White Paper

Just as Humans Use Words to Communicate, Machines and Events Talk in Time Series Data!

## WHY BUSINESSES SHOULD UNDERSTAND THE LANGUAGE OF TIME SERIES DATA?

EXECUTIVE SUMMARY

● Time series data is the fastest growing data in the world today

● Time series data is being collected in huge volumes by sensors, transactional systems, events logs, video streams, audio recordings, images, and many more sources

● Time series data records very detailed footprints that reveal root cases or opportunities

● Every company has massive volumes of underutilized time series data

● The invaluable insights locked in time series data can be used to save costs and optimize revenues

● The time series insights can be monetized by monitoring for recurring patterns in real time

● Business professionals can use Trendalyze self-service tools to discover, search, and monitor for time series patterns without having to rely on data scintists to extract the insights or to IT to build the monitoring applications

TREND ALYZE

The Fastest Growing Data Is Time Series Data

Today sensors and connected smart devices are generating vast amounts of data about machine operations, business processes, human behaviors, environmental and other conditions. Most of this data, whether it comes from monitoring devices, transactional systems, logs, video streams, images, and other sources, is or can be converted to time series data. Time series data is visualized on line charts (such as stock trading or EKG charts) which make it easy to see how patterns change over time. Business professionals are aware of the benefits of visualizing and analyzing data patterns over time as trends can alert them of dangers they can avoid or opportunities they can capitalize on.
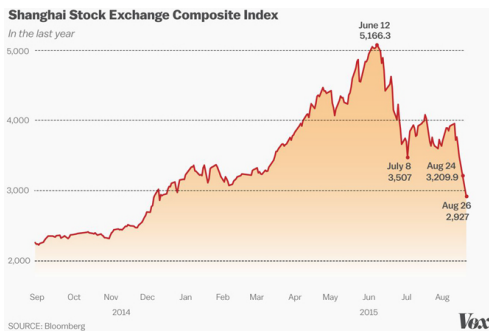
Stock Market Downward Trend
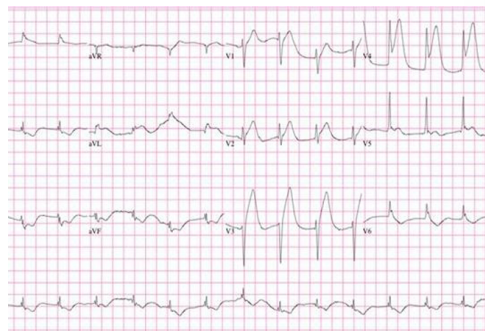
ECG/EKG for Heart Attack

Time series data has been collected and analyzed for many years but not at the level of detail that we get today. Due to advances in technologies time series data is collected at very granular intervals – milliseconds, seconds, minutes, hours. This makes the time series very long, having millions of data points that do not fit on a single screen. Furthermore, the data contains many dimensions that describe it, such as, for example, stores and products in retail; patients' age, gender and demographic attributes; machine operating conditions; etc. When all dimensions are taken into consideration, there are frequently millions of time series to be analyzed.

The biggest challenge for businesses is how to extract insights from the millions of time series. To do so, each time series must be analyzed individually. But the sheer volume of data makes this impossible for humans to do. This is where Trendalyze's new motif discovery approach to time series analytics comes into play.

Why Time Series Data Is Critical for Your Business

Today sensors and other data collection technologies capture the detailed footprints of every phenomenon. Like DNA, these detailed footprints contain the root causes of events that we want to understand and control. In fact, DNA data itself can be converted to time series data where machine algorithms can quicklydiscover recurring

patterns, called motifs, that explain the causes of many diseases and also predict predispositions to diseases.
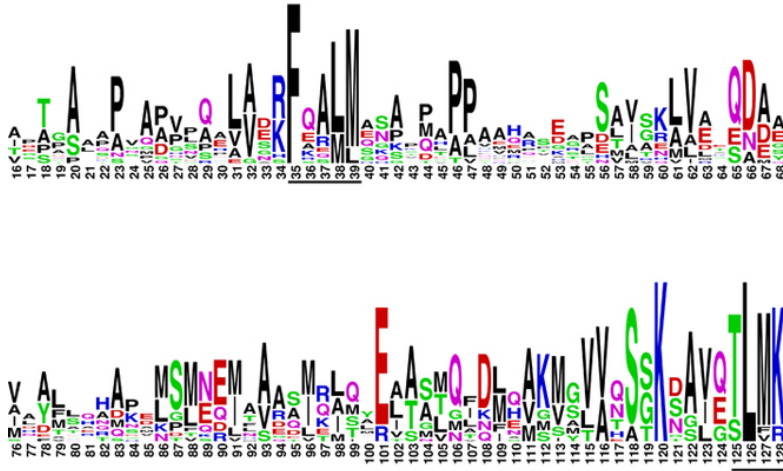
DNA as a Time Series

When business professionals can detect and analyze easily the footprints of the processes and events that they manage, they gain valuable insights. When they can search and monitor for valuable motifs within millions of time series in real time, they can monetize the insights by achieving better outcomes. In the past, business professionals depended on statisticians and data scientists to build complex models to extract insights from the data. They also depended on IT professionals to build the monitoring applications to improve their decision making. Trendalyze's Google-like search approach to time series pattern discovery puts the power of discovery and monitoring straight into the hands of the business professionals who make the decisions.

The Similarity Between Trendalyze Motif Discovery and Google Search

Motif discovery is extremely easy to use by business professionals as it works like Google search. Humans use words to talk and write, and we use Google daily to search for relevant information within billions of web pages on the web. Google takes words as search terms and finds pages that match the search criteria.

You can think of words as a type of time series pattern. Each letter in the alphabet has a sequential number - A is 1, B is 2, and Z is 26. When you replace the letters with their sequential number, each word becomes a time series data that can be plotted on a line chart and the shape of the numeric motif can be used as a search criteria.

Just as humans use words to communicate, machines and events talk in time series data. Each motif within the time series data has a meaning. A downward trend within stock data may indicate a market crash; within medical data, it can indicate a deteriorating condition; within machine operations data, it can indicate a failure.

TRENDALYZE

By searching and monitoring for motifs, business professionals can find relevant information about processes and events. For example, a Holter ECG heart monitor records millions of data points about heart rate and rhythm over a period of 12, 24 or 48 hours. Since there are about 40 known motifs for various heart conditions, it is easier and faster to search for these motifs in order to diagnose a patient.

By understanding the language of time series data, business professionals can retrieve or receive meaningful information from machines, conditions and events monitors, and use this information to improve outcomes, save costs and increase revenues.

The Difference Between Motif Discovery and Machine Learning

As we mentioned earlier, machine learning (ML) has traditionally been used to analyze trends within time series data but it is complex, time consuming, costly, and frequently inaccurate. The method applies machine algorithms onto large volumes of historical data to train models to recognize patterns within new data. To facilitate the learning, the historical data requires some preparation which takes time, effort and data science expertise. For example, data scientists have to flag in the historical data all the things that the model will predict in the future -- a transaction is flagged as fraudulent or not; a customer is flagged as a churn or not; etc. The success of an ML model depends on its accuracy, i.e., how well it predicts. The lower the accuracy, the more predictions will be false and false predictions drive up business costs. If a model predicts 50% of the time falsely that a particular transaction is fraudulent, that doubles the fraud investigation costs. Since real life data contains a lot of variation, accuracy is hard and costly to achieve.

Here is how motif discovery alleviates some the problems associated with machine learning:

1) Model training vs no training: Machine learning requires large amounts of historical data to train the predictive model. On the other hand, motif discovery can use just one historical observation, i.e., the first observed failure of a machine, or even a hypothetical pattern, i.e., a hand-drawn pattern leading to a machine failure, to search for recurring motifs. Training is not required because motif discovery looks for matching patterns. As in Google, if you provide a search term, the engine will find the matching shapes.

2) Model fit vs match: Machine learning fits all historical data onto an extrapolated (think of it as average) line that represents all data points. Hence, the model accuracy depends on how close the fitted line is to the original data. The higher the variation, the lesser the accuracy of the predictions. And real time data tends to have a lot of variation, which is why so many models fail to deliver ROI. On the

TRENDALYZE

other hand, matching is quite precise. Matching compares motifs and evaluates mathematically the similarity of their shapes. Like Google, the motif search results are ranked based on the relevance, i.e., the similarity of the matched motifs to the search motif.
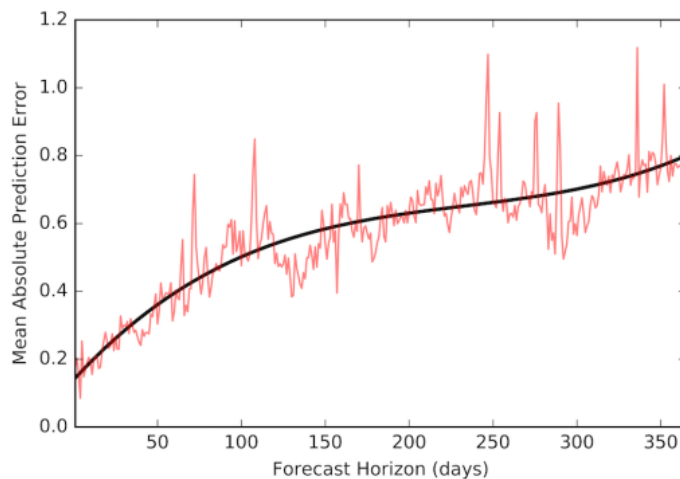
The fitted line by the model vs the error
due to variation in the data

3) Model refresh vs. library updates: Models are always temporary as conditions and their underlying data patterns change over time. Thus, to maintain the model accuracy as new data comes in, models have to be retrained periodically which is both costly and time consuming. Alternatively, business professionals discover and update motifs continuously in the course of their daily work and simply update the Trendalyze motif library. The Trendalyze motif library is a knowledge sharing platform that facilitates the learning and enrichment of domain-specific time series patterns.

4) Data scientists vs. business professional: Machine learning requires statistical knowledge to prepare and transform the data, to estimate and fine tune the model, and much more. However, business professionals can spot motifs with Trendalyze's interactive tools or with our machine profiling without any additional training. Who doesn't know how to use search? With its new approach, Trendalyze puts the power of patterns discovery and prediction directly into the hands of business professionals.

Trendalyze is Easy to Use, Scalable and Cost Effective

Trendalyze is an easy-to-use web-based platform with interactive visualization and data exploration tools. It allows business professionals to quickly find the motifs (also called micro trends) that lead to outcomes they want to understand and manage. The platform can ingest and manage massive volumes of time series data, all the while using a simple Google-like search function to discover motifs across all business dimensions.

Trendalyze gives businesses the power to improve outcomes, save costs and optimize revenues -- all based on the discovery of micro trends captured in huge volumes of detailed data. It's easy to monetize micro trends, while using predictive pattern detection to watch for problems before they happen.

How Time Series Patterns Tell the Lies on the Lie Detector
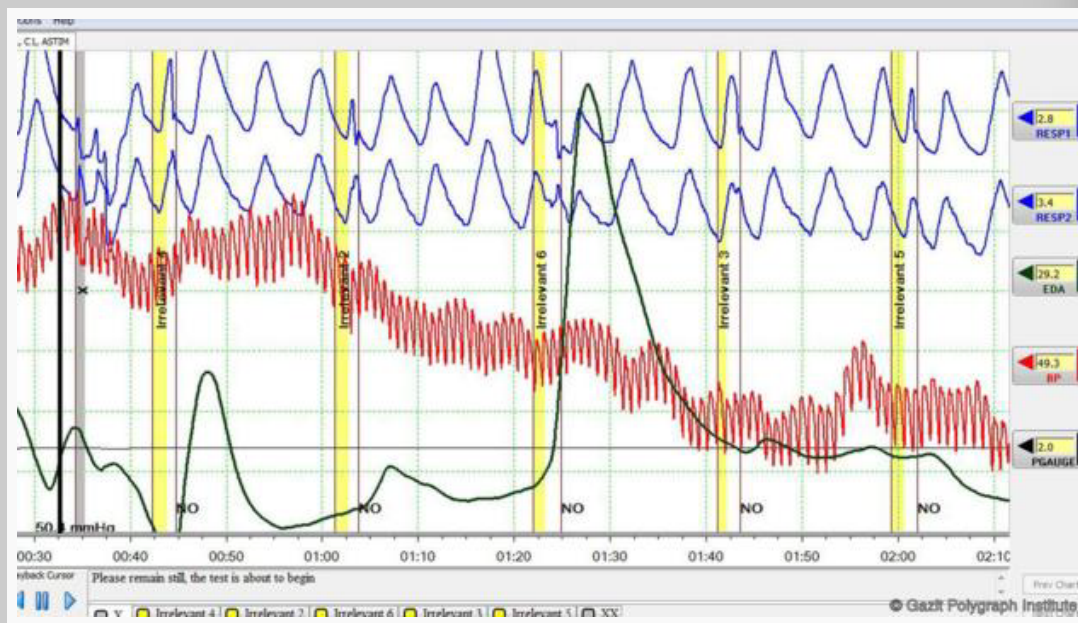
A lie detector is a technology that records some physiological functions of a person answering questions to ascertain the truth or falsehood of the person's statements. The above graph shows the readings of the polygraph test which are displayed as time series. The x-axis on the graph indicates time. The yellow columns indicate the questions being asked. Each time series represents the readings of some physiological function: the blue lines represent respiratory rate; the red line represents pulse and blood pressure; the thin black line represents physical movement in the chair; the thick black line represents galvanic skin response. It change of the pattern of the thick black when question six was answered indicates a possible lie.

Adapted from "Can You Beat a Lie Detector?" By Tiffanie Wen, featured on BBC.com