Introduction

- A Time Series is a collection of observations, capturing a variable of interest, over time.
- Analysis accounts for the fact that data points collected over time, have an internal structure: autocorrelation, trend, seasonality, outliers, etc. (Figure 1)
- Primary uses of time series analysis include Sales Forecasting, Stock Market Analysis, Census Analysis, Anomaly Detection, etc.
- Aim to develop an open-source python library to facilitate a new Subset Scanning approach for accurate, time efficient, and computationally inexpensive event detection for large data sets.
- Existing algorithms show high accuracy and efficient optimization over subsets, requiring only linear Time Subset Scanning.
- When spatial component exists, the implemented ‘Proximity-Constrained Subset Scanning’ approach substantially improves the computation time and result accuracy for spatial event detection.
- Utilize the designed library to assist in problem solving.

Figure 1: Stock prices for two of the technology giants. Generic time series data has an interest variable recorded in order of time, and past values are assumed to effect the current and future value.

Spatio-Temporal Event Detection

- Work with spatio-temporal data to identify the most anomalous subset(s) for the stream(s) of interest.
- For the Univariate case, we detect the subset of locations having the highest anomaly score, based on a scoring function, a priority function, and a max subset window size (given by the user).
- For the Multivariate case, we follow the same process as Univariate, But also incorporate the ‘Coordinate Ascent’ algorithm, to find the local maxima.
- This can give considerable leverage in detecting disease outbreaks before they occur, almost as early as 2-days in advance.

Figure 2: Visualization of the Spatial Temporal data. Every cell holds the value for the corresponding Location, Time and Stream indices.

Effect of Piracy on Box Office

- Motion picture studios care about how pre-release piracy affects Box Office sales of movies.
- Initial results have shown that pre-release piracy leads to a 19.1% decrease in Box office sales, as compared to post-release piracy. But, this is not homogenous, and depends on the quality and timing of the piracy leak.
- Aim to find which types of leaks, at which times, are most harmful, and identify if there are any positive effects of piracy enjoyed by movie studios.
- Aim to tap into other relevant latent variables, so as to explore the relationship between piracy and box office sales, other than causal relation.

Future Work

- Combine open source piracy consumption data to gain additional predictive power, beyond information carried by the results from past similar movies.
- Extend the analysis to revenue from Home Entertainment Systems.

References