Anomaly Detection for Time Series

Tingyi Zhu   Manager: Hui Zang
tingyi.zhu@huawei.com

Problems

We investigated problems in anomaly detection for time series, including

- Outliers Detection in Single Time Series
- Online Outlier Detection for Time Series
- Unusual Series Detection among Multiple Time Series

A UI is designed to realize the detection methods.

Modeling of Single Series

Given: A time series X_T
Find: Outlier points at time t

The classical outlier detection technique of single time series is based on the general ARIMA model:

$$X_t = \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t.$$

- $\phi(B)$: autoregressive polynomial
- $\alpha(B)$: differencing filter
- $\theta(B)$: moving average polynomial

Model with outliers at time $t_1, t_2, \ldots, t_m$:

$$X_t^* = \sum_{j=1}^{m} \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t.$$

- $L_j(B)$ depends on outlier types
- $I_t(t_j) = 1$ there's outlier at time $t = t_j$, and 0 otherwise.
- ω_j denotes the magnitude of the jth outlier effect

Types of outliers:

- Additive Outliers (AO): $L(B) = 1$;
- Level Shift (LS): $L(B) = \frac{1}{1-B}$;
- Temporary Change (TC): $L(B) = \frac{1}{1-\delta B}$;
- Seasonal Level Shift (SLS): $L(B) = \frac{1}{1-B^s}$;
- Innovational Outliers (IO): $L(B) = \frac{\theta(B)}{\phi(B)\alpha(B)}$.

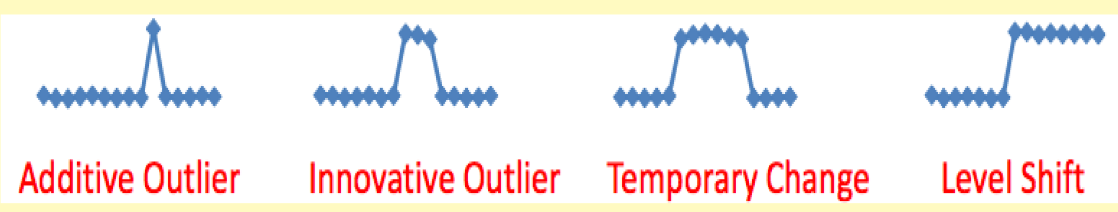

Figure 1: Types of outliers

Outlier Detection Algorithm

Step1: Fitting the series by an ARIMA model (`forecast` package in R), obtain the initial parameters $(\phi(B), \theta(B), \alpha(B))$ estimation of the model.

Step2: Compute test statistics τ for each time point to find the possible outliers and their effects

Step3: Obtain the adjusted series by removing the outlier effects

Step4: Go back to step 1, refit the model

References

[1] C. Chen, L. Liu. Joint Estimation of Model Parameter and Outlier Effects in Time Series, *Journal of American Statistical Association*,1993

[2] R. Hyndman, E. Wang, N. Laptev. Large Scale Unusual Time Series Detection, *ICDM, 2015*

Online Outlier Detection

Given: A live series, acquire new data point in real time
Goal: Upon arrival of the new data, identify whether it's outlier

Framework:

Step1: Initial series $x_1, \ldots, x_n$ as training sample

Step2: Use prediction algorithm, which can be adapted to online setting, to predict x_{n+1}

- Linear model: Stochastic Gradient Descent (SGD)
- Nonlinear model: Online Support Vector Regression (SVR)

Step3: Set criteria of identifying outliers or outlier events, based on the prediction $\hat{x}_{n+1}$ and the new data point coming in x_{n+1}

Step4: Update the training series to $x_1, \ldots, x_n, x_{n+1}$, and predict x_{n+2}

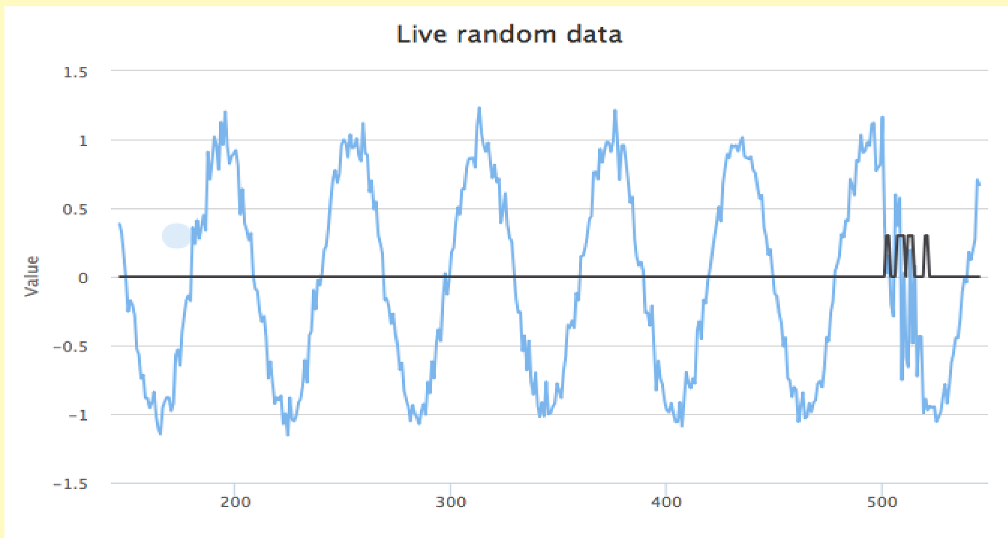

Figure 2: Online outlier detection demo

Figure 2 shows a demo for online series outlier detection, where the blue line is a live series and the black line is the indicator of outliers

Unusual Series Detection

Given: A database of time series
Find: All anomalous time series

Application: Internet companies monitoring the servers (CPU, Memory), find unusual behaviors (ex. server intrusion)

Framework:

Step1: Extract 15 features from each series in the database, most of which capture the global information of the time series

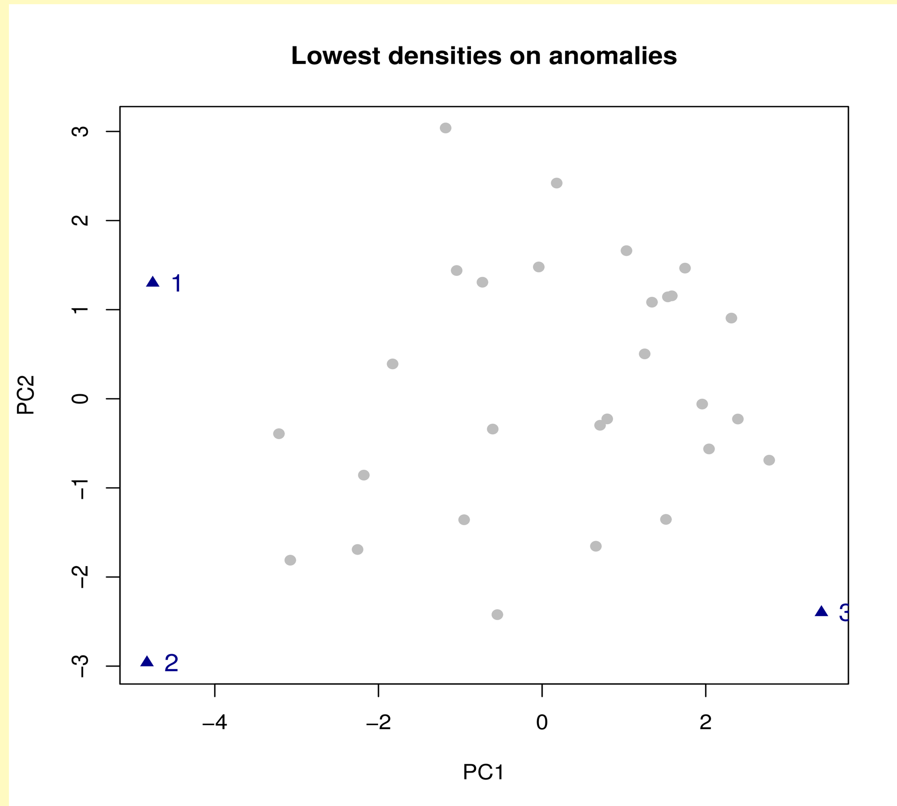

Figure 3: Feature space projected to dim 2 by PCA

Step2: PCA to reduce dimension: The first 2 PCs are sufficient to capture most of the variance

Step3: Implement multi-dimentional outlier detection algorithm to find outlier series

- Density based
- α-hull
- One-class SVM

UI for Outlier Detection

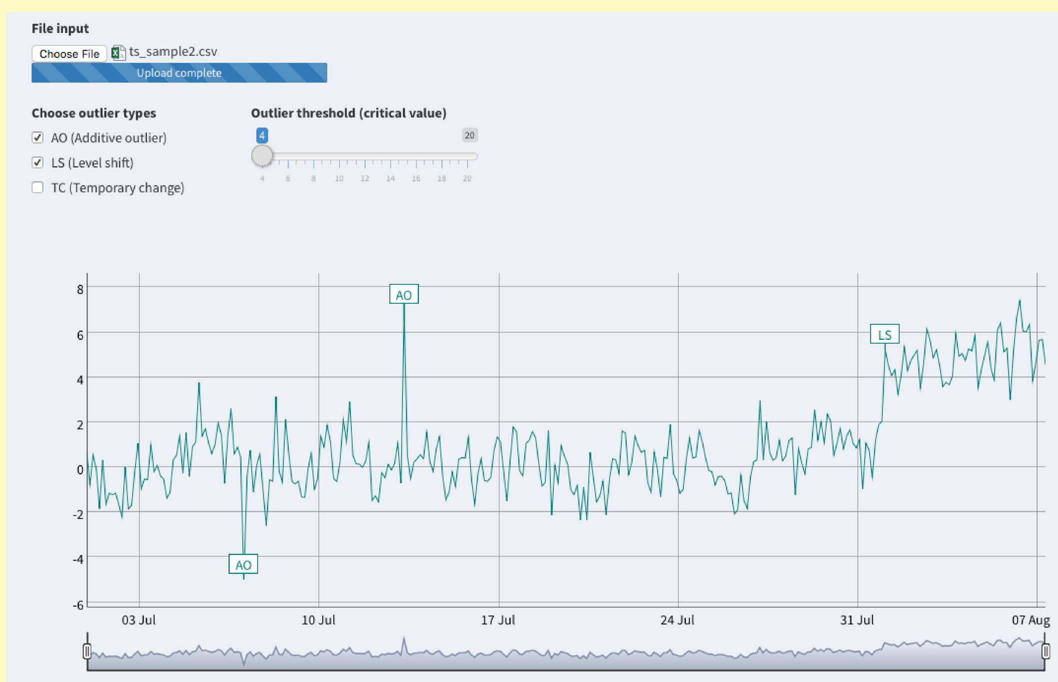

Figure 4: Simulated ARMA series with AO and LS outliers.

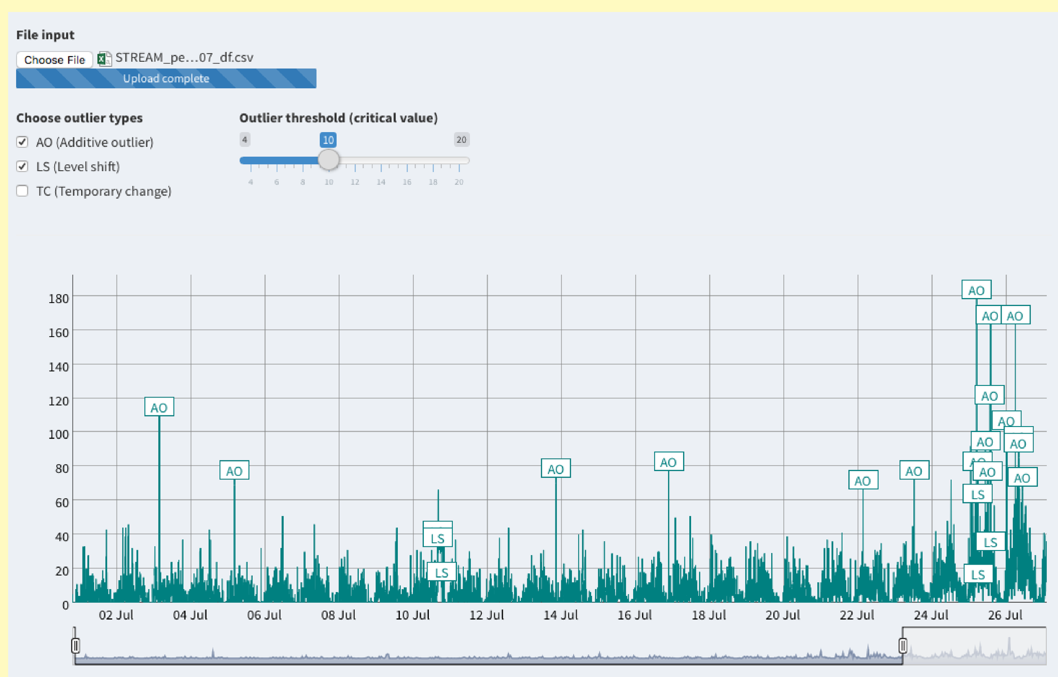

Figure 5: Video requests of carrier in July 2014

Figure 4 and Figure 5 are the interface of detecting outliers in single time series, with functions of uploading data files, choosing types of outliers to display, setting the threshold of outlier identification, etc.

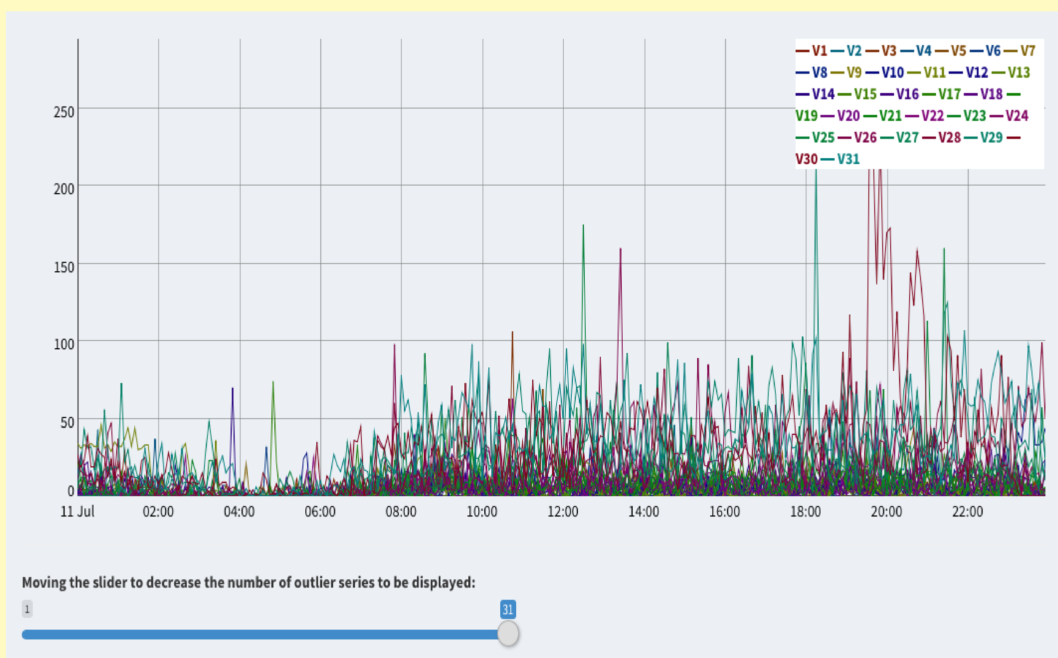

Figure 6: Daily curves of video requests in July 2014

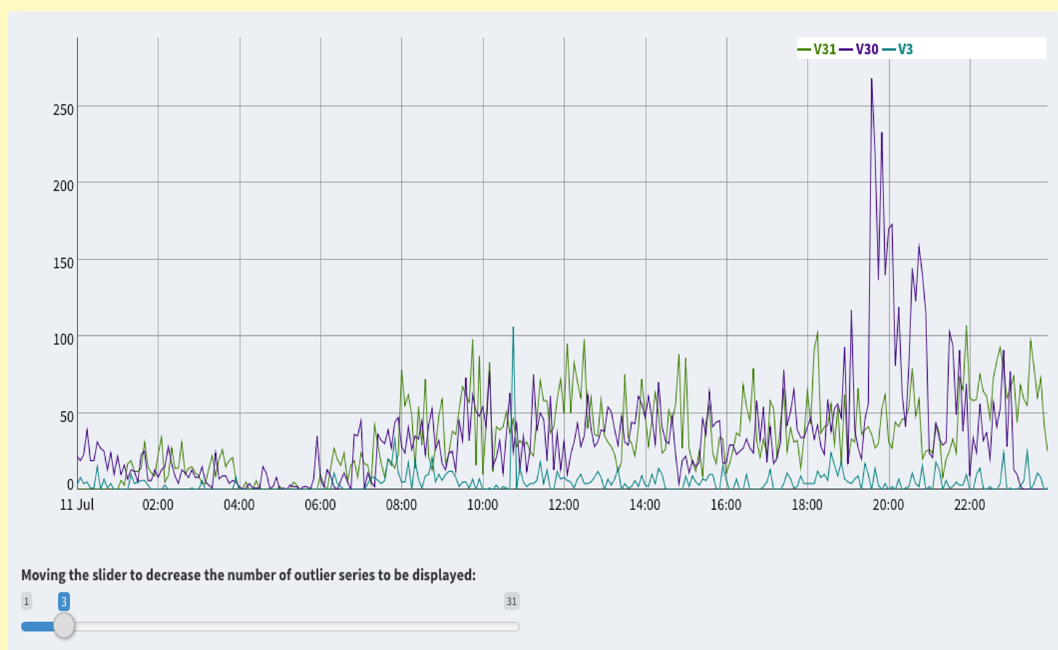

Figure 7: The 3 most unusual curves: 3rd, 30th and 31st day

Figure 6 and Figure 7 are the interface for unusual series detection in time series database. The data are the daily videos request curves split the monthly data of Figure 2. With the upload of multiple time series data, the slider below can be moved to set the number of unusual series to be displayed