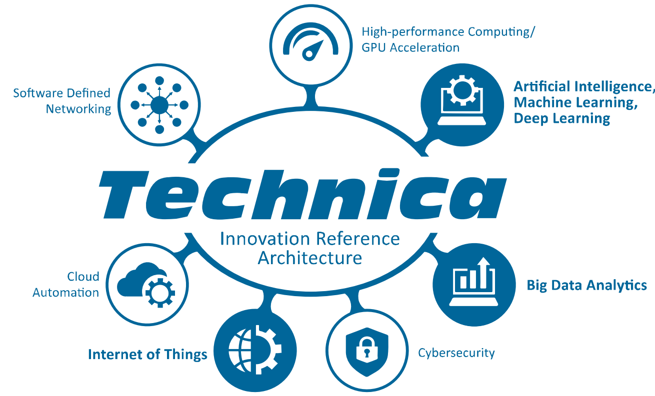
Technica™

Exponential growth of IoT Big Data is coming. This paper highlights improvements in the predictive capability of time series analysis when combined with deep learning's neural networks. Technica is implementing time series analysis with deep learning into FUNL, its GPU-accelerated graph analytics engine.

But… what is time series data? Why is it important? And how can it provide a way forward for the volume, variety, and velocity of IoT data?

The Technica Innovation Platform White Paper Series presents advanced topics that will drive competitive advantage for next-generation IT over the next three-to-five years.



TIME SERIES ANALYSIS (TSA) AND DEEP LEARNING

Time series data records quantities that represent or trace the values taken by, a variable over a period such as a month, quarter, or year. In other words, time series data describes *something* over time. We are surrounded by time series data. What makes time series data so important is that it allows one to see not only a single quantity in isolation, but at the same time, it gives us the typical "context" for the measured amount.

Other white papers in this series have described advances in Graphics Processing Units (GPU)-accelerated deep learning. This paper builds upon that topic and highlights time series analysis with deep learning's neural networks. The paper will also highlight Technica's FUNL product, a GPU-accelerated graph analytics engine, which is in the process of implementing time series analysis with deep learning.

With the coming exponential growth in IoT Big Data, time series analysis will become more important. At the same time, analyzing the volume, velocity, and variety of IoT Big Data will become an incredible challenge. Time series analysis with deep learning offers the ability to improve time series analysis and strengthen its predictive capability, incorporating Big Data.

TIME SERIES DATA EXAMPLES

- Enterprise operational data such as monitoring data, performance metrics, log events, or CPU utilization
- Financial data like stock charts or the Dow Jones Industrial Average
- Climate data: Temperature, Carbon Dioxide (CO_2) levels in the atmosphere, sunspot activity, etc.
- Health care information such as blood pressure and electro cardiograms plotted by doctor visits or the number of births in a city
- Activity tracking such number of steps during a given day
- Sensor data from all manner of Internet of Things (IoT) devices like auto traffic monitoring devices, electricity meters, and sensors that report the maintenance information on planes, trains, ships, trucks, and automobiles
- Web tracking such as Clickstream data

TRAITS OF TIME SERIES DATA

A time series is a sequence of measurements or observations of a system that varies in time. **Figure 1** presents a six-month stock chart for Facebook, Inc. Time is portrayed on the X-Axis, while the Y-Axis lists the price in US dollars. The value of $129.96 is a discrete variable—the closing of price of Facebook stock every day. Other time series charts might present continuous values from a perpetual signal, e.g. real-time values from second-to-second.

Figure 1 – Facebook Stock Price
(Source: stockcharts.com)

Major components of nearly every time series are *trend, seasonality,* and *noise*. The trend in Figure 1 is positive towards price over time—there is an increase in price during the last six months.

Figure 2 presents time series data recording the number of call to an enterprises call center during a nearly three-year time period. It is clear from examining Figure 2, that the chart displays seasonality—there is a regular, periodic pattern over time. Namely, call center activity decreases in the March to May timeframes. Figure 2 also highlights noise—some form of random variation. For example, in the second November, there were over 45,000 calls to the call center. Through various algorithms performed on the time series data the results can be smoothed. The dark line in chart presents the smoothed values.

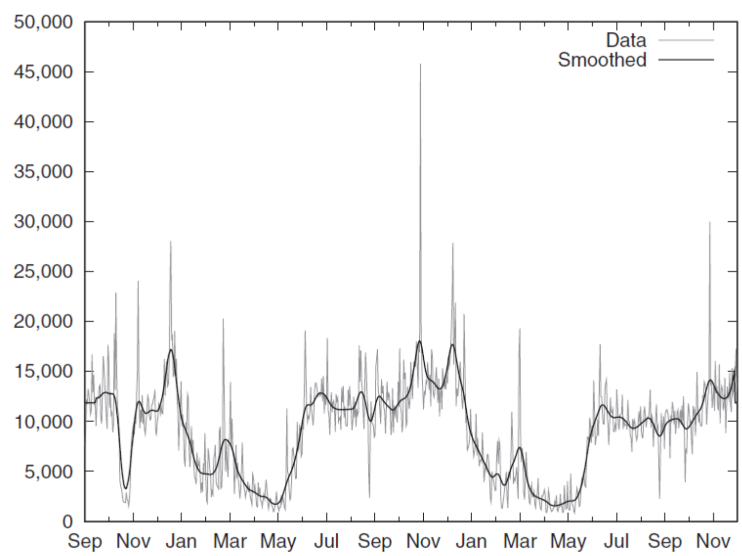


Figure 2 – Calls Placed to Call Center
(Source: Data Analysis with Open Source Tools, by Philipp K. Janert)

While entire courses are taught on time series data and analysis, it is important to get a flavor for some basic terminology and standard operations on time series data.

Figure 3 highlights some of these concepts. Multivariate time includes measures of more than one variable, for example the price of Facebook, Twitter, and General Motors could be compared. Time series tensors are more complex, but in a nutshell describe linear relationships between geometric vectors, scalars, and other tensors. Observations are points along the line or curve. Finally, sets of time series data can be gathered and compared.

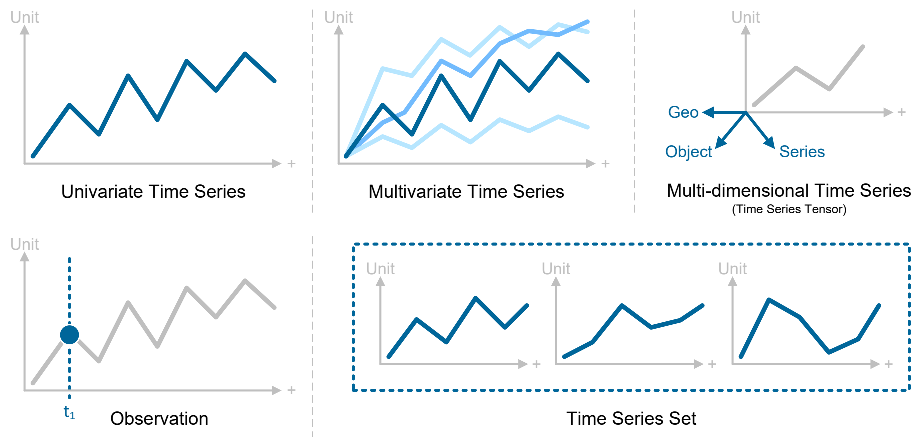


Figure 3 – Basic Terms for Time Series

Figure 4 portrays some standard operations on time series data.

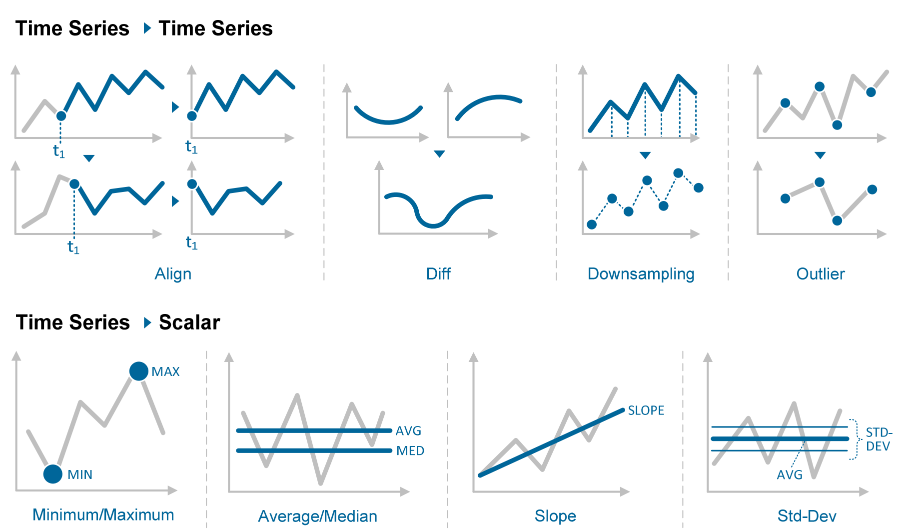


Figure 4 – Illustrative Operations on Time Series

While it is beyond the scope of this paper to explain all the operations, it is apparent that numerous operations can be performed on time series data to glean insight and turn data in to actionable knowledge.

TIME SERIES ANALYSIS (TSA)

Given the traits of time series data, what does TSA entail? The process of analyzing time series data can be broken down into three basic tasks:

- **Description** – Seeks to identify the components of the time series including trend, seasonality, and noise
- **Prediction** – Given current values, this task seeks to forecast future values
- **Control** – Includes monitoring the values over time in light of the TSA predictions

TSA involves examining the past with the task of description, while prediction looks to the future, and control concentrates on the present. With this background on time series data and TSA, we now turn our attention to deep learning.

DEEP LEARNING

The previous sections of this document highlighted TSA within the framework of conventional statistical methods. Motivated by the human brain's biological functions, deep learning is a rapidly growing field within machine learning to perform predictions, classifications, and pattern recognition. To date, the biggest use cases for deep learning have been in image recognition/ classification and speech recognition. However, the utilization of deep learning for TSA is very promising.

Neural networks are at the core of deep learning. There are many different architectures and permutations for neural networks that are optimized to perform a given task. Most of the advances in image recognition have been accomplished with convolutional neural networks (CNNs). Recurrent Neural Networks (RNNs) have thus far proved best for TSA.

The benefits of applying deep learning to TSA are tangible. Error rates from predictions based on deep learning can be reduced as compared to standard statistical models. The need for domain experts is lessened, because—over time—predictions from deep learning can out-perform models based on conventional statistics and expert models. For example, finance (price prediction) is one of the greatest employers of deep learning experts.

TSA AND IOT

As mentioned in the Introduction, time series data is everywhere. The biggest and most advanced use case in the future will be the application of time series analysis to IoT data.

A revolution in sensor–to–insight data flow is rapidly changing the way we perceive and understand the world around us. Much of the data generated by sensors, as well as a variety of other sources, benefits from being collected as time series.

Although the idea of collecting and analyzing time series data is not new, the astounding scale of modern datasets, velocity of data accumulation, and variety of new data sources present a huge challenge. TSA in conjunction with deep learning offers an innovative way forward.

Technica is in the process of adding time series analysis with deep learning to FUNL. FUNL provides analytics with GPU-accelerated machine learning and graph analytics. The company is a partner of NVIDIA and knowledgeable with NVIDIA GPU programming language, called CUDA. While other architectures can be supported, Technica has employed a Long Short Term Memory Recurrent Neural Network (LSTM-RNN) model within the latest development version. The network is trained on historical time series data to provide predictive analytics.

An impediment to usage of neural networks for TSA, is that predictive success is highly dependent upon fine-tuned parameters. This knowledge is only possessed by deep learning experts. FUNL's TSA implementation enables a data analyst unfamiliar with deep learning or neural network architecture to perform time series analysis. It hides the complex details of neural network parameters, allowing the data analyst to only think about the predictive data.

SUMMARY

Time series data is ubiquitous. The volume, variety, and velocity of time series data from IoT demands new methods for TSA. GPU-accelerated deep learning is one such method. With improved predictions, the utility of TSA on all aspects of enterprise computing will have unforeseen, but revolutionary impacts upon the enterprise. Technica's implementation of deep learning for TSA within FUNL enables the enterprise to leverage the coming data deluge from IoT Big Data. This can be done without having to develop in-house deep learning algorithm expertise.

Technica provides professional services, products, and innovative technology solutions to the Federal Government. We specialize in network operations and infrastructure; cyber defense and security; government application integration; systems engineering and training; and product research, deployment planning, and support.

Technica™
22970 Indian Creek Drive, Suite 500
Dulles, VA 20166
703.662.2000