5-2013

What's the big deal with Big Data?

Singapore Management University

What's the big deal with Big Data?

Big Data does not necessarily cost big bucks, and smaller companies can benefit from it too.

On 23 April 2013, at 1:07pm U.S. Eastern Standard Time, the Dow Jones Industrial Average suddenly plunged 130 points in two minutes, wiping out $136 billion from the S&P 500 index. Within another two minutes, those losses were recovered. The trigger which set off the tumultuous set of events was a tweet from The Associated Press (AP): "Breaking: Two explosions in the White House and Barack Obama is injured."

It is now well known that AP's Twitter account was hacked, which the news agency quickly made clear when it saw the havoc the tweet had wrought on financial markets worldwide. In the aftermath of what is now known as "The Flash Crash of 2013", much attention was focused on High-Frequency Trading (HFT) and how computer algorithms that were beyond the control of humans had caused widespread panic.

"The data analysis that was run by algorithms was basically machines reading tweets to look for combinations of keywords," says **Antony Creighton** of Wellington Management. "These algorithms saw a tweet that said, 'White House bombed. Obama injured.' That's all it needed."

"What happens next time is (the algorithms will be adjusted): 'We'll need two sources', which is traditional journalism. So that's how these systems develop internally – every failure is actually a source of improvement. The more data you run, and the faster the data, the better the road-testing of the concept will be."

Big Data = Big potential

Creighton was speaking to *Perspectives@SMU* recently at [BFI's Knowledge Series: "Future Themes: Big Data, Health Care and the Power of Women" and "Evaluating Asian Real Estate"](#), where Big Data took centre stage. In the same way that financial firms' computer algorithms picked up and acted on the AP tweet, companies have developed similar software to cull the ever-increasing amounts of data available to them both online and from other interactions, to make sense and, they hope, profit.

"There are companies - branded consumer companies - which have developed technologies to scan the broad variety of data sources, Facebook, Instagram, Twitter, anything that carries language or images, for mention or images of their own products," says Creighton. "What they're doing is testing the success and popularity of their own products by seeing how often they crop up ."

"If facial recognition (software) can distinguish between two faces, it can definitely recognise the can of a brand of soft drink from another; it's the same technology, just simpler. Companies track that, and they track their competitors' appearances as well. So if they are falling behind, they will do something about it – they'll outbid the other company for positioning on the latest blockbuster movie."

Big Data = Big money?

It is a lot of information to sift through, on top of the videos, podcasts, and other forms of data that flood the internet these days. Google Executive Chairman Eric Schmidt made the famous proclamation in 2010 that "there were 5 exabytes of information created between the dawn of civilisation through 2003, but that much information is now created every 2 days, and the pace is increasing".

One exabyte is a million terabytes, and one terabyte is a thousand gigabytes. To put that into perspective, a single DVD can carry up to 4.7 gigabytes of information. Surely only the biggest companies can afford to manage and capitalise on Big Data, right?

"It seems that way because we talk about huge volumes of data, but the technology that you use is cheap technology," Creighton clarifies. "For a small and medium-sized enterprise, it's not an issue because they can rent space rather than own, just as they can on the cloud. In fact, the cloud is a kind of representation of Big Data storage."

Creighton adds, "The other thing is Hadoop, which is the core software, an open source software – it's free. What isn't free is the developer time to build something. Again there's a Pareto answer here, which says that 80 percent of the objectives will be met by 20 percent of the tasks. In other words, everybody wants the same things. There will be websites where you can go and pay a relatively small amount of money for that app."

"It may be expensive per project, but it is relatively cheap on an annual basis if you go to an agency or a consultant who will do it for you. (Hence) people shouldn't be thinking, 'I'm not a big company, and this is Big Data so we need not be concerned, every business could benefit from understanding what motivates their customers.'"

Big Data = Big Brother?

While the cost of running Big Data analytics on customer profiles may not be a problem, privacy issues are. With any commercial transaction, be it the purchase of a car or signing up for childcare services, you would be telling the product or service provider something about yourself i.e. you own a car or have children. On top of that, there is also financial and personal information, including your annual income. How do you prevent Big Data from becoming Big Brother?

"It's very, very difficult," Creighton concedes. "Right now, we're at the point in our use of technology where people are pretty willing to give away stuff about themselves in return for 'free' access to a newspaper website without knowing where that's going, who will obtain it, or what it may be used for." "At some point also, no doubt because it's human nature, some criminal activity will take place that takes advantage of people's naivety. When that happens, there will be an outcry and a demand for legislation."

Big Data = Big benefits

That's not to say that Big Data does not bring any benefits. "The digitisation of medical records, particularly in the United States and Europe, is a long way behind where it should be," says Creighton, citing administrative and budget constraints. "What's happening is that they're building machines now that can read the doctors' notes and digitise them."

"And then you have the records of when people have been in hospital, when someone was born, when they go to see their GP. If you can bring all of that together into a mosaic, you'd know a lot about that person's health, and you'll be able to forecast what their health issues might be in the future."

In a sense, Big Data becomes Big Medicine, which might be a good thing.