Smart Monitoring System for Anomaly Detection on Business Trends in Alibaba

Zhaogang Wang
zhaogang.wzg@alibaba-inc.com
About me

• Senior Specialist of GOC (Global Operation Center) Team in Alibaba Group
  • Business trend monitoring
  • Business fault diagnosis and root cause analysis
  • Data warehouse for infrastructure and operation data

• Before I joined Alibaba
  • Senior Engineer of SRE Team in Baidu
Introduction to Alibaba Group
About business trends monitoring in Alibaba

- **Business faults management**
  - Mapping business functions to business trends

- **Faults Priority Definitions**
  - Orders per minute on Taobao decreased by XX% or above => P1 Fault
  - Transactions per minute on Alipay decreased by X% to XX% => P2 Fault

- **Business trends monitoring**
  - Business faults can be found by anomaly detection on business trends
Features of businesses trends

- Cyclicity
- Holiday Effect
- Noise and interference
Challenges of anomaly detection on business trends

• How to adopt the characteristics of different business trends?

• How to meet the artificial standards of faults?

• How to get all the configurations in automation?
Summary of anomaly detection approaches

- Local trend based
  - Static threshold
  - Dynamic threshold
  - Local regression

- Historical trend based
  - Trend prediction
    - Segment average of historical data
    - Time series decomposition
      - Holt-winters
      - STL (Seasonal Trend LOESS)
    - Machine Learning
      - Deep Learning (LSTM)
Our choice

• Our choice
  • STL (Seasonal Trend on LOESS)

• Advantages of STL on business trends time series
  • Suitable for cyclical data
  • Suitable for data with drifting trend
  • Robust to local noises and interference

How to get a good “prediction”

• A good “prediction”
  • Accurately fits business trends
  • Smooth and stable
Using STL directly on original data...

- **Drawbacks**
  - Effected by noise
  - Not smooth or stable
  - Not enough sensitive to recent trends

- **Solutions**
  - Customized data preprocessing
Customized data preprocessing

Remove history noises

Smooth the data

Smooth the data again:
Use recent trends to adjust the outline of historical data

Complete the "future" data.
A better “prediction” is born
Anomaly detection based on predicted curve

• The traditional N-sigma law
  • Anomaly point: residence > N * sigma

• N == 3?
  • Sigma varies with the time segment
  • Sigma varies with the business trend

• We need
  • Different N for each time segment and each business trend
How to determine the “N”s

- Divide the time segments by residence for each business trend
- Initialize the N for each time segment
- Adjust the N according to manual feedback
Manual feedback loop

• About the label data
  • Label data from the operators’ team
    • Effectiveness of the anomaly points
    • Quantity of the label data

• How to utilize the label data
  • Adjust the N parameter according to the label data
  • Tolerant the errors in the label data
Evaluation

• Anomaly detection
  • Precision: 80%
  • Recall: 80%

• Configuration cost
  • Auto parameter initialization
  • Auto parameter adjustment
    • When the business trend changes
Future work

• Lightweight anomaly detection for system metrics
• Early warning for business faults
• Fault diagnosis and root cause analysis
Q & A