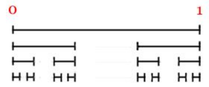
A Novel Approach for Assessing the Predictability of Time-Series

Peter MacNeil, Colin Hascup, Yuesong Huang, Zhelin Sheng
Math & Data Science Dept University of Rochester

Introduction

Time series data is ever present in today's world. From forecasting the weather, retail sales, or the stock market, it is important to understand what the characteristics are of time series that can be forecasted with a high degree of accuracy using modern techniques of forecasting. In this research, we explored different "metrics of forecastability"; values calculated from the original time series data that help us understand the future forecastability of a time series. Two industry standards, Q1, and Q2 are currently used for such purposes, and we had a goal of beating the performance of these metrics. The time series data used to draw conclusions from was retail sales data from a large retailer. Fractal behavior is often found in time series data that has seasonality, such as sales data, and this motivated us to explore a fractal metric from a previous paper entitled "Fractal dimension, approximation and data sets". This metric called discrete s-energy is used to approximate the fractal dimension of a data set for purposes of dimensionality reduction. The fractal features often present in time series data motivated us to test a discrete s-energy as a metric of forecastability and compare against the industry standard as well.

Motivation for exploring fractal behavior in time series data:
Suppose that we have a time series representing sales of a retail store going back 40 years. Suppose that one wished to look at the times when the sales were in the top 5% of all sales in a given year, and it turned out that this happened every July, December and April, and that during those months it happened during the first week, and during that week it happened on Fridays and Saturdays, and that on those days, the sales peaked in the mornings. As you can see, this structure is highly reminiscent of Cantor set construction:



While this type of a phenomenon has been extensively studied in terms of seasonality, we believe that a fractal perspective can be of considerable value in view of the fact that if the specific months, weeks, days, and times in the hypothetical above change, while their relative number remains roughly the same, the seasonality considerations no longer apply, while the fractal dimension analysis is still valid and effective.[1]

Methods

Employed various metrics to see how complex different sets of time series were.
Metrics Used:

Industry Standard metric - Minimum of Q1 and Q2 where Q1, Q2 and μ_Δ are:

$$Q_1 = \frac{(\frac{1}{N} * \sum_{i=1}^{N} (f(i) - \mu)^2)^{1/2}}{\mu} \qquad Q_2 = \frac{\sqrt{\frac{1}{N-1} * \sum_{i=1}^{N} (|f(i) - f(i-1)| - \mu_\Delta)^2}}{\mu_\Delta} \qquad \mu_\Delta = \frac{1}{N-1} * \sum_{i=2}^{N} |f(i) - f(i-1)|$$

CKS Complexity Proxy - CKS complexity, gauges the complexity of strings based on the shortest amount of turing code that can generate them. However, directly computing CKS complexity is notoriously difficult, and impossible in infinite cases. As a workaround, We employed LZMA and ZLib compression as an approximate measure. LZMA and ZLib, are high-ratio compression algorithm, reduces sequences by identifying patterns. If a sequence compresses significantly, it indicates lower inherent complexity. Conversely, less compression may suggest greater complexity.

Discrete-S Energy - Since all the time series data is of dimension >= 1 , Discrete 1-Energy was employed. We first had to compress the time series to [0,1] and then calculated

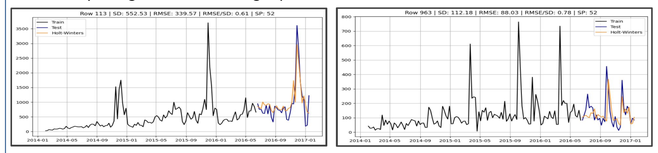
$$I_s(P_n) = n^{-2} \sum_{p \neq p' : p, p' \in P_n} |p - p'|^{-s},$$

Where Pn is a finite point set in $[0,1]^d$, in this case d = 1, with s = 1 and n = length of time series.

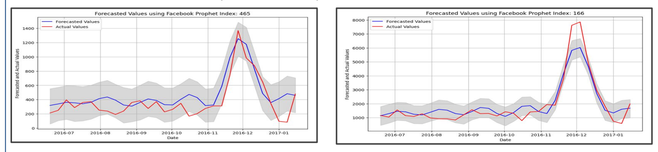
Methods Cont.

Then we predicted the Time Series using well known forecasting methods.

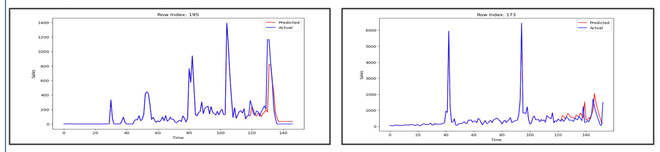
Methods Used:
Triple Exponential Holt-Winters Model - a time series forecasting method capturing level, trend, and seasonality using three smoothing equations.



Facebook Prophet - an open-source forecasting tool by Facebook, handling seasonality, holidays, and trends with additive or multiplicative components.

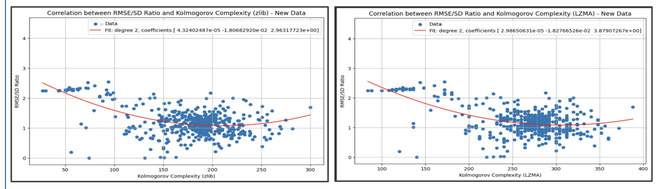


LSTM Neural Network - a recurrent neural network variant, using long short-term memory units, excelling in capturing long-term dependencies in time series.

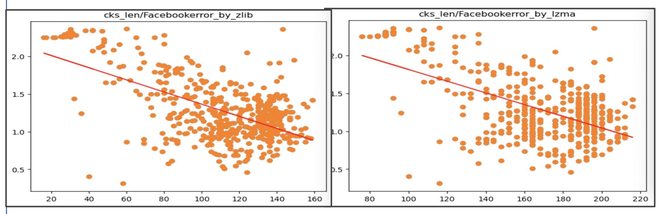


We then graphed the Correlation between the RSME/SD of the forecasting method and the CKS complexity proxies using both ZLIB and LZMA.

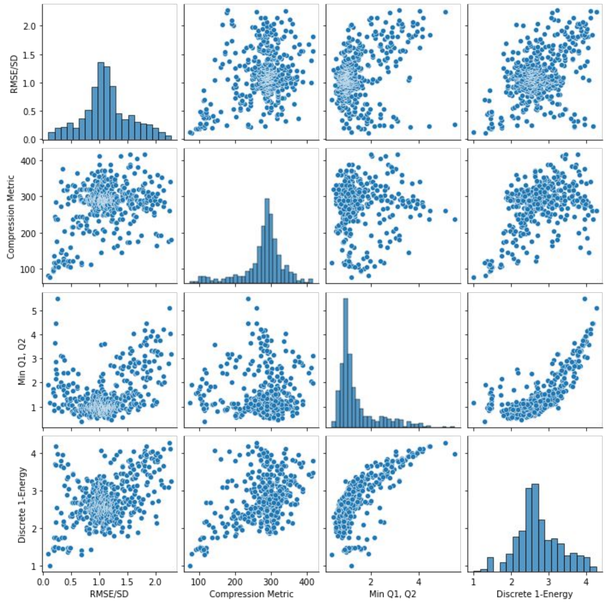
Triple Exponential prediction error correlation to ZLIB (Left) LZMA (Right):



Facebook Prophet prediction error correlation to ZLIB (Left) LZMA (Right):



Results



We run a linear regression (OLS) to find the the correlation between all the variables (discrete s-energy, industry standards, Proxy CKS Compression Metric, RMSE/SD of the best forecasts.

We can see some correlation between the following two variables: discrete s-energy and industry standards (with correlation coefficient 0.588) ; discrete s-energy and Proxy CKS Compression Metric(0.215); discrete s-energy and RMSE/SD (0.211); Proxy CKS Compression Metric and RMSE/SD (0.114); industry standards and RMSE/SD (0.132).

Conclusions and Discussion

The correlation between discrete 1-energy and RMSE/SD is the largest among other two tools, Proxy CKS Compression Metric and industry standards, indicating that a relatively high discrete 1-energy is the most indicative feature that a time series (of our type) will not be forecastable. All the metrics are correlated with each other, except the CKS Proxy and the industry standard

Future plans:

1. Given such a high correlation between discrete s-energy and industry standards, we are interested in further assessing relationships between industry standards and discrete s-energy. What is the underlying theoretical relation?
2. The time series data used in this research is limited in scope, and we must assess these metric's performance on other types of time series data.

References

[1] Betti, L., et al. "Fractal dimension, approximation and data sets." arXiv preprint arXiv:2209.12079 (2022).