Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

Introduction

Implementations of classical statistical methods such as Holt-Winters filtering are widely available in statistical packages and open source software. Often these methods appear (at least visually) to model data effectively. However, the complexities of real time series data often render these methods ineffective for accurate anomaly detection and prediction. In this technical brief, we explore two approaches to time series data modeling and anomaly detection. We discuss how anomaly detection on real-world time series data requires specific modeling capabilities, and we compare anomaly detection results obtained with each approach.

Example Time Series Data Set with Anomalies

Figure 1 below shows an application transaction count obtained from an online retailer over a 7½ week period. A key use case for this retailer is gaining early notification when this transaction count is unusual. Primarily employed to detect revenue-impacting conditions or outages, the retailer also values understanding any time the transaction count significantly deviates from its 'normal' baseline.

Note that from a data science perspective, this data set is really quite simple. It's a univariate time series with daily and weekly periodicity. That said, it's a good example data set for our comparison, because it's easily understood. As an added benefit, we've had one of the retailer's operations analysts "mark-up" the data with the actual anomalies present in the data. The set of marked-up anomalies in a data set is sometimes called the "ground truth" anomaly set. The analyst identified four important anomalies:

1. Abnormally high transaction rate on the Thursday in the fourth full week of data
2. Abnormal high nightly minimum rates over the entire fifth full week
3. Brief but significant drop in transaction rate on the Thursday in the fifth full week
4. Unusually low rate compared to typical Mondays:

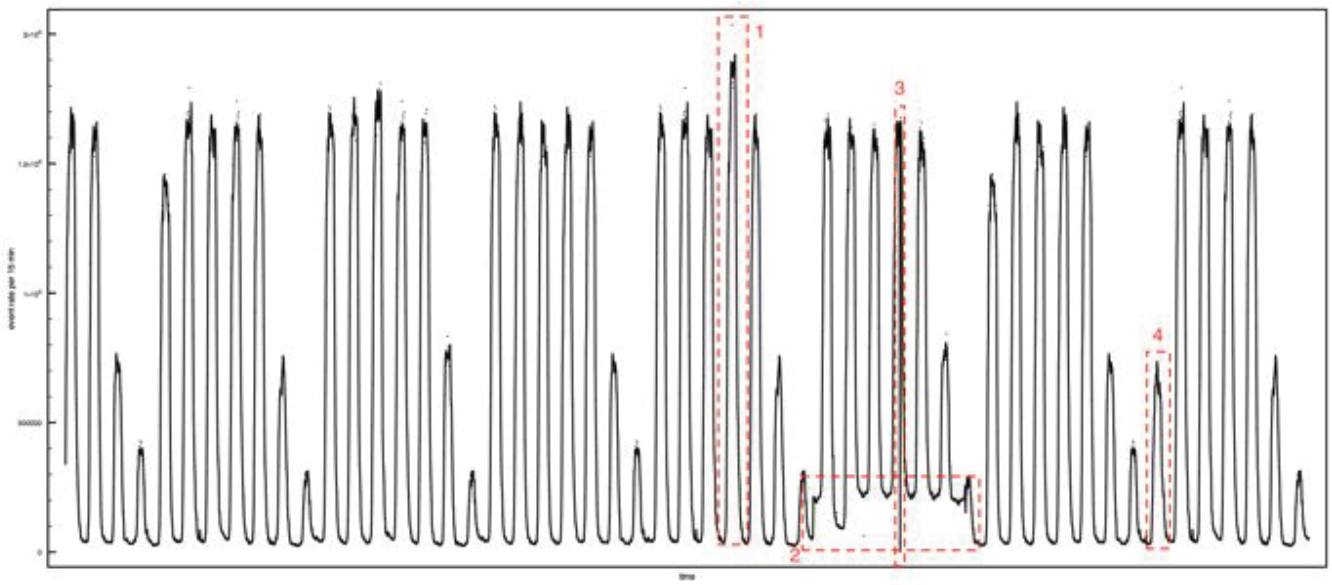

Figure 1: Four "ground truth" anomalies in the sample data set

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

Exponential Smoothing

Initial implementations of time series anomaly detection are often based on classical statistical methods. Several recent implementations by Netflix[1], Splunk[2], elastic[3], RRDtool[4], and others, for example, reference exponential smoothing methods such as "Holt-Winters Filtering" as a method for forecasting and anomaly detection.

These methods are similar to a "moving average," where a prediction is calculated as a weighted function of previous values. Using raw data sequence $\{x_t\}$, predictions made by a simple exponential smoothing algorithm $\{s_t\}$ can be given by:

$$s_t = \alpha x_t + (1 - \alpha) s_{t-1}, t > 0$$

Holt-Winters filtering is a generalization of these methods that can deal with time series containing trend and seasonal variation. In this case, three smoothing parameters are used for level, trend and seasonal variation (α, β, and γ).

Figures 2 and 3 below illustrate R's HoltWinters function[5] run with default settings on our example data set.

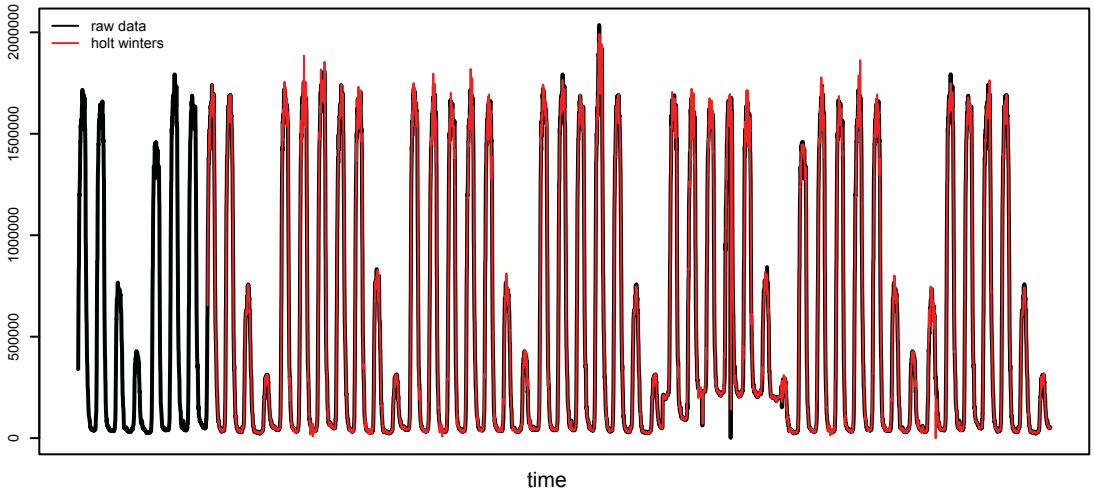

Figure 2: Raw data and visually accurate Holt-Winters predictions for example data set

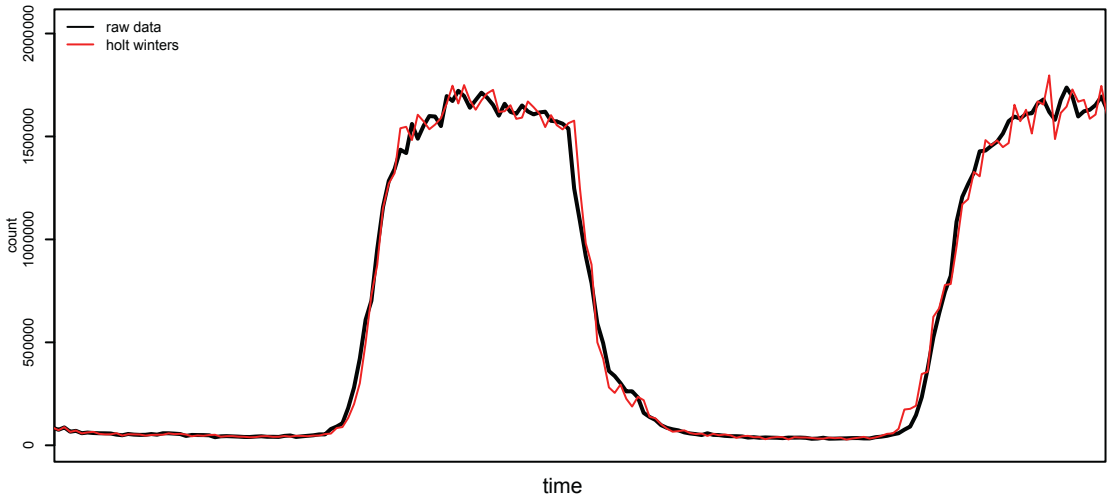

Figure 3: Zoomed detail of Figure 2 data highlights multiple incongruities between raw data and Holt-Winters predictions for example data set.

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

The results shown in Figure 2 create the impression of accurate modelling and goodness of fit, however upon closer inspection, Figure 3 shows that there are numerous instances of seemingly minor incongruities between the raw data and the predicted values.

Limitations of Exponential Smoothing for Anomaly Detection

Generally, a value is considered anomalous if it is unpredictable, given a set of related historical values. Exponential smoothing methods, such as Holt-Winters filtering, can be used for anomaly detection by comparing a new value to relevant predictions.

In Figure 4 below, the Holt-Winters prediction, bounded by a 95 percent confidence interval, is illustrated by the red line, while the upper and lower bounds are illustrated in blue. Red dots highlight detected anomalies when the actual values occur outside the upper and lower bounds (greater than the upper bound in this case).

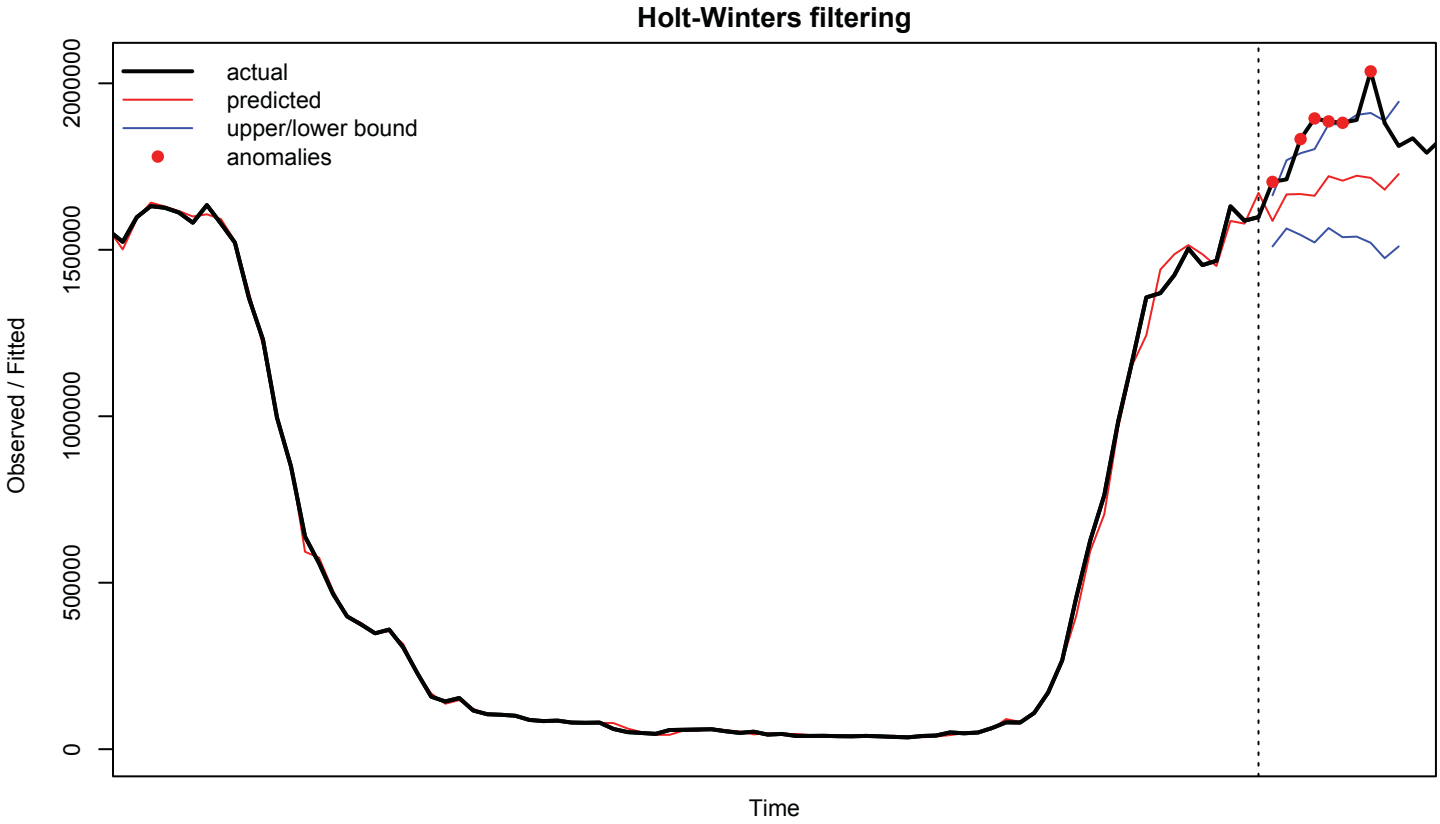

Figure 4: Example predictions using Holt-Winters filtering at time indicated by dotted line vs actual values

Since the Holt-Winters prediction is a weighted function of previous values, with a higher weight applied to the most recent values, predicted results typically lag behind changes in values, and they generally overfit values, whether they are outliers, noise or normal. An overfit model will tend to have more false positives as the model tries to "follow" anomalous data points. This characteristic of exponential smoothing limits the accuracy of these techniques when it comes to anomaly detection.

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

Anomaly Detection Comparison on Example Data Set

For this comparison, we run our retailer-supplied 7 ½ week transaction data set through two anomaly detection implementations and compare the results. The first is based on R's Holt-Winters filtering, and the second is based on Prelert's Anomaly Detective® engine.

Anomaly Detection Results Using R's Holt-Winters Filtering

Holt-Winters results were obtained using the HoltWinters[5] and predict.HoltWinters[6] functions in R with default parameters. This configuration attempts to find the optimal values of α, β, and γ by minimizing the squared one-step prediction error across the entire dataset. Experimental results with fixed values of α, β, and γ were generally more prone to false positives and negatives.

Figure 5 shows the anomaly detection results obtained from the Holt-Winters implementation. While it appears that the ground truth anomalies in the data set are detected, it's also quite apparent that there are a large number of false positive anomalies that are flagged.

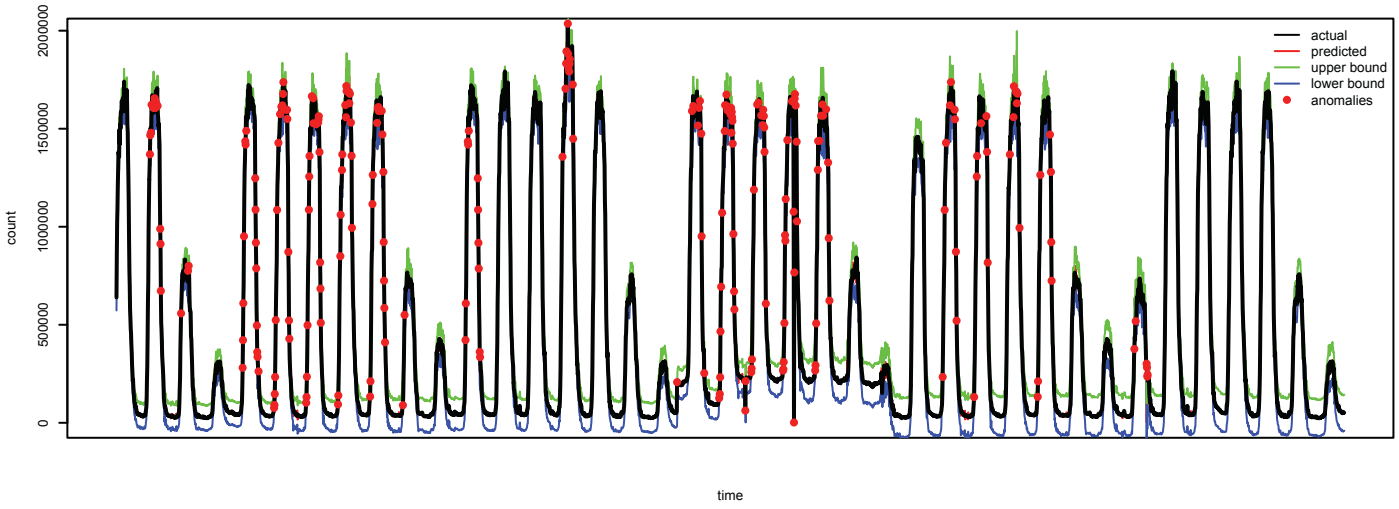

Figure 5: Anomaly detection results of Holt-Winters filtering vs actual values, showing true positive and false positive anomalies.

As explained the previous section, the tendency of Holt-Winters filtering to detect false positive anomalies when it was configured for best fit was realized in this example data set. Holt-Winters filtering is designed for making reasonably accurate short-term predictions of smooth signals based on local estimates of the value, its gradient and a single periodic component. Its specific weakness for anomaly detection is that a given configuration must choose between lag and robustness, either setting the smoothing low and resulting in the over-fit and false positives seen here, or setting the smoothing high and typically obtaining a bad fit, especially on more complex data sets.

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

Anomaly Detection Results Using Prelert Analytics

Prelert uses a fundamentally different approach to modeling time series data, which overcomes the limitations of exponential smoothing and prediction methods such as Holt-Winters. Unlike the problem Holt-Winters attempts to solve—trying to determine what will the series value be in a short time, Prelert's algorithms solve the harder problem of what should the series value have been in a short time. To solve this problem, Prelert's algorithms have been designed to accurately identify and distinguish "noise" and genuine anomalies. Figure 6 shows the anomaly detection results obtained from the Prelert Anomaly Detective Engine.

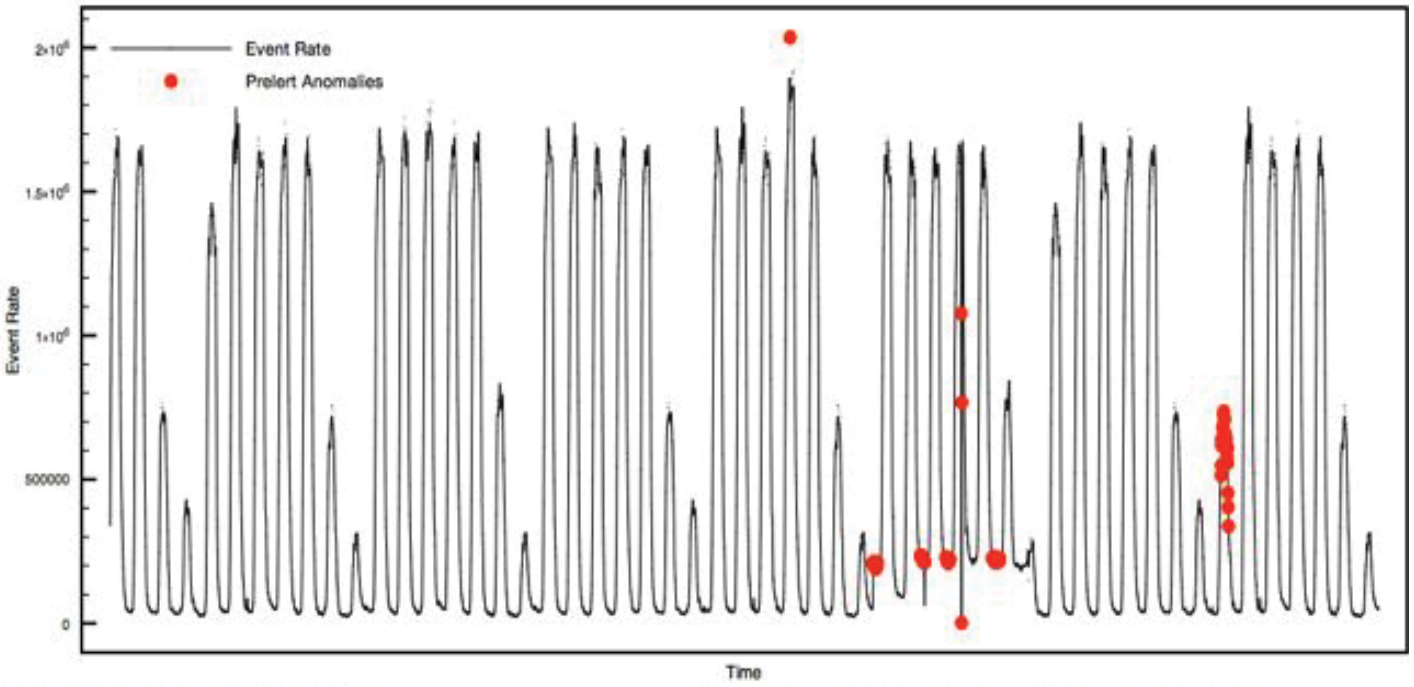

Figure 6: Anomaly detection results from Prelert's Anomaly Detective showing true positive anomalies

Prelert's unsupervised modeling process involves automatic identification and modeling of the trend components (accounting for multiple periodicities and changes in other statistical properties of the data throughout each period) and careful modeling of the residuals using multi-modal probability distributions. Bayesian measures of evidence are used for model selection that can embrace the uncertainty in the data and allow multiple models to be run concurrently. While this process adapts quickly to significant changes in the time series, and ages out old models over time, the process is also careful to avoid fitting noise and genuine anomalies. This is achieved by reducing the impact of very unexpected values on both our trend and residual estimators.

Prelert's rigorous approach is more robust to arbitrary real-world signals than simple statistical methods, and results over a large corpus of varied customer data have proven Prelert's accuracy and scalability.

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

Side-by-Side Result Comparison

Looking at a specific time period:

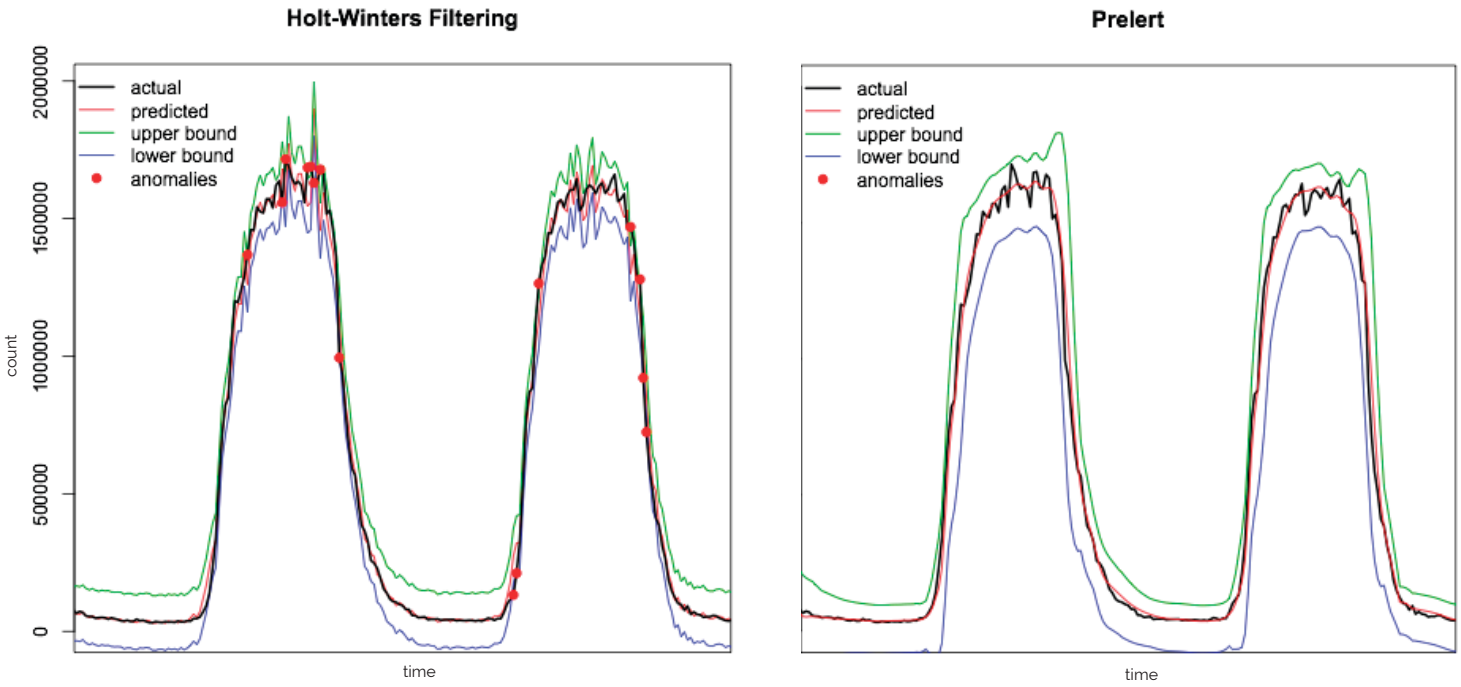

Figure 7: Holt-Winters Filtering result, left, illustrates over-fitting and false positives. Prelert Anomaly Detective result, right, illustrates accurate fit and zero false positives

Summary

Implementations of classical statistical methods such as Holt-Winters filtering are widely available in statistical packages and open source code. Often these methods appear (at least visually) to model data effectively. However, the complexities of real time series data often render these methods ineffective for accurate anomaly detection and prediction.

In addition, online (e.g. real-time) implementation of these classical methods to analyze streaming data is often not feasible, as these methods may require repeated passes over the data set, something which is impractical or impossible in real-time streaming environments.

The simple example presented in this brief shows that a robust anomaly detection system for arbitrary data types must rely on approaches that go beyond classical statistical methods. Prelert's Anomaly Detective solution has been designed to accurately model large volumes of streaming data in real-time, yielding accurate anomaly detection results.

Prelert Analytics vs. Holt-Winters Filtering
For Anomaly Detection

References

1 http://techblog.netflix.com/2014/12/introducing-atlas-netflixs-primary.html

2 http://docs.splunk.com/Documentation/ITSI/latest/ReleaseNotes/holt-winterforecasting

3 https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations-pipeline-movavg-aggregation.html#_additive_holt_winters

4 https://www.usenix.org/legacy/event/lisa2000/full_papers/brutlag/brutlag.pdf

5 https://stat.ethz.ch/R-manual/R-devel/library/stats/html/HoltWinters.html

6 https://stat.ethz.ch/R-manual/R-devel/library/stats/html/predict.HoltWinters.html

About Prelert

Prelert is the leading provider of behavioral analytics for IT security and operations teams. The company's solution analyzes an organization's log data, finds anomalies, links them together and lets the data tell the story behind advanced cyberthreats and IT performance problems. Leveraging machine learning anomaly detection and other behavioral analytics capabilities, the solution automates the analysis of massive data sets, eliminating manual effort and human error. Hundreds of progressive IT organizations rely on Prelert to detect advanced threat activity, reduce false positive alerts and enable faster root cause analysis. Prelert lets your data tell the story.