



POSITION PAPER

What can Machine Learning bring to investing?

07/03/16

Dr. Arnaud de Servigny¹
Elyes Ben Bouzid²

www.bramham-gardens.com

AGENDA

A) Introduction - What are the main differentiators between Machine Learning & traditional statistics?

B) Five fields of innovation:

1. Machine Learning & Smart Ranking
2. Machine Learning & Non Linear Patterns
3. Machine Learning & Abstraction / Adapting – the Deep Learning route
4. Machine Learning & The usage of soft data
5. Machine Learning & Network / Graph theory – a new approach to dynamic dependencies

C) Not being afraid of dealing with large datasets

D) Consequences in terms of investment strategies

¹ a.deservigny@bramham-gardens.com

² e.benbouzid@bramham-gardens.com

A) Introduction

There are still many sources of imprecision related to this new field of Financial Analytics³.

Some people think that it applies to high frequency data only and that it almost leads to disruptive effects on the functioning of markets by giving a competitive hedge akin to a form of front-running to the most equipped traders in the market.

First of all, it is wrong to closely link Machine Learning to high frequency trading. It is true that the assumption of efficient markets swallowing immediately new information looks a bit naïve and academics like Andrew Lo who are investigating the notion of adaptive markets⁴, work in the right direction in this respect. New techniques will help us to understand better the market dynamics. This being said, high frequency is not a big friend of sophisticated algorithms. In order to deploy some of the interesting new techniques that have recently been developed in Machine Learning, one needs some computational time, which is not fully consistent with high frequency trading.

As a consequence, it is the view of the authors of this note that Machine Learning will impact all trading frequencies.

The core objective of Machine Learning is prediction. In finance, there are two intertwined actors: investors and investments. One cannot look at time series purely as the outcome of physical / mechanical events. Such series incorporate a behavioral element that cannot be avoided. The particularly low signal-to-noise ratio that characterizes finance does not disappear because of the usage of these new machine-learning techniques. What these techniques bring though is the capability to adapt quickly to changing environments and patterns. This adaptive stance is the testimony of an increased modesty, in the sense that it is acknowledged that very few stable “laws of nature” can be found in finance, but conditional patterns are not out of scope.

After many years of wonderful promises around Artificial Intelligence, the effective delivery has been pretty weak so far. Neural Networks have not revolutionized decision making as a science. These algorithms were meant to compete with the human brain, but unlike it, they generally failed in two directions: learning from past history in order to adapt and create new neuronal nodes as per needed and articulate an adequate degree of abstraction in order to capture more conceptual patterns. Pinocchio never started to live by itself.

So after having articulated in a defensive manner what Financial Analytics is not: i.e. a new trick from high frequency trader, the next generation of econo-

³ We call Financial Analytics the new discipline using Machine Learning coupled with large datasets applied to finance and money management.

⁴ http://papers.ssrn.com/sol3/papers.cfm?abstract_id=602222

physics or else the second attempt to mimic the human brain in a somewhat unclever robotic manner, how can Financial Analytics help?

The first thing to note is that Machine Learning has been developed outside of finance with a mindset that differs fundamentally from the modern finance paradigm of efficient markets and asset returns driven by random walks. Again, what Machine Learning is trying to do is to predict future situations / events. The interesting by-product of this new endeavor is that unlike in finance, in Machine Learning there is no common technique taught to and used by all financial market participants irrespective of the organization they work for (no Ito Lemma, no Garch processes, etc.).

The second interesting aspect is that the world of Machine Learning is very much tied to open architecture. Knowledge is not frozen or academically ring-fenced, it spreads virally at a fast pace, much quicker than what academic journals generally allow. Codes are usually offered in simple languages R or Python. Working papers are quickly translated into open access packages.

Machine learning effectively starts where traditional statistics ends. Linear regressions, linear patterns, normalized linear regression such as logit, probit, etc. are part of the old world. What people look for in machine learning is related to non-linear patterns. In this vein of what is still supervised learning, models developed are called Support Vector Machine⁵, Random Forest⁶ and are combined with regularization techniques to force parsimony (Lars, Lasso, etc.) and Gradient Boosting techniques to speed up the convergence towards an optimal model. A by-product of this stream of innovation has been to introduce some form of hierarchical conditionality on factors, in the wake of Cart / tree models. Not to forget ensemble methods, i.e. averaging over different uncorrelated models to improve the overall performance.

Another stream of thinking has been relying heavily on Bayesian inference in order to avoid making assumptions on the shape of the patterns themselves.

Unsupervised learning has gained significant traction too. It is looking at gathering a complex system of data, and analyzing it over time in a way to be able to detect trends but also anomalies.

The developments related to computer vision have also contributed very much to the development of unsupervised learning in a direction that is commonly called Deep Learning. Deep Learning relates to an attempt to identify objects that require a certain degree of abstraction, such as being able to identify dogs of different breeds, different sizes, etc. Being able to tell that the animal one sees is a dog and not a cat is trivial for the human brain, but not for the machine.

Lastly, an important stream of activity related to Machine Learning consists in discrete ranking meant to maximize prediction. Google has been built using its

⁵ Vapnik, V. *Statistical Learning Theory*. Wiley-Interscience, New York, (1998)

⁶ Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). *The Elements of Statistical Learning* (2nd ed.). Springer. ISBN 0-387-95284-5.

“PageRank” algorithm and with it outplaying competitors. Netflix is known for its focus on optimal choice prediction. Amazon has been able to double the success rate of its recommendations versus more traditional peers. Microsoft Xbox relies on ranking players from bottom to top.

Predicting accurately expectations, desires, skills has become the name of the game.

Before getting into practical applications, there is a question that requires a precise answer: how complex are all these models? Finance is extremely concerned with black boxes, with overfitting and the first reaction is to avoid techniques that could look suspicious from that standpoint. There is another issue related to this, i.e. the risk of misselling: can we explain to clients in a sufficiently clear way what is at stake so that damaging misunderstanding can be avoided during the life of investments?

The answer to these questions is naturally somewhat subjective, but here is an attempt to provide a reasonable answer:

There is a trend in Machine Learning to replicate, on a large scale, on large datasets, simple operations: For instance a Random Forest is made of the aggregation of simple binary decision trees. Ranking methods are generally quite simple but they are trained and updated on very large samples. So from a conceptual perspective, in many instances, we should not talk of a black box. From an execution perspective however, the computing power required is often quite large, and it is important that it is so in order to maximize the robustness of the findings.

The fact that Machine Learning has come first in our day-to-day life before entering finance is a plus in the sense that for instance the ranking algorithm in the play-station from Microsoft Xbox which is called “Trueskill”, is not perceived as a black box. Google has overtaken Yahoo thanks to its algorithms and nobody questions its equations. The conclusion comes from effective experimenting.

This being said, in finance more than anywhere else, there is a certain degree of randomness, which makes not everything predictable. In addition, individual human behaviors and aggregated human behaviors are sometimes rather stochastic as Shefrin documented it, talking about the difficulty to characterize in a stable manner the utility function of a Representative Investor⁷. In other words, Machine Learning will significantly outperform traditional statistical techniques where there is predictability to be found, but will do equally poorly when there is not.

⁷ <http://www.amazon.com/Behavioral-Approach-Pricing-Academic-Advanced/dp/0123743567>

B) Five fields of innovation

It is hard to guess in which direction people will go in order to build new strategies. What follows is a first attempt by the authors to provide guidance on what looks possible, often based on their own trials and testing.

It should not be seen as a comprehensive / exhaustive summary of possibilities.

1. Machine Learning & Smart Ranking

In a financial world largely guided by indices, outperforming these indices has become a goal within the financial community. Many traditional techniques rely on equal weighting, maximizing diversification, using certain filters such as value, size, etc. Machine Learning should enable us to have a more detailed understanding of all the constituent of an index and therefore to select a subset of it based on a more refined filtering process than the crude features typically applied nowadays.

The world of sports: horseraces, football, tennis, basketball, chess are all focused on developing advanced ranking algorithms. There is a lot to leverage from this area.

It should be noted that the language used here differs from the features used in finance. In finance, the approach looks quite mechanical with the corresponding features that are used usually linked to financial indicators such as the price to book ratio or technical indicators such as momentum crossings.

In smart ranking, what people are after is rather resilient skill, offense or defense capabilities, a measure of the quality of the language associated with a firm on the web, etc.

2. Machine Learning and non linear patterns

Finding data factors that provide a resilient predictive power is often very hard. Many seasoned portfolio managers rightly question the simplistic algorithms that will make the price of an index or of a stock depend on the last issue of PMI numbers in the US for instance. They know that it sometimes matters but other times not. What a seasoned portfolio manager will do is to identify the cases when such an indicator is at play or not. Like a chess player he or she will anticipate various scenarios and conclude. Like in a chess contest, it is possible to create scenarios based on historical information and to infer complex decision trees. Machine Learning does this and does it in a manner to simplify things as much as possible in order to balance accuracy and robustness. Basically things boil down to a “meteorological” representation of financial markets and of aggregated human behaviors. We

have been able to see that machines using these types of algorithms now outperforms average as well as excellent human players⁸.

3. Machine Learning & Abstraction / Adapting – the Deep Learning route

Multilayered neural networks combined with Markov Chains offer a stylized replication of the human brain that look much more plausible than what has been done so far. The firm Numenta⁹ offers an interesting perspective in this direction.

Using this type of techniques, it should be possible to better understand how the market adapts to new information and to identify which asset is reacting more quickly than others to the information inflow.

This approach does not particularly require a high frequency stance, but rather focus on more complex situations than the traditional mean reversion effects, such as changes in regimes, etc.

4. Machine Learning & The usage of soft data

The availability of large streams of information is seen as a new Eldorado. In finance, there has been such a trend towards transparency that a lot of the traditionally used information is on Bloomberg or Datastream. By bringing new sources of data at a cost that is not prohibitive, access to information becomes a new target. Being able to cope with the fuzziness of this soft data requires adequate filters, but a lot of people are working on dealing with them in areas like marketing and more generally in the context of business analytics. Again here, the focus on investors rather than investments opens some interesting avenues, whereby investing is not primarily a matter of human biases but rather of human preferences.

5. Machine Learning & Network / Graph theory – a new approach to dynamic dependencies

Everyone in finance is fully aware of the limits of correlations. The attempt to broaden the topic in the directions of dependencies with copulas, or frailty models just seem quite naïve nowadays. When LinkedIn or Facebook look to understand group dynamics using network models, it is hard to think that the network of SP500 companies that share a relatively common shareholding structure, and that have strong commercial links cannot be approached as a network, both from a risk management perspective¹⁰, thereby creating systemic risk or from an investment perspective, looking for joint dynamic pricing movements.

⁸ <http://www.nature.com/news/google-ai-algorithm-masters-ancient-game-of-go-1.19234>

⁹ <http://numenta.com>

¹⁰ See for instance: <http://toreopsahl.com>

C) Not being afraid of dealing with large datasets

Finance has generally been fascinated with elegant closed-form solutions, which are parsimonious both in terms of data requirement and computational time.

The Machine Learning approach starts from a different angle, as it is precisely looking to make sense of data.

It would not be true to consider that just dumping raw data in large computers does the trick. The analysis and the transformation / filtering / de-noising of the raw data plays an extremely important role and constitutes a good part of the know-how of the analyst.

Once this is achieved, being able to understand what type of tool in the toolkit would best fit the sample in order to have it deliver sound prediction is part of the art. In this respect, global contest platforms such as Kaggle¹¹ show that there is a large field for expertise / competition there.

It is interesting to note in this respect that the combination and averaging of different techniques often improves performance.

At this stage, robustness matters. It can be too easy to overfit and reach some sort of in-sample perfection due to the breadth of the data used. Looking at things from a true out-of-sample perspective is critical. In this respect rerunning several times a model in order to get to good out-of-sample results is not state of the art. Out-of-sample data needs to be untouched and used with great parsimony.

The problem in finance is that datasets are always the same ones and are very well known. Rigor commands that the methods implemented should be reviewed with care (number of parameters / factors involved, speed of convergence of the models, intuitive results, etc.).

Overall, there are 3 requirements at play:

1. An in-depth understanding of all the factors / pieces of input used
2. The usage of adapted and intuitive modeling techniques given the nature of the data and the recourse to model averaging to increase the robustness of the results
3. A proper out-of-sample testing with a detailed analysis of the results, especially focused on crisis periods.

With these rigorous aspects in place, the interest of such strategies is that they can be implemented in real life using a back test with some credibility, although real life behavior remains the ultimate test.

¹¹ <https://www.kaggle.com>

D) Consequences in terms of investment strategies

We can see five consequences going ahead:

1. There will be more diversity in terms of strategies. At present there is quite some herding around traditional methods to allocate money. More recently aggregated investment styles have gained a lot of traction within the financial community (smart beta). We foresee more diversity in the future, with people operating at more granular levels, at different frequencies, uncovering complex and temporary patterns.
2. The fee model should evolve. As such strategies are largely automated end-to-end, running costs are much lower than traditional active management. As a result, the fee model should bend towards more performance sharing and less of yearly flat fees. The current environment of very low interest rates could prove a real booster for such new modes of remuneration.
3. The asset management industry will be impacted, with the high fee active management feeling the competition heat of the new passive-active model drawing from machine learning. The paradigm will shift away from paying for people to paying for demonstrable added value and skill.
4. As the methodological toolkit progresses, investment strategies based on Financial Analytics will evolve and improve. The job of money managers will be to monitor and evaluate who is innovative and state of the art. The growing pace of progress will mean that the pace required to make investment decisions will also have to accelerate, without being able to wait for 3 years of real track-record before starting to invest.
5. As long as passive and traditional active investors represent the bulk of investing, there is ample room for Financial Analytics.

Bramham-Gardens is a new and nimble Financial Research Lab, which operates in the field of machine learning applied to finance.

It has strong roots in Imperial College – London.

Its founders have a lot of experience in money management and in academia.

www.bramham-gardens.com