Opprentice: Towards Practical and Automatic Anomaly Detection Through Machine Learning

Dapeng Liu, Youjian Zhao, Haowen Xu, Yongqian Sun, Dan Pei, Jiao Luo, Xiaowei Jing, Mei Feng
KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality.
KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality

KPI anomalous (unexpected) behaviors → Potential failures, bugs, attacks...
KPIs and Anomaly Detection

KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality

KPI anomalous (unexpected) behaviors → Potential failures, bugs, attacks...

Anomaly detection matters: Find anomalous behaviors of the KPI curve
→ Diagnose and fix it
→ Avoid further influences and revenue losses
KPIs and Anomaly Detection

**KPIs** (Key Performance Indicators):
A set of performance measures that evaluate the service quality

**Page views (PV) of Baidu**

**Anomaly detection matters:**
Find anomalous behaviors of the KPI curve
→ Diagnose and fix it
→ Avoid further influences and revenue losses

---

**IMC’ 15** The Dark Menace: Characterizing Network-based Attacks in the Cloud

**IMC’ 15** Dissecting UbuntuOne: Autopsy of a Global-scale Personal Cloud Back-end
How to Build the Anomaly Detection System

**Domain experts (Operators)**
- Responsible for the KPIs
- Knowing the KPI behaviors well

**Developers**
- Building the detection system
- Knowing several anomaly detectors

- Simple threshold
- Historical Average
- Wavelet
- Holt-Winters
- …
In practice, it is more complex
How to Build the Anomaly Detection System

In practice, it is more complex

Describe anomalies

Operators

Developers

Select detectors & Tune parameters

Detection System

Wavelet
Moving Average
Holt-Winters
...

Operators

Developers
How to Build the Anomaly Detection System

In practice, it is more complex

Describe anomalies

Operators

Anomalies

Developers

Detection System

Select detectors & Tune parameters

- Wavelet
- Moving Average
- Holt-Winters
- ...

2015/12/3
Dapeng Liu (liudp10@mails.tsinghua.edu.cn)
How to Build the Anomaly Detection System

In practice, it is more complex

Operators

Describe anomalies

Anomalies

Detection System

Developers

Select detectors & Tune parameters

Wavelet
Moving Average
Holt-Winters

...
How to Build the Anomaly Detection System

Challenges

1. Operators have difficulties to precisely and formally define anomalies in advance

2. Selecting and combining suitable detectors are tricky

3. Detectors are not intuitive to tune

Operators

Developers

Describe anomalies

Detect anomalies

Select detectors & Tune parameters

Wavelet

Moving Average

Holt-Winters

...
Opprentice

(Operators’ apprentice)
A More Natural Way
Design Goal

Operators

Accuracy preference (Precision & recall)

Provide

Label

Anomaly Detection

Opprentice
Design Goal

Operators
Accuracy preference (Precision & recall)
Provide

Anomaly Detection

VS.

Label

Anomaly Detection

Opprentice

Select and tune detectors

Alarm

Detection System

ARIMA

Holt-Winters

Describe anomalies and accuracy preference (FPR & FNR)

Operators

Developers
Outline

- Background and Motivation
- Key Ideas
- Results
- Conclusion
Detector model:

\[
\text{data point} \xrightarrow{\text{a detector with parameters } \{p\}} \text{severity} \xrightarrow{s\text{Thld}} \{1, 0\}
\]
Key Ideas

Detector model:

\[
\text{data point} \xrightarrow{\text{a detector with parameters } \{p\}} \text{severity} \xrightarrow{s\text{Thld}} \{1, 0\}
\]

For example

\[
\text{severity} = \frac{|\text{value} - \mu|}{\sigma}
\]
Key Ideas

Detector model:

\[
\text{data point} \xrightarrow{a \text{ detector with parameters } \{p\}} \text{severity} \xrightarrow{s\text{Thld}} \{1, 0\}
\]

For example

\[
\text{severity} = \frac{|\text{value} - \mu|}{\sigma}
\]
Detector model:

\[
\text{data point} \xrightarrow{\text{a detector with parameters } \{p\}} \text{severity} \xrightarrow{sThld} \{1, 0\}
\]

For example:

\[
\text{severity} = \frac{|\text{value} - \mu|}{\sigma}
\]

Anomaly feature
Key Ideas

Detector Configurations

- Time series decomposition
- HW 0.2 0.2 0.2
- HW 0.5 0.7 0.7

Differencing
- last day
- last season

WMA
- WIN30

Differencing
- last slot
- last season
- last day

EWMA
- 0.7

Historical average
- 4 season

Extract features
(KPI data with different parameters)

2015/3/21

Dapeng Liu (liudp10@mails.tsinghua.edu.cn)
Key Ideas

Detector Configurations

<table>
<thead>
<tr>
<th>Detector / # of configurations</th>
<th>Sampled parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple threshold [24] / 1</td>
<td>none</td>
</tr>
<tr>
<td>Diff / 3</td>
<td>last-slot, last-day, last-week</td>
</tr>
<tr>
<td>Weighted MA [11] / 5</td>
<td>points</td>
</tr>
<tr>
<td>MA of diff / 5</td>
<td></td>
</tr>
<tr>
<td>EWMA [11] / 5</td>
<td>α = 0.1, 0.3, 0.5, 0.7, 0.9</td>
</tr>
<tr>
<td>TSD MAD / 5</td>
<td></td>
</tr>
<tr>
<td>Historical average [5] / 5</td>
<td>win = 1, 2, 3, 4, 5 week(s)</td>
</tr>
<tr>
<td>Historical MAD / 5</td>
<td></td>
</tr>
<tr>
<td>Holt-Winters [6] / 4 &lt; 64</td>
<td>α, β, γ = 0.2, 0.4, 0.6, 0.8</td>
</tr>
<tr>
<td>SVD [7] / 5 × 3 = 15</td>
<td>row = 10, 20, 30, 40, 50</td>
</tr>
<tr>
<td>Wavelet [12] / 3 × 3 = 9</td>
<td>win = 3, 5, 7 days, freq = low, mid, high</td>
</tr>
<tr>
<td>ARIMA [10] / 1</td>
<td>Estimation from data</td>
</tr>
</tbody>
</table>

In total: 14 basic detectors / 133 configurations

KPI data

Extract features

(Detectors with different parameters)

Historical average-4 season

EWMA-0.7

WMA-WIN30

Differencing-last slot

Differencing-last season

Differencing-last day

Time series decomposition

HW 0.2 0.2 0.2

HW 0.5 0.7 0.7
Classification in the feature space
(Supervised machine learning)
Key Ideas

Operators

Classification in the feature space (Supervised machine learning)
Address Challenges of Designing Apprentice

- Labeling overhead
- Solution: an effective labeling tool
Address Challenges of Designing Apprentices

- **Labeling overhead**
  - Solution: an effective labeling tool

- **Incomplete anomaly types in the historical data**
  - Solution: incremental re-training with new data
Address Challenges of Designing Apprentice

- Labeling overhead
  - Solution: an effective labeling tool
- Incomplete anomaly types in the historical data
  - Solution: incremental re-training with new data
- Class imbalance problem
  - Solution: adjusting classification threshold (cThld) based on the preference
Address Challenges of Designing Apprentice

- Labeling overhead
  - Solution: an effective labeling tool

- Incomplete anomaly types in the historical data
  - Solution: incremental re-training with new data

- Class imbalance problem
  - Solution: adjusting classification threshold (cThId) based on the preference

- Irrelevant and redundant features
  - Solution: random forests
Design Overview

Training a classifier

See the paper for full details
Design Overview

Training a classifier

Detecting anomalies

See the paper for full details
Outline

- Background and Motivation
- Key Ideas
- Results
- Conclusion
Evaluation

### Data sets
1. PV
2. #SR
3. SRT

### Detection approaches
- Random forest
- Basic detectors
- Static combinations
- Other machine learning

### Training sets
- All historical data (Incremental retraining)
- First 8-week data
- Recent 8-week data

### Accuracy metrics
- PC-Score
- SD(1,1)
- F-Score
- Default cThld

### cThld predictions
- EWMA (Apprentice as a whole)
- 5-Fold cross-validation

### Other approaches

### Apprentice

### § 4.3
Detection approaches

### § 4.4
Training sets

### § 4.5
Accuracy metrics

### § 4.6
cThld predictions

### § 4.7
Labeling time vs. tuning time

### § 4.8
Detecting lag and training time

---

2015/12/3
Dapeng Liu (liudp10@mails.tsinghua.edu.cn)
Evaluation

Detection approaches
- Random forest
- Basic detectors
- Static combinations

Data sets
1. PV
2. #SR
3. SRT

Training sets
- All historical data (incremental retraining)
- First 8-week data
- Recent 8-week data

Accuracy metrics
- PC-Score
- SD(1,1)
- F-Score
- Default cThld

cThld predictions
- EWMA (Apprentice as a whole)
- 5-Fold cross-validation

Labeling time vs. tuning time
Detecting lag and training time

Opprentice
Other approaches

Other machine learning

Dapeng Liu (liudp10@mails.tsinghua.edu.cn)
Random forests vs. Basic Detectors and Static Combinations

![Graph showing comparison between Random forests and Basic Detectors]

- Random forest
- Basic detector
- Majority-Vote
- Normalization schema

Precision vs. Recall graph showing the performance of different detection methods.
Evaluation

Data sets
1. PV
2. #SR
3. SRT

Detection approaches
- Random forest
- Basic detectors
- Static combination

Other machine learning

Training sets
- All historical data (Incremental retraining)
- First 8-week data
- Recent 8-week data

Accuracy metrics
- PC-Score
- SD(1,1)
- F-Score
- Default cThld

cThld predictions
- EWMA (Apprentice as a whole)
- 5-Fold cross-validation

Labeling time vs. tuning time
Detecting lag and training time

Apprentice
Other approaches

2015/12/3
Dapeng Liu (liudp10@mails.tsinghua.edu.cn)
Random Forests vs. Other Learning Algorithms

(The order of features is based on mutual information)
Evaluation

Data sets
1. PV
2. #SR
3. SRT

Detection approaches
Random forest
Basic detectors
Static combinations
Other machine learning

Training sets
All historical data (Incremental retraining)
First 8-week data
Recent 8-week data

Accuracy metrics
PC-Score
SD(1,1)
F-Score
Default cThld

cThld predictions
EWMA (Apprentice as a whole)
5-Fold cross-validation

Labeling time vs. tuning time
Detecting lag and training time

See the paper for full details
Apprentice as a whole

Apprentice achieves

40% more points inside the preference regions than 5-Fold cross-validation

23%

110%
Opprentice as a whole

Opprentice achieves

40% more points inside the preference regions than 5-Fold cross-validation

23% 110%
Opprentice is an **automatic** and **accurate** machine learning framework for KPI anomaly detection.

Opprentice **bridges the gap** in applying complex detectors in practice.

The idea of Opprentice, i.e., using machine learning to model the domain knowledge, could be a very promising way to automate other service managements.
Thank you

liudp10@mails.tsinghua.edu.cn

On the job market 😊