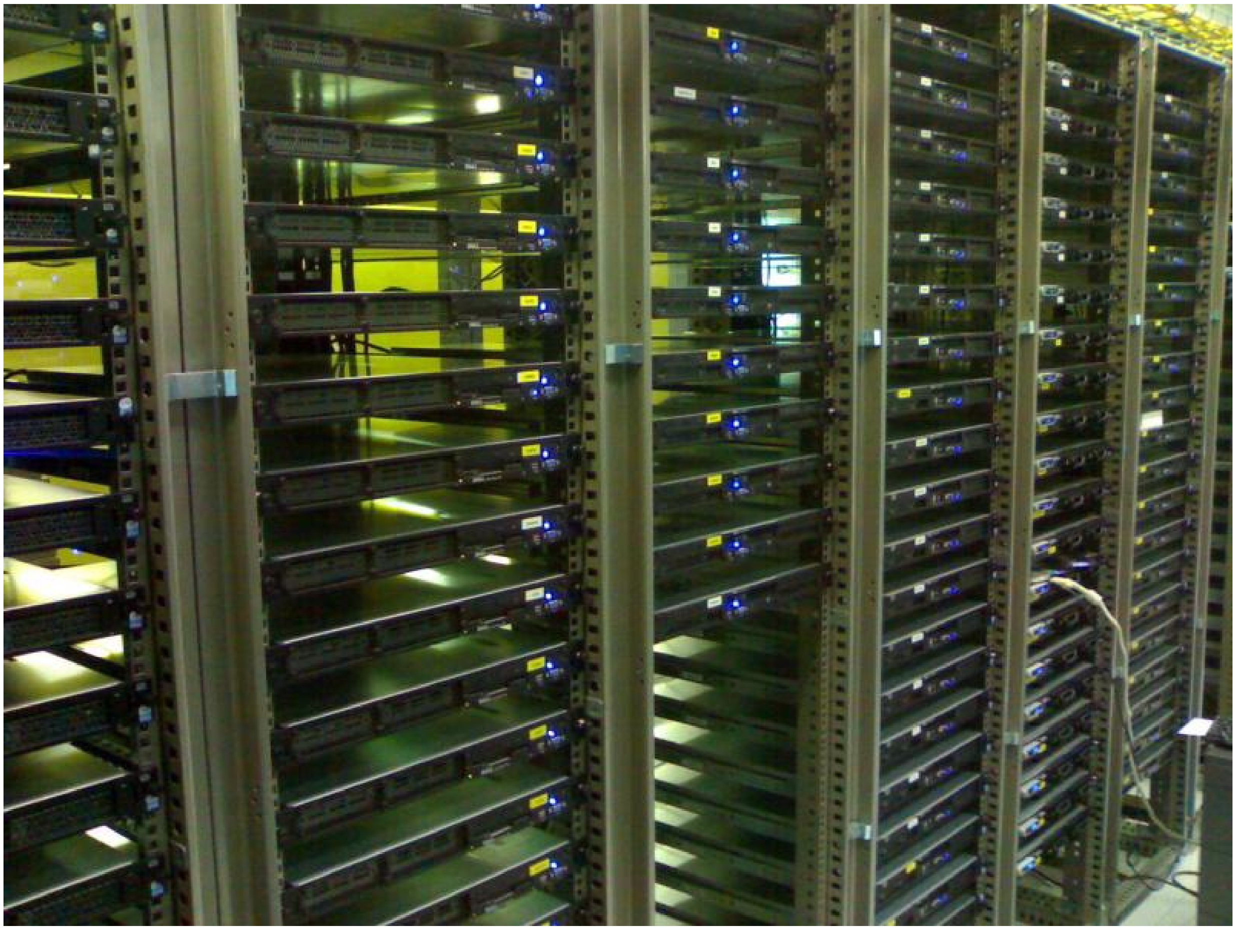
Machine learning and big data know it wasn't you who just swiped your credit card

November 26 2015, by Jungwoo Ryoo



It takes a lot of computing power. Credit: Stefano Petroni, CC BY-NC-ND

You're sitting at home minding your own business when you get a call from your credit card's fraud detection unit asking if you've just made a purchase at a department store in your city. It wasn't you who bought expensive electronics using your credit card – in fact, it's been in your pocket all afternoon. So how did the bank know to flag this single purchase as most likely fraudulent?

Credit card companies have a vested interest in identifying financial transactions that are illegitimate and criminal in nature. The stakes are high. According to the Federal Reserve Payments Study, Americans used credit cards to pay for 26.2 billion purchases in 2012. The estimated loss due to unauthorized transactions that year was US$6.1 billion. The federal Fair Credit Billing Act limits the maximum liability of a credit card owner to $50 for unauthorized transactions, leaving credit card companies on the hook for the balance. Obviously fraudulent payments can have a big effect on the companies' bottom lines. The industry requires any vendors that process credit cards to go through security audits every year. But that doesn't stop all fraud.

In the banking industry, measuring risk is critical. The overall goal is to figure out what's fraudulent and what's not as quickly as possible, before too much financial damage has been done. So how does it all work? And who's winning in the arms race between the thieves and the financial institutions?

Gathering the troops

From the consumer perspective, fraud detection can seem magical. The process appears instantaneous, with no human beings in sight. This apparently seamless and instant action involves a number of sophisticated technologies in areas ranging from finance and economics to law to information sciences.

Of course, there are some relatively straightforward and simple detection mechanisms that don't require advanced reasoning. For example, one good indicator of fraud can be an inability to provide the correct zip code affiliated with a credit card when it's used at an unusual location. But fraudsters are adept at bypassing this kind of routine check – after all, finding out a victim's zip code could be as simple as doing a Google search.

Traditionally, detecting fraud relied on data analysis techniques that required significant human involvement. An algorithm would flag suspicious cases to be closely reviewed ultimately by human investigators who may even have called the affected cardholders to ask if they'd actually made the charges. Nowadays the companies are dealing with a constant deluge of so many transactions that they need to rely on big data analytics for help. Emerging technologies such as machine learning and cloud computing are stepping up the detection game.

Learning what's legit, what's shady

Simply put, machine learning refers to self-improving algorithms, which are predefined processes conforming to specific rules, performed by a computer. A computer starts with a model and then trains it through trial and error. It can then make predictions such as the risks associated with a financial transaction.

A machine learning algorithm for fraud detection needs to be trained first by being fed the normal transaction data of lots and lots of cardholders. Transaction sequences are an example of this kind of training data. A person may typically pump gas one time a week, go grocery shopping every two weeks and so on. The algorithm learns that this is a normal transaction sequence.

After this fine-tuning process, credit card transactions are run through the algorithm, ideally in real time. It then produces a probability number indicating the possibility of a transaction being fraudulent (for instance, 97%). If the fraud detection system is configured to block any transactions whose score is above, say, 95%, this assessment could immediately trigger a card rejection at the point of sale.

The algorithm considers many factors to qualify a transaction as fraudulent: trustworthiness of the vendor, a cardholder's purchasing behavior including time and location, IP addresses, etc. The more data points there are, the more accurate the decision becomes.

This process makes just-in-time or real-time fraud detection possible. No person can evaluate thousands of data points simultaneously and make a decision in a split second.

Here's a typical scenario. When you go to a cashier to check out at the grocery store, you swipe your card. Transaction details such as time stamp, amount, merchant identifier and membership tenure go to the card issuer. These data are fed to the algorithm that's learned your purchasing patterns. Does this particular transaction fit your behavioral profile, consisting of many historic purchasing scenarios and data points?

The algorithm knows right away if your card is being used at the restaurant you go to every Saturday morning – or at a gas station two time zones away at an odd time such as 3:00 a.m. It also checks if your transaction sequence is out of the ordinary. If the card is suddenly used for cash-advance services twice on the same day when the historic data show no such use, this behavior is going to up the fraud probability score. If the transaction's fraud score is above a certain threshold, often after a quick human review, the algorithm will communicate with the point-of-sale system and ask it to reject the transaction. Online purchases go through the same process.

In this type of system, heavy human interventions are becoming a thing of the past. In fact, they could actually be in the way since the reaction time will be much longer if a human being is too heavily involved in the fraud-detection cycle. However, people can still play a role – either when validating a fraud or following up with a rejected transaction. When a card is being denied for multiple transactions, a person can call the cardholder before canceling the card permanently.

Computer detectives, in the cloud

The sheer number of financial transactions to process is overwhelming, truly, in the realm of big data. But machine learning thrives on mountains of data – more information actually increases the accuracy of the algorithm, helping to eliminate false positives. These can be triggered by suspicious transactions that are really legitimate (for instance, a card used at an unexpected location). Too many alerts are as bad as none at all.

It takes a lot of computing power to churn through this volume of data. For instance, PayPal processes more than 1.1 petabytes of data for 169 million customer accounts at any given moment. This abundance of data – one petabyte, for instance, is more than 200,000 DVDs' worth – has a positive influence on the algorithms' machine learning, but can also be a burden on an organization's computing infrastructure.

Enter cloud computing. Off-site computing resources can play an important role here. Cloud computing is scalable and not limited by the company's own computing power.

Fraud detection is an arms race between good guys and bad guys. At the moment, the good guys seem to be gaining ground, with emerging innovations in IT technologies such as chip and pin technologies, combined with encryption capabilities, machine learning, big data and, of course, cloud computing.

Fraudsters will surely continue trying to outwit the good guys and challenge the limits of the fraud detection system. Drastic changes in the payment paradigms themselves are another hurdle. Your phone is now capable of storing credit card information and can be used to make payments wirelessly – introducing new vulnerabilities. Luckily, the current generation of fraud detection technology is largely neutral to the payment system technologies.

This story is published courtesy of The Conversation *(under Creative Commons-Attribution/No derivatives).*

Source: The Conversation

Citation: Machine learning and big data know it wasn't you who just swiped your credit card (2015, November 26) retrieved 18 April 2024 from https://phys.org/news/2015-11-machine-big-wasnt-swiped-credit.html