ACTA SCIENTIFIC COMPUTER SCIENCES

Volume 2 Issue 3 March 2020

Review Article

Machine Learning Driving Forecasting Paradigm

Bahman Zohuri[1]* and Farhang Mossavar Rahmani[2]

[1]*Research Associate Professor, Electrical Engineering and Computer Science Department, University of New Mexico, Albuquerque, New Mexico USA*

[2]*Professor of Finance and Director of MBA School of Business and Management, National University, San Diego, California, USA*

***Corresponding Author:** Bahman Zohuri, Research Associate Professor, Electrical Engineering and Computer Science Department, University of New Mexico, Albuquerque, New Mexico USA.

Received: March 03, 2020;
Published: March 10, 2020

Abstract

The future of any business from banking, e-commerce, real estate, homeland security, healthcare, marketing, the stock market, manufacturing, education, retail to government organizations depends on the data and analytics capabilities that are built and scaled. The speed of change in technology in recent years has been a real challenge for all businesses. To manage that, a significant number of organizations are exploring the Big Data (BD) infrastructure that helps them to take advantage of new opportunities while saving costs. As necessity of any business to be resilience, one needs Forecasting with a paradigm that fits to that business day-to-day operation using their incoming daily and timely information-driven by those data while comparing them with existing historical data to do Data Analytics (DA) and Data Predictive (DP) which will be derivative the observation of these data. Give the speed of incoming in real-time at sheer volume, leave us no choice but using Artificial Intelligence (AI) and consequently Machine Learning (ML) as its foundation and together with Deep Learning (DL) will enhance our predictive analytic to be to augment a forecasting model into our business to make it more resilience. In this article, we discuss these topics.

Keywords: Artificial Intelligence; Machine Learning; Deep Learning; Resilience System; Forecasting and Related Paradigm; Big Data; Fuzzy Logic

Introduction

As we as author have repeatedly expressed Power is Knowledge and in order to have knowledge, we need Information and for that information we need Data as depicted in figure 1 [1].

Timely transformation of information is also critical for the survivability of an organization by having the right information at the right time will enhance not only the knowledge of stakeholders within an organization but also providing them with a tool to make the right decision at the right time. It is no longer enough to rely on a sampling of information about the organizations' customers. The decision-makers need to get vital insights into the customers' actual behavior, which requires enormous volumes of data to be processed. We as authors believe that Big Data infrastructure is the key to successful Artificial Intelligence (AI) deployments. Although the is intelligent too, however, cannot make a human decision. So, its output is unbiased real-time. Big data solutions have a direct impact and changing the way the organization needs to work with help from AI and its components ML and DL.

Furthermore, providing a tool such as Forecasting Paradigm (FP) into hand of stakeholder and decision-maker on top of the capabilities as mentioned earlier allows them to be more resilience against their rivals and completion within domain of same business, therefore provides them with better Promotion Optimization (PO) including an innovative and smarter Key Performance Indicator (KPI) to measure their performance against their competition for a better survivability in the their today's competitive world [2].

Figure 1: Depiction of Data, Information, and Knowledge is Power in Four Dimensions [1].

Implementation a good and appropriate ML with right methodology and algorithm will provided such opportunity as a layer under their AI technology [3].

Most predictive modeling algorithms will take some number of observations as input and predict a single output value. As such, they cannot be used directly to make a multi-step time series forecast. This applies to most linear, nonlinear, and ensemble machine learning algorithms (Figure 2).

In today's growth of modern technology and the world of Robotics, significant momentum is driving the next generation of these robots that we now know as Artificial Intelligence (AI). This new generation is attracting tremendous attention from scientists and engineers. They are eager to move them to the next generation that is smarter and more cognitive, which we now call them Super Artificial Intelligence (SAI) [3].

Figure 2: Multi-Step Time Series Forecasting with Machine Learning.

Paradigm of forecasting

Our principle goal of writing this short communication article is to define a process and alimony of forecasting using machine learning that can help the decision-makers, implementing of the different methodology of forecasting techniques, not only to be able to forecast but even to be able predict any events that may be adversary or useful to the business. And its daily operation of any targeted industries, countries, and the world as well and as accurately as possible.

Understanding the data by using analytic methodology driven via machine learning allows us to use the right variables to be able to model the right forecasting paradigm as well as to identify any errors in our forecasting model. Moreover, to be able to filter any noise from our historical and incoming present data from every direction that enables our machine learning approach either in a supervised or semi-supervised manner through its layer of deep learning.

Most traditional forecasting models rely on fitting data to a pre-specified relationship between input and output variables, thereby assuming a specific functional and stochastic process underlying that process.

For any business to deal with these variables in an effective both from the cost point of view and performance, one needs to take a new approach to forecast by employing several machine learning algorithms, a method that is data-driven and imposing limited restrictions on the nature of the true relationship between input and output variables.

Bear in mind that the study of historical data is the basis for the analysis of trends, and there is a reasonable probability that the outcomes might not be as accurate as one expects, or even accurate at all. But the study in itself has and offers some value.

This exercise would enable the decision-makers and stakeholders with help from forecasting paradigm to better understand the environment in which they are working and learn about the strengths and weaknesses of their organizations and enterprise in dealing with an unknown future. Such capability may even result in preparing these actors by far out better to encounter such unknown variables; thus, the speed of processing of data within the real-time frame is essential to be ahead of the ball. Thus, we have no choice except to take advantage of Machine Learning (ML). In addition, in today's fast paste growing data at the level of Big Data (BD)we no longer can employ the Business Intelligence (BI), but we are left with the only choice that deploying Artificial Intelligence (AI) [3] with its machine learning capability as new and innovative tool (Figure 3).

Figure 3: Presentation Artificial Intelligence in Relation with Machine Learning and Deep Learning.

Every day businesses are facing the vast volume of data. Unlike before, processing this amount of data is beyond Master Data Management (MDM) to a level that we know it as Big Data (BD) that are getting around at the speed of the Internet. Since our daily operations within any organization or enterprise are expanding the Internet of Things (IoT) dealing with these data, either structured or unstructured, that is also growing at the same speed, thus processing these data for extracting the right information for the proper knowledge growing accordingly. The momentum existing behind such growing speed is way beyond the capability of traditional statistical forecasting, thus demands implementation of machine learning for most trusted forecasting and related paradigm.

With the demand of Power to make a decision with minimum risk based on Knowledge of Information from Data accumulated in Big Data (i.e., Figure 1) repository, we need real-time processing of the data coming to us from Omni-direction perspective. These data are centric around the BD and get deposited at the speed of Internet-driven mainly by IoT.

Thus, at this stage, we need to understand, what is the Big Data (BD) and why it matters when we are bringing the AI into play toward real-time processing of data with infrastructure around the Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. Big data usually has three characteristics. They are volume (the amount of data), velocity (the rate at which the data is received), and variety (types of data) [3].

Moreover, in most cases, the forecasting usually is a time series in real-time, where we need to apply machine learning with techniques such as Elastic Net, Super Learner, Recurring Neural Network (RNN) algorithms on macro data are among the seven, broadly representative, advanced and emerging of any targeted business. We find that these algorithms can outperform traditional statistical models from the processing perspective, thereby offering a relevant addition to the field of economic forecasting.

One of the most critical applications of Machine Learning Driving Forecasting Paradigm can be in any economy of countries with high GDP, where Forecasting macroeconomic variables are crucial to developing a view on a country's economic outlook. Understanding the current and future state of the economy integrating the historical and present data enables timely responses and policy measures to maintain economic and financial stability and boost resilience to episodes of crises and recessions [4-6].

Suppose you are taking under consideration and investigating as well as studying your company's financial health situation, measured quarterly utilizing multiple matrices. In this case, the data is a sequence of one or more values per time step. This approach is called a time series. Where in the financial world there exists multiple value per time step (e. g., the company's revenue, debt, and so one), such time series is called multivariate time series, while in case of one variable tracking events such as number of active users per hour on your website, or the daily temperature in your city is considered to be univariate time series.

A typical task is to predict future values, which we call it forecasting and, in contrast, is another common task could be the application to fill in the blank, to predict or rather "postdict" missing value from the past, and this is called imputation. Figure 4 shows 3 univariable time series, each of them 50-time steps long and the goal is to be able to forecast the value at the next time step represented by the X for each of them using a computer language technique such as Python [7].

Figure 4: Time Series Forecasting [7].

Looking at data in real-time and near real-time, we can clearly distinguish day and night for each 24-hour interval in any use case, would improve our observation of ongoing events and enhance us to see any discrimination that needs to be under consideration. When evaluating a prediction task on a time series like the above examples, we usually want to learn from the past and predict for the future. Such approach allows us to make the best supervised AI, ML, and DL integration into our forecasting paradigm.

In summary, a Machine Learning model can incorporate large amounts of data to forecast and allows the forecaster to be more enabled.

In the case of our example of economy, forecasting macroeconomic conditions can be very challenging and requiring forecasters to make many discretionary choices about the data and methods they use. Although forecasters underpin the choices they make about models and complexity with economic intuition and judgment, these assumptions can be flawed. Machine learning approaches, on the other hand, automate as many of those choices as possible in a manner that is not subject to the discretion of the forecaster.

However, bear in mind that, there are certain flaws associated with Machine Learning (ML) aggregating via its Deep Learning layer (i.e., Figure 3) that may have some acute impact in our model of forecasting as well [8].

Machine learning

Machine Learning (ML) is a subset of AI, and it consists of the techniques that enable computers to figure things out from the data and deliver them to the AI system (Figure 3).

Machin learning is a supplier of AI. It has been changed in the past few decades. As stated in the SAS cite5, "Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data.

The iterative aspect of machine learning is important because as models are exposed to new data, they are able to adapt independently. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that has gained fresh momentum".

In summary, Machine Learning (ML) can also be defined as the process of solving a practical problem by:

- Gathering dataset, and
- Algorithmically building a statistical model based on that dataset.

That statistical model is assumed to be used somehow and some way to solve the practical problem.

In other words, the present short communication here introduces an innovative approach to use cases such as macro-economic forecasting based on both existing data and incoming real-time data to be able to forecast and to contribute to the field of such use case or any anomaly cases as well. Establishing an algorithm using Python language capability with the functionality of the Predictive Interpolation (PI) factor, which is built into it, tailored explicitly for macro-economic forecasting and addresses non-linearity and non-ergodicity of the economy. The problem of high dimensionality in the context of scarce data is the way it should be going forward with the incoming technology of Artificial Intelligence (AI) as well as Super Artificial Intelligence (SAI).

As we have stated before, Most predictive modeling algorithms will take some number of observations as input and predict a single output value. As such, they cannot be used directly to make a multi-step time series forecast. This applies to most linear, nonlinear, and ensemble machine learning algorithms

With the advent of big data, both the amount of data available and our ability to process it has increased exponentially. The ability of machines to learn and thus appear ever more intelligent has increased proportionally. Even so, machines aren't independent thinkers (yet).

Yes, machine learning may identify previously unidentified opportunities or problems to be solved. But the machine is not autonomously creative. The machine will not spontaneously develop new hypotheses from facts (data), which is not evidence. Nor can the machine determine a new way to respond to emerging stimuli. This is due to the fact today's AI nor NR are thinkers without human in the lope.

Remember: the output of a machine learning algorithm is entirely dependent on the data it is exposed to. Change the data, change the result.

Conclusion

A variety of forecasting methods often apply to any particular supply chain scenario. Smart supply chain planners use multiple forecasting methods tuned to perform well at different phases of a product life cycle, chosen to exploit best the available historical data and degree of market knowledge.

The key is to pick the most effective and flexible forecasting models, blend their best features, and shift between them as needed to keep forecast accuracy at its peak. This very popular white paper delves into the details of eight forecasting methods, including why, when, and how they should be used to realize the greatest overall improvements in forecast accuracy.

All these conclusions would be most effective if we choose the most innovative and effective machine learning tool, along with deep learning and artificial intelligence.

Bibliography

1. Z Anwar., *et al*. "Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: A brief review". *Journal of Radiation Research and Applied Science* 7 (2014): 163-173.
2. Y Zheng., *et al*. "Overview of biomass pretreatment for cellulosic ethanol production". *International Journal of Agricultural and Biological Engineering* 2.3 (2009): 51-68.
3. M Foston J., *et al*. "Biomass characterization: recent process in understanding biomass recalcitrance". *Ingram Micro* 8 (2012): 4-22.
4. JS Kim., *et al*. "A review on alkaline pretreatment technology for bioconversion of lignocellulosic biomass". *Bioresource Technology* (2015).
5. H Wulfhorst., *et al*. "Tippkeotter, Compositional analysis of pretreated (beech) wood using differential scanning calorimetry and multivariate data analysis". *Tetrahedron* (2016): 1-6.
6. SR Decker., *et al*. "High-throughput screening of recalcitrance variations in lignocellulosic biomass: total lignin, lignin monomers, and enzymatic sugar release". *Journal of Visualized Experiments* 103 (2015): e53163.
7. J Lindedam., *et al*. "Evaluation of high throughput screening methods in picking up differences between cultivars of lignocellulosic biomass for ethanol production". *Biomass and Bioenergy* 66 (2014): 261-267.

8. ASluiter B Hames., *et al*. "Determination of structural carbohydrates and lignin in biomass laboratory analytical procedure". Golden, CO: National Renewable Energy Laboratory (2008).

9. FC Moreira-Vilar., *et al*. The acetyl bromide method is faster, simpler and presents best recovery of lignin in different herbaceous tissues than klason and thioglycolic acid methods (2014).

10. Ragauskas AJ., *et al*. "The path forward for biofuels and biomaterials". *Science* 311.5760 (2006): 484-489.

11. Yan L., *et al*. "Hot water pretreatment of lignocellulosic biomass: an efective and environmentally friendly approach to enhance biofuel production". *Chemical Engineering and Technology* 39.10 (2016): 1759-1770.

12. Pu Y., *et al*. "Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments". *Biotechnol Biofuels* 6.1 (2013): 1.

13. Meng X., *et al*. "An in-depth understanding of biomass recalcitrance using natural poplar variants as the feedstock". *Chem Sus Chem* 10.1 (2016): 139-150.

14. Chang VS and Holtzapple MT. "Fundamental factors affecting biomass enzymatic reactivity". *Applied Biochemistry and Biotechnology* 84 (2000): 5-37.

15. A Sluiter., *et al*. "Determination of structural carbohydrates and lignin in biomass laboratory analytical procedure". Golden, CO: National Renewable Energy Laboratory (2008).