

USING CREDIT CARD FRAUD DETECTION SOLUTIONS IN THE PROCESS INDUSTRIES

By Sebastian Amos -GAMBICA - Sector Head, Process Instrumentation & Control



Many people reading this will have had "the call". Today, you may receive it by

text. "Have you just made a purchase for a 56" HD TV in Currys, Singapore? "Press 1 for yes. 2 for no". Then starts the real conversation and in most cases the conclusion is validated, the card stopped and life goes on.

Detecting fraudulent transactions is arguably the most useful use case to individuals for big data at Amex, as it is for most financial services companies. Try as they do, it's quite difficult for fraudsters to create transactions that totally mimic real transactions in every detail, and machine learning algorithms are quite good at picking out these anomalies⁽¹⁾

If you replace "Credit Card" with "Process Variable" or "Machine Status" and "Fraudulent" with "Concerning", there seems to be the possibility of merging technologies.

ANALYSIS

Where are the process industries today on this potential path to better use of big data to improve the process and better manage maintenance? A recent paper from Don Rozette published in Control Global magazine⁽²⁾ recognises that within most organisations, maintenance divisions are

"siloes" and the process team can be oblivious to the maintenance issues, thus things get fixed but no real review is done on a) why they failed and b) how can the process be changed to remove the failure. There is also recognition that 75% of assets operate on a "run to fail" basis.

With more and more data available within the historians, the first analysis to be done requires assets to be classified and prioritised as a minimum requirement to focus on the 25% that cannot operate on the "run to fail".

This is a view of today. Big data and data analytics, by default will remove the silos.

Big data, Industrial Internet of Things (IIoT) and Data analytics have enormous potential within the process industries and well beyond just machine management.

The amount of data that can be produced by any process automation system is now at a level that simple KPI graphs and alarm systems are flooded with data and manually interpreting this is becoming near impossible. As the Industrial Internet of Things (IIoT) develops over the next 10 years, the data will continue to increase. There is, however, time on our side. The process industries tend to lag the cutting-edge technologies for two reasons. One, developers tend to wait for the systems to become stable before deploying them into an industrial context and two, the turnover of new technology is slow. New greenfield plant will take advantage as soon as it can, but brownfield upgrades, particularly in the UK, tend to be changed out only when a device fails. So, what can we learn from technology available today in other sectors and how should it be developed and deployed? Again, we can look at the technologies used in the credit card industry.

Learning what's legit, what's shady - An extract from a report in "theconversation.com"⁽³⁾

"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."

In summary, actual data is compared against model data and when the two show a difference, an alarm is raised. In the process industries, we do something like this already. We check temperatures and pressures. We check yields and quality. Some of this is performed in real time, some is performed in the laboratories.

We have asset management systems that store data about runtime and propose maintenance cycles. We can look at machine vibration analysis, that while these can alarm and stop machines effectively, thus protecting the machine, the process and people, we still need to have the full process offline while repairs are made. Rarely do we compare a formal high definition model against the actual process, though there are some excellent examples for pipeline leak detection systems.

If we can collect large quantities of data via the IIoT, and create models and learning software, as with credit card fraud, we should be able to predict failure and address it before the process shuts down. Process management is now entering the world of automated data forensic analysis.

The benefits here can be enormous. Oil rig shutdown can cost between \$1m and \$5m per day in lost production. Black starting a rig can take 3 days.

Recalls on products from soap powder to pharmaceuticals can be reduced as the products would not have left the factory.

So, what are the nuances in the process we can use big data for. As examples, we would be able to see when valves start to stick, not necessarily due to valve alarms but measurements taken around the area the valve is installed. We could be looking at flows, temperatures and pressures, with the self-educating software flagging a process error that would provide a focus for detailed examinations. The possibilities are endless.

HOW MUCH DATA ARE WE TALKING ABOUT?

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. (3) We can compare this to a real-time process control system.

A 10,000 I/O HART enabled system could generate 50,000 data points/second which include some locally derived information plus timestamps. Assuming each point is 512 bytes, this equates to 2.2 Terabytes per day. Currently we rarely store that level of data as historians tend to look at max/min/avg over predefined timescales, but that has always been the way around too much data.

We are now approaching data storage and processing capabilities, outside of the financial capabilities of all but the largest organisations. IIoT and Big Data processing now need to look elsewhere and cloud computing is the current solution. Data needs to be stored and analysed and only the results piped back to the users.

HOW CLOUD COMPUTING WORKS

The following summarises an article from "How Stuff Works" (4) Currently, nearly every office based employee in an organisation requires a computer complete with Operating System and applications to perform their job. This is, for example, Windows, Office Suite with email, Internet Access and specific programmes to enable the job to be done. This covers everything from control software development tools, finance tools, Enterprise management etc. The list goes on.

This is a complex and, especially in a large organisation, an ever-changing environment with serious network and security management infrastructures.

Cloud computing changes the focus of where the heavy work is done with efficiencies made through virtualisation that better exploit processing power. Employees can now use basic web browsers or mobile devices to complete their tasks.

How the cloud works in detail is probably not relevant for this discussion. Just assume it does.

TYPES OF CLOUD COMPUTING

There are three main types of cloud computing:

1. Infrastructure as a Service (IaaS) means you're buying access to raw computing hardware over the Net, such as servers or storage. Since you buy what you need and pay-as-you-go, this is often referred to as utility computing. Ordinary web hosting is a simple example of IaaS: you pay a company serve up files for your website from their servers.

2. Software as a Service (SaaS) means you use a complete application running on someone else's system. Web-based email and Google Documents are perhaps the best-known examples.

3. Platform as a Service (PaaS) means you develop applications using Web-based tools so they run on systems software and hardware provided by another company. So, for example, you might develop your own ecommerce website but have the whole thing, including the shopping cart, checkout, and payment mechanism running on a

merchant's server. App Cloud (fro salesforce.com) and the Google App Engine are examples of PaaS." PaaS lends itself to the requirements of Big Data analysis in the Process Industries.

PROS

- Lower upfront costs and reduced infrastructure costs.
- Easy to grow your applications.
- Scale up or down at short notice.
- Only pay for what you use.
- Everything managed under SLAs.
- No more slow internet on a Monday Morning as upgrades are sent to ever machine.
- Overall environmental benefit (lower carbon emissions) of many user efficiently sharing large systems. (But see the box below.)

CONS

- Higher ongoing operating costs. Could cloud systems work out more expensive?
- Greater dependency on service providers. Can you get problems resolved quickly, even with SLAs?
- Risk of being locked into proprietary or vendor-recommended systems? How easily can you migrate to another system or service provider if you need to?
- What happens if your supplier suddenly decides to stop supporting a product or system you've come to depend on?
- Potential privacy and security risks of putting valuable data on someone else's system in an unknown location?
- If lots of people migrate to the cloud, where they're no longer free to develop neat and whizzy new things, what does that imply for the future development of the Internet?
- Dependency on a reliable Internet connection.

WHERE ARE WE TODAY?

The large computing based organisations all offer cloud services. Microsoft (5), IBM (6) and Oracle (7) lead the way, however, the data analysis tools available are still mainly focused on supply chain management, CRMs and building services. Whilst IBM and Oracle have products with a consumer focus, Microsoft does have examples of configuration on its Azure platform for predictive maintenance(5).

The focus for the Process Industries is therefore Platform as a Service (PaaS) but the focus moves to the applications that can help us.

We started this article looking at the low hanging fruit benefits of big data, cloud computing and learning analytics within the maintenance sectors.

There are products available that are using these technologies already. Microsoft has a complex system which can be reviewed – see Ref 8. However, there are offerings from SMEs that could be considered closer to the end user. Based on a technology generically called "Prognostics"

- An engineering discipline focused on predicting the time at which a system or a component will no longer perform its intended function (9). Solutions such as Prognosys TM from Senseye (10) and Cassantec (11) use machine learning and advanced analytics extracting data from multiple sources to achieve this. The actual analytics tend to be the domain of the latest breed of Data Scientists.

These solutions, including the Microsoft offering, have a focus on

identifying from historical, real-time, repair and failure histories the anticipated failure of specific items. This can work well for production lines and some process lines but it still does not reach to where it is believed/hoped that the IIoT can take the process industries.

WHAT NEEDS TO BE DONE AND NEXT STEPS

Firstly, we need to see beyond the uses within maintenance. The target today is as we stated at the start:

- Process anomaly detection
- Product continuous quality management
- Self-correcting systems
- Planned maintenance as part of both the process and maintenance engineers brief.

We have the design tools available today, lifted from the Credit Card company structured design approach. They are still closely aligned with software analysis procedures developed by the data scientists where the picture below, borrowed from Microsoft (5), shows it can be easily used as the basis for industrial needs.

Developing this will enable the new systems to collect large quantities of data, analyse it either against models, automatically update and modify the models and then either ask predefined "sharp questions" e.g. will this system fail in the next 36 hours, or offer a heat map that simply drives the process and maintenance teams to act on potential issues before these start affecting quality and yield.

None of this can be done by human review today as the data quantities are too large. However, the IIoT will be providing the data and the Data Scientists already understand the learning software applications from banking to enable the crossover.

In 10 years time, there will still be alarm systems and reports but the process and management of the process will be ordered by the intelligent data crunchers operating on the cloud.

To reach this aim, companies must maintain a rolling upgrade path on all their equipment and replace it as it fails with IIoT enabled hardware.

IT/OT infrastructure need to be quietly upgraded to meet the higher demands for bandwidth and radio communications. Computer systems need to migrate to cloud based services

A ten-year period will bring plants to a ready to go status, the application designers will have their systems ready and the two will meet. ●

References

1. www.datanami.com/2016/05/03/credit-card-companies-evolving-big-data/
2. www.controlglobal.com/articles/2016/big-data-analytics-to-manage-asset-performance/?utm_source=hs_email&utm_medium=email&utm_content=39554198&_hsenc=p2ANqtz-c9qnNr3qoCxeweBlj_
3. www.theconversation.com/machine-learning-and-big-data-know-it-wasnt-you-who-just-swiped-your-credit-card-48561
4. www.explainthatstuff.com/cloud-computing-introduction.html
5. Microsoft - www.azure.microsoft.com/en-gb/
6. IBM - www.ibm.com/big-data/us/en/
7. Oracle - www.oracle.com/uk/cloud/get-started.html
8. www.gallery.cortanaintelligence.com/Tutorial/Predictive-Maintenance-Template-with-SQL-Server-R-Services-1
9. www.en.wikipedia.org/wiki/Prognostics
10. www.senseye.io/solution/-what
11. www.casantec.com/

