks have a well-established history with fraud detection research. Their major downside is the high computation power needed to train and operate the neural network.

1 Artificial intelligence was founded as an academic discipline and research area in the summer of 1956 at Dartmouth College (Hanover, United States).

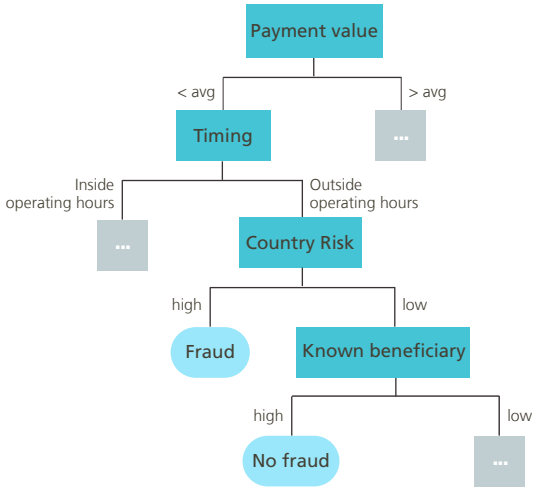
Chart 6

Key constructs in artificial intelligence (simplified examples)

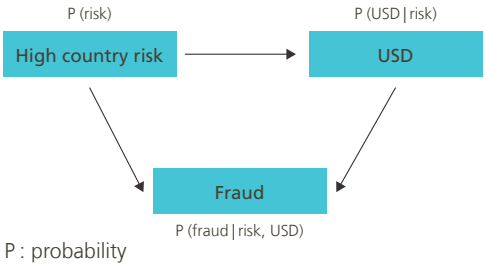
(a) Logistic regression



(b) Decision Tree

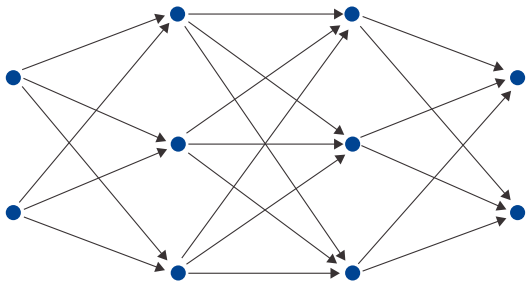


(c) Bayesian belief network



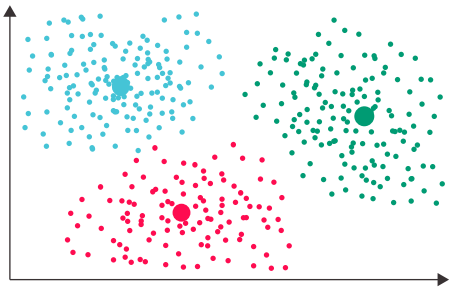
P : probability

(d) Artificial neural network



Neurons

(e) K-means clustering



Unsupervised learning

While the vast majority of AI-based fraud detection tools learn from labelled input-output pairs, several researchers reported on the use of unsupervised learning. The objective of unsupervised learning is to find patterns and a classification structure in large data sets without an explicit link between the input data and the output variable, e.g. identifying customers that exhibit similar payment behaviour (output variable, no predefined groups) based on their recent payments (input data).

A common unsupervised learning objective is clustering observations/objects, i.e. grouping data points that are highly similar to and rather different from data points in other groups. Clustering algorithms like the K-means use

iterative techniques to optimise the grouping based on predefined similarity metrics. This clustering technique has been used to group customers with similar behaviour and provide them with "standardised" payment controls to detect fraud.

But anomaly detection is probably the most interesting unsupervised learning technique objective for fraud detection. This technique monitors behaviour over time using different baselines.

- *Peer group analysis* detects users that are starting to behave in a distinctly different manner from users that were previously identified as highly similar. Increasing dissimilarity can be identified through both externally defined criteria or internal criteria which summarise previous behaviour. Peer group analysis techniques typically flag the most deviating payment behaviour as a transaction that merits closer investigation;
- *Break point analysis* aims at identifying changes in the payment behaviour of a single customer. The algorithms compare recent payment behaviour with historic data to identify material changes for the user (e.g. significant increase in the level of spending) which may not be captured with traditional rules or outlier detection techniques.

Reinforcement learning

Reinforcement learning refers to a set of AI-techniques that aim to learn how to optimise policies by trial and error. The algorithms interact with the environment and try to maximise a certain metric, e.g. optimising the return on investment of a portfolio. Reinforcement learning is typically applied in environments characterised by limited training data, vaguely specified end states or where learning about the environment is possible only through interaction. The currently limited applications of reinforcement learning have been reported in research.

Fraud detection challenges

AI-based fraud detection tools have the potential to identify potentially fraudulent payments more accurately and rapidly compared to a manual review or matching of transactions. As a result, these fraud detection tools may significantly reduce the losses (directly) related to payment fraud. However, there are some limitations.

Typical limitations of AI classification functions

As payment fraud detection is a typical classification problem, these techniques will typically be subject to issues related to the training set provided to the algorithm and to business constraints. The former typically manifest themselves when there are too little examples of fraudulent payments to learn patterns (i.e. skewed class distribution), and/or when the training set is not a good representation of all potential fraudulent behaviour, resulting in limited detection capabilities for new data sets (i.e. overfitting).

Payment processing efficiency or the time needed to initiate, clear and settle payments, is increasingly under scrutiny. While parameterisation of the algorithms is crucial in achieving maximum accuracy, (parameterisation of) the fraud detection algorithms should allow for computational efficient classifications in support of contemporary business requirements. New initiatives in the retail market like the TARGET Instant Payment Settlement (TIPS, see section 3.1) service aim at settling payments within seconds. Similarly, the Global Payments Innovation (GPI, see section 4) initiative has resulted in a significantly increased efficiency for cross-border payments, with over 50 % of the payments credited to the end beneficiaries within 30 minutes.

Ideally, payment fraud detection algorithms correctly classify all payment transactions into fraudulent and non-fraudulent categories. However, in reality there will always be false positives (legitimate payments marked as fraudulent) and false negatives (fraudulent payments marked as legitimate). False positives could result in important losses. Market research reported significant losses for retailers as a result of false declines and reputational damage for the card-issuing financial institutions.

Concept drift: fraudsters changing tactics

Supervised AI-based fraud detection techniques are extensively trained to highly accurately identify previously observed fraudulent behaviour. As the sophistication of these fraud detection techniques increases, the incentives for fraudsters to change their behaviour increase. These fraudsters have proven to be highly effective in detecting and subsequently circumventing geo-blocking and time and value constraints in the detection models.

Consequently, supervised AI-based detection techniques should be frequently updated to learn evolving trends in fraudulent behaviour. Adapting tools to changing fraudster tactics can be hard. First, retraining may demand significant computational power in case the tool is based on e.g. neural networks. Secondly, computational efficient techniques such as Bayesian belief networks require a strong prior knowledge and understanding of typical and abnormal behaviour.

Securing the fraud detection tool

Forensic cybercrime analyses of recent payment fraud-related incidents at banks have not only evidenced a significant global increase in the adversaries' level of understanding of business-oriented controls but also of the implemented security measures. Security experts reported a growing sophistication of cyberattack tactics to effectively disable and circumvent control and security measures. It is not considered unimaginable that hackers may succeed in bypassing or impacting AI-based fraud detection tools.

Network-based fraud detection services aim at mitigating the risks related to compromised environments in individual financial institutions. Payments are screened and validated in external IT environments before being cleared and settled. Credit card schemes have since long implemented network-based fraud detection techniques, and continue to further innovate in this area (e.g. MasterCard Decision Intelligence). SWIFT's Payment Control Service is an example of network-based fraud detection for wholesale payments. These examples are further explained in the box below.

Fraud detection in policy frameworks and strategies

Automated fraud detection has been recurring in recent policies, strategies and industry initiatives to reduce payment fraud.

Transaction risk analysis: risk-based exemptions for payment initiation

With the revision of the Payment Services Directive, the Commission defined a series of general security principles, including the need for a strong customer authentication (SCA). The EBA's interpretation, as described in its recent regulatory technical standards (applicable after 14 September 2019), requires verification of the customer's identity using at least two factors out of three (e.g. something you know, something you have and something you are). However, two-factor authentication might disrupt the customer experience and inhibit frictionless processing, which according to marketing research could result in significant levels of shopping-cart abandonment for online merchants.

An exemption of strong customer authentication can be obtained for small value transactions (less than € 500) if a merchant's acquiring bank has a sufficiently good fraud rate (based on the exemption threshold values) and if transaction risk analysis is implemented. This transaction risk analysis should consider geo-location, previous patterns of expenditure and all other relevant data items, making it an interesting use case for AI-algorithms. Note that even though the merchant's acquirer can claim the transaction risk analysis exemption, the issuer has the final decision and can turn down the request.

CPMI endpoint security strategy: Fraud detection in wholesale payments

Recent cyberincidents have highlighted the increasing sophistication of fraud in the wholesale payment ecosystem. Cyberattackers succeed in exploiting security weaknesses in the ecosystems endpoints (i.e. infrastructures of the connected financial institutions), resulting in both material financial risks to individual institutions and systemic risks to the ecosystem.

In response, the Committee on Payments and Market Infrastructures (CPMI) has proposed a holistic endpoint security strategy to encourage and coordinate industry initiatives. The fourth element of the strategy explicitly addresses the adoption of payment fraud detection techniques, yet another important business case for AI.

SWIFT's Customer Security Programme contains an example of the CPMI endpoint security operationalisation, the Payment Control Service being a concrete implementation of element four. The adoption of (AI-based) fraud detection algorithms is further stimulated through the advisory control (i.e. Transaction Business Control) to restrict transaction activity to validated and approved counterparties and within the expected bounds of normal business.

BOX 14

Examples of network-based fraud detection services

MasterCard Decision Intelligence (retail payment fraud detection)

MasterCard's Decision Intelligence implements data-driven algorithms to analyse and learn a specific account's spending behaviour, which after time enables the detection of abnormal behaviour. With this Decision Intelligence service, MasterCard aims at improving the accuracy of real-time approvals and reducing false declines, reducing operational expenses like chargebacks and improving customer experience.
A wide variety of account data – like customer value segmentation, risk profiling, location, merchant characteristics and time of the day – are leveraged to provide the card issuer with a predictive fraud risk score. The issuer can incorporate this predictive score in its existing fraud mitigation framework and solutions. Alternatively, the issuer could opt for MasterCard's holistic service that makes real-time decisions tailored to individual accounts.

SWIFT's Payment Control Service (wholesale payment fraud detection)

The Payment Control Service (PCS) supports SWIFT participants in detecting and preventing high fraud risk payments. As sophisticated cyberattackers would be able to circumvent payment screening controls in a compromised IT environment, the service is SWIFT-hosted with a zero footprint in the participants' IT environment.

SWIFT participants design a payment risk policy that enables real-time monitoring and alerting and blocking of sent payments. The business rules in the payment risk policy describe the expected payments behaviour, e.g. the rules cover typical characteristics like timing, thresholds, beneficiaries and currencies.

Advanced algorithms enable the identification of behavioural patterns and stimulate continuous improvement of the payment risk policy.