Big Data & Deep Learning

Focus: financial forecasting

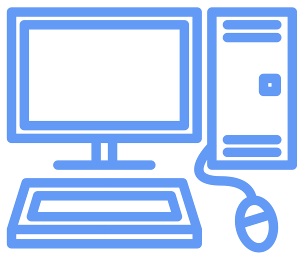
Jurgen Palsma
jurgen.palsma@gmail.com

Image rendered with Google's Autodraw - an online machine learning based application for recognising drawings.

April 2018

Acting as a global record and highway of information, the Internet revolutionised the way people are connected and share experiences. As people are getting more and more access to the internet, their online and offline connections and actions are recorded to an incredible extent. As a result, these connections and events have become a new primordial resource in our economy and societies. In this report, we will explore how two major technical buzzwords, “Big Data” and “Deep Learning”, are part of the foundation for this revolution by briefly defining these terms, and linking them to real world applications, with emphasis on their use in the world of finance.

Big Data



The buzzword “big data” is a conjunction of two terms: “data” - by which we mean characters and numbers logged in registers, and “big” - which is often misleading; it does not only represent the *volume* (size) of data, but represents the *gigantism of all of its dimensions*, such as its *velocity* and *variety*. This is well explained by Gandomi et. al.[1] which presents these three dimensions as “The three Vs of big data”. To try and grasp this gigantism, we can look at some example of “big data” through the most used social media, *Facebook*[2]. When it comes to *Volume*, Techcrunch[3] estimated in 2012 (six years ago!) that Facebook, at the time, processed “500+ terabytes of data each day”. In terms of *Velocity,* Facebook[4] says it handles 1.40 billion worldwide daily active users, which upload more than 900 million photos[5] everyday onto their systems . And finally, in terms of *Variety*, Facebook stores not only basic information and data types such as your name(s) and photo(s), but all kinds of media, events, locations, “likes”, and much more.

As a result, it can become comprehensible that for the basic Facebook user, the use of the platform is free. Or at least free of charge, as users pay for the service not with cash, but with what we called the new oil of the internet, their data. As a matter of fact, The Guardian[6] reports that in 2016, an average Facebook user earned the company from $1.22 if he lived in a developing country, and up to an average of $13.54 if he lived in the US or in Canada. How Facebook makes money out of this virtual oil is through a modern discipline called Data Science, or the arts and techniques of making data useful. Indeed, data is only as valuable as what you make of it, and as refining oil can yield both high and low value fuels, so can Data Science. One of the best data refining used nowadays alongside big data is deep learning.

Deep Learning

Deep Learning is a sub-category of *Machine Learning Algorithms*. These algorithms are statistical models designed for *pattern recognition* in datasets: in other words, they allow to discover regularities in data in order to extract knowledge from it. More precisely, *Deep Learning* is a subfield of *Artificial Neural Networks*, which are nature-inspired machine learning algorithms which are inspired from animal brains and their *neural networks*. As the brain *learns* when it encounters stimulation (through our senses, for example) with their neural network, deep learning algorithms perform in the same way; by giving them repeated stimulations through what we call an *input layer* of neurons, they perform repeated calculations on these inputs through the *hidden layer(s)*, and output extracted data through the *output layer*.

In order to learn patterns, these networks can perform either *supervised* or *unsupervised* learning; Unsupervised deep learning algorithms allow to *cluster* data together, and perform *pattern analysis*, in order to discover recurrent patterns in data. On the other hand, supervised deep learning algorithms allow to categorise and predict data, given a repeated stimulation of the network by *example data samples*, it allows to predict knowledge of an *unknown data sample* depending on what the system learned by these examples.

Focus Big data and deep learning in Finance

Big data and deep learning can be used to optimise and replicate human behaviour - such as image recognition.[7] However, they can also solve problems and tasks which can *hardly be performed by humans*.[8] Indeed, the human mind can be slowed down or overwhelmed by large amount of complex and abstract data, or simply by bias from emotions. As a result, big data and deep learning have become extremely attractive in industries with problems of these types; notably the *Financial Sector*.

The Big (Data) Picture: Algorithmic Trading

A concrete example of the use of big data in the financial sector is how it is used for algorithmic trading. The following figure shows how in less than 10 years, the percentage of market volume of algorithmic trading rose from 15% to 85%. In other words, this means that by 2012, over 85% of trading was done with the help of computers

and algorithms.

Figure 1: The boom of algorithm-processed trading[9]

Volume

Historically, finance "was a small data discipline".[10] Financial investors and traders would often have to rely simply on four indicators for their stock or share: its price at the open, high, low, and close of a certain exchange. Today however, you can get overwhelmed with the enormity of big data in finance; even in 2011, according to IBM,[11] the New York Stock Exchange captured over 1 TB of information per day.

Velocity

Historically, trading was done in face to face, on paper, and this kind of process could take from a few minutes to an hour. Nowadays, financial trading platforms are competing between each other to try and reach as much data velocity as possible: In 2012, Gresham Computing, a financial platform, declared "being able to process 500,000 transactions per second". [12]

Variety

Financial data is not just prices anymore. As a matter of fact, businesses and individuals try and incorporate as much *different* data as possible to get the best tuned algorithm and market knowledge as possible. An example of this incredible variety of data is how some companies like Neuralist[13] sell *textual data* from the press, investor presentations, company quarterly reports, etc… This text is then treated by Natural Language Processing algorithms[14, 15].

An academic example of Deep Finance

The current technical setting of finance is perfect for deep learning. Its volume, velocity and variety is the perfect mix to continuously feed complex deep networks, as a key feature of deep neural networks is that *the more data you give them, the more accurate they are.* This gives deep learning an advantage over other algorithms: as we've seen, financial data is more than abundant, freely available with a high stream, and can be complex and rich if necessary.

For example, Ding et. al.[16] use deep neural networks for "event driven stock market prediction". Their work is separated in three parts, each with its separate type of neural network.

First of all, they extract "*events*" from media (Reuters and Bloomberg financial news) data through text extraction techniques. This way, they get daily timestamped events with an action (for example, an acquisition), an actor (for example, Google) and an object (for example, a small start up). You might see the first application of neural networks here: with this kind of very abstract data, there is no way for a computer to understand it. As a result, they used the neural networks ability to *extract valuable information from complex, abstract data* by using a *Neural Tensor Network*[17]*.* Their neural network took words as inputs, and provided *embedded events* as outputs; an enriched list of similar events.

The second part of their predictive system took these *embedded events*, and processed them through a *Convolutional Neural Network*: a type of deep learning algorithm which is used to *reduce the dimensionality of data*. This way, it allowed them to retain only the most significant events provided by the first part of the system.

Fast Data

It is interesting to note at this point that short term events were fed directly to the last part of the system. This shows us how in today's context of Big Data in finance, the volatility of the markets is increasing and as a result, new and fresh data can get more valuable

These selected events are then processed by the third and last part of the system: a "simple" neural network composed of a hidden layer and an output layer, which mapped the most significant events into a simple binary output: this output represented if a stock price will increase or decrease during that day.

The result of their paper is promising: this complex combination of deep learning network provided, in their experiments, 65% accuracy of individual stock prediction (which is very high accuracy for predicting stocks).

What's next?

Ding et. al.'s work is just one of the thousands of applications of big data and deep learning in the world of finance. In academia, there has been research done on for example, calculating mortgage risk with deep learning[18], or managing portfolios[19].

Although most businesses keep their use of deep learning as secret as possible, in order to avoid leaking information, key actors are actively seeking machine learning and big data specialists, such as JP Morgan which recently published a comprehensive report to attract data scientists in their company[20]. Other businesses, like Qplum[21], share news on their deep learning techniques through reports[21], which they use for prediction or cleaning and extracting data (just like Ding et. al.)

The deep learning and big data disciplines have become so famous in finance that international summits[22] are being organised, and an experienced programmer can even get a glimpse of deep learning in action by following many of the available online tutorials[23].

It is undeniable that in the future, big data will still play a major role in finance. Even if the expansion of its size and complexity can become a problem in some applications, more and more techniques are discovered to extract high quality knowledge and predictions from big data.

On the other hand, it's important to note that deep learning and neural networks, although being revolutionary techniques, are not necessarily the go-to method for all financial analysts and traders. Deep learning models can be complex to understand, compared to other machine learning models such as decision trees, and are often highly computationally expensive, rendering them unattractive for small finance firms for example.

References, cited works, and links

1. Gandomi, Amir, and Murtaza Haider. “Beyond the Hype: Big Data Concepts, Methods, and Analytics.” *International Journal of Information Management*, vol. 35, no. 2, 2015, pp. 137–144., doi:10.1016/j.ijinfomgt.2014.10.007.

2. *Facebook*, facebook.com.

3. Constine, Josh. “How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500 Terabytes Ingested Every Day.” *TechCrunch*, TechCrunch, 22 Aug. 2012, techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/.

4. *Facebook Reports Fourth Quarter and Full Year 2017 Results, https://s21.q4cdn.com/399680738/files/doc_news/Facebook-Reports-Fourth-Quarter-and-Full-Year-2017-Results.pdf*

5. Miller, Rich, et al. “Inside Facebook's Blu-Ray Cold Storage Data Center.” *Data Center Frontier*, 21 Mar. 2017, datacenterfrontier.com/inside-facebooks-blu-ray-cold-storage-data-center/

6. Gibbs, Samuel. “How Much Are You Worth to Facebook?” *The Guardian*, Guardian News and Media, 28 Jan. 2016, www.theguardian.com/technology/2016/jan/28/how-much-are-you-worth-to-facebook.

7. Simonite, Tom. “Facebook Creates Software That Matches Faces Almost as Well as You Do.” *MIT Technology Review*, MIT Technology Review, 19 Sept. 2014, www.technologyreview.com/s/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/.

8. Heaton, J.b., and Nick Polson. “Deep Learning for Finance: Deep Portfolios.” *SSRN Electronic Journal*, 2016, doi:10.2139/ssrn.2838013.

9. Glantz, Morton, et al. *Multi-Asset Risk Modeling: Techniques for a Global Economy in an Electronic and Algorithmic Trading Era*. Elsevier Academic Press, 2013.

10. Aldridge, Irene. “Trends: All Finance Will Soon Be Big Data Finance.” *The Huffington Post*, TheHuffingtonPost.com, 6 Apr. 2015, www.huffingtonpost.com/irene-aldridge/trends-all-finance-will-s_b_6613138.html.

11. “The four V’s of Big Data”. *IBM Big Data & Analytics Hub, www.ibmbigdatahub.com/infographic/four-vs-big-data*

12. Verge, Jason. “Financial Platform Hits 500,000 Transactions Per Second.” *Data Center Knowledge*, 11 Dec. 2012, www.datacenterknowledge.com/archives/2012/12/11/gigaspaces-accelerates-financial-clouds-survey-says-big-year-for-big-data.

13. “NLP Dataset for the Stock Market.” *NLP Dataset for the Stock Market*, neuralyst.com/

14. “Natural-Language Processing.” *Wikipedia*, Wikimedia Foundation, 31 Mar. 2018, en.wikipedia.org/wiki/Natural-language_processing.

15. Wigglesworth, Jason. “AI decodes trading signal hidden in jargon”, *The Financial Times*, 26 Oct. 2017

16. Xiao Ding et. al., “Deep learning for event-driven stock prediction”. IJCAI’15 Proceedings of the 24th international conference on artificial intelligence. Pp 2327-2333. 25 Jul. 2015

17. Socher et. al., “Reasoning With Neural Tensor Networks for Knowledge Base Completion”. https://www-cs.stanford.edu/~danqi/papers/nips2013.pdf

18. Sirignano et. al., “Deep learning for mortgage risk”. *SSRN Electronic Journal*, 23 Jun. 2016

19. Heaton, J.b., and Nick Polson. “Deep Learning for Finance: Deep Portfolios.” *SSRN Electronic Journal*, 2016, doi:10.2139/ssrn.2838013.

20. Butcher, Sarah. “JPMorgan's Massive Guide to Machine Learning Jobs in Finance.” *EFinancialCareers*, 26 Dec. 2017, news.efinancialcareers.com/uk-en/285249/machine-learning-and-big-data-j-p-morgan/.

21. *Qplum | Invest in Trend Following, Algorithmic Trading Strategies*, www.qplum.co/documents/deep-learning-trading.

22. *Deep Learning in Finance Summit London*, www.re-work.co/events/deep-learning-in-finance-summit-london-2018.

23. Heinz, Sebastian. “A Simple Deep Learning Model for Stock Price Prediction Using TensorFlow.” *Medium*, ML Review, 9 Nov. 2017, medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877.

Google AutoDraw was used to generate the picture in the cover. It’s a really interesting machine learning based online software that “predicts” what you are drawing - https://www.autodraw.com/