Generic and Robust Localization of Multi-Dimensional Root Causes

Zeyan Li, Chengyang Luo, Yiwei Zhao, Yongqian Sun, Kaixin Sui, Xiping Wang, Dapeng Liu, Xing Jin, Qi Wang, Dan Pei

ISSRE 2019
Background

- KPI: key performance indicator

Anomaly happens, and we need to find the root cause
Motivation

Raw log for an order:

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Province</th>
<th>ISP</th>
<th>Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019.10.15</td>
<td>Beijing</td>
<td>China Mobile</td>
<td>PC</td>
</tr>
</tbody>
</table>

Province
ISP
Device

China Unicom
Beijing
Shanghai
Guangdong
China Mobile
PC
Cellphone

Beijing & China Unicom
Multi-dimensional Data

- Cuboid: a way to slice the multi-dimensional data
- Attribute combination: elements in a cuboid
Multi-dimensional Data

- Cuboid: a way to slice the multi-dimensional data
- Attribute combination: elements in a cuboid
Multi-dimensional Data

- Cuboid: a way to slice the multi-dimensional data
- Attribute combination: elements in a cuboid
Problem Statement

The KPI of the whole cube is abnormal, but where is the root cause?

Root cause is a set of attribute combinations
Challenge: Huge Search Space

Root Cause: a set of attribute combinations

How many potential root cause for a simple 2-d data?

\[2^2 + 7 + 14 - 1\]
## Previous Approaches

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Root Cause Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>single attribute</td>
</tr>
<tr>
<td>Recursive Adtributor (Master Thesis, 2018)</td>
<td>none</td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>one or two attribute combinations</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>none</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>all attribute combinations of the root cause in one cuboid</td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>those which cause the same changes are in one cuboid</td>
</tr>
</tbody>
</table>
# Previous Approaches

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>fundamental &amp; derived (quotient)</td>
</tr>
<tr>
<td>Recursive Adtributor</td>
<td>fundamental &amp; derived (quotient)</td>
</tr>
<tr>
<td>(Master Thesis, 2018)</td>
<td></td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>fundamental only</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>fundamental &amp; derived</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>fundamental only</td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>fundamental &amp; derived (quotient, product)</td>
</tr>
</tbody>
</table>

**Diagram:**

- **Total Volume**
  - China Mobile
  - China Unicom

- **# Orders**
  - fundamental, additive

- **% Success Rate**
  - derived, not additive

---

**Notes:**

- iDice and HotSpot rely on addition, thus cannot handle derived measures.
## Previous Approaches

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Change Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>significant</td>
</tr>
<tr>
<td>Recursive Adtributor (Master Thesis, 2018)</td>
<td>significant</td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>significant</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>any</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>significant</td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>any</td>
</tr>
</tbody>
</table>

![Graph](image)
## Previous Approaches

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter Fine Tuning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>no</td>
</tr>
<tr>
<td>Recursive Adtributor</td>
<td>yes</td>
</tr>
<tr>
<td>(Master Thesis, 2018)</td>
<td></td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>no</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>yes</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>no</td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>no</td>
</tr>
</tbody>
</table>

Some approaches perform badly without parameter fine tuning.
# Previous Approaches

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>very short</td>
</tr>
<tr>
<td>Recursive Adtributor (Master Thesis, 2018)</td>
<td>short</td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>very short</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>always too long</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>sometimes long</td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>short</td>
</tr>
</tbody>
</table>

Some approaches cost too much time.
## Previous Approach

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Root Cause Assumption</th>
<th>Measure</th>
<th>Change Magnitude</th>
<th>Parameter Fine Tuning</th>
<th>Time Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adtributor (NSDI, 2014)</td>
<td>single attribute</td>
<td>fundamental &amp; derived (quotient)</td>
<td>significant</td>
<td>no</td>
<td>very short</td>
</tr>
<tr>
<td>Recursive Adtributor</td>
<td>none</td>
<td>fundamental &amp; derived (quotient)</td>
<td>significant</td>
<td>yes</td>
<td>short</td>
</tr>
<tr>
<td>(Master Thesis, 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iDice (ICSE, 2016)</td>
<td>one or two attribute combinations</td>
<td>fundamental only</td>
<td>significant</td>
<td>no</td>
<td>very short</td>
</tr>
<tr>
<td>Apriori (TON, 2017)</td>
<td>none</td>
<td>fundamental &amp; derived</td>
<td>any</td>
<td>yes</td>
<td>always too long</td>
</tr>
<tr>
<td>HotSpot (IEEE Access, 2018)</td>
<td>all attribute combinations of the root</td>
<td>fundamental only</td>
<td>significant</td>
<td>no</td>
<td>sometimes long</td>
</tr>
<tr>
<td></td>
<td>cause in one cuboid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squeeze (ISSRE, 2019)</td>
<td>those which cause the same changes are in</td>
<td>fundamental &amp; derived (quotient,</td>
<td>any</td>
<td>no</td>
<td>short</td>
</tr>
<tr>
<td></td>
<td>one cuboid</td>
<td>product)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Design Goals

<table>
<thead>
<tr>
<th>Root Cause Assumption</th>
<th>Measure</th>
<th>Change Magnitude</th>
<th>Parameter Fine Tuning</th>
<th>Time Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squeeze</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Squeeze* has no impractical assumptions
handles both fundamental and derived measures
handles anomalies with any change magnitude
does not need parameter fine tuning
is consistently fast in all cases
Core Idea: Generalized Ripple Effect (GRE)

With idea from HotSpot[IEEE Access 2018], we propose generalized ripple Effect.
Core Idea: GRE & Deviation Score

\[ \text{deviation score} = 2 \frac{f - v}{f + v} \]

forecast value: \( f \)  
real value: \( v \)

Beijing & China Mobile  
Beijing & China Unicom
Beijing  
Shanghai  
Guangdong

\( f = 30, v = 15, ds = \frac{2}{3} \)
\( f = 20, v = 10, ds = \frac{2}{3} \)
\( f = 10, v = 5, ds = \frac{2}{3} \)

PDF

Deviation Score

should in the same bin
Core Idea: GRE in Real World Cases

# successful orders drops down after an update

By manually analysis, root cause is ServiceType=020020

Their deviation scores are in the same bin, which supports GRE
Core Idea: GRE in Real World Cases

# successful orders drops down

4 root cause attribute combinations

The data shows that deviation scores of the same root cause are in the same bin
Generalized Ripple Effect

Does GRE holds for both fundamental and derived measures? **Yes.** Please see the details in the paper.
Core Idea: Generalized Potential Score

Evaluate how likely a set of attribute combination is the root cause

\[ GPS = 1 - \frac{\text{avg}(|v(S_1) - a(S_1)|) + \text{avg}(|v(S_2) - f(S_2)|)}{\text{avg}(|v(S_1) - f(S_1)|) + \text{avg}(|v(S_2) - f(S_2)|)} \]
Core Idea: Generalized Potential Score

Sorted candidates list: (Beijing,*), (Shanghai,*), (Guangdong,*), (Zhejiang,*), ....

S1: abnormal

→ forecast value and real value should be close
→ \( f(S2) - v(S2) \sim 0 \)

S2: normal

→ KPI value should be expected by GRE
→ \( \frac{v(\text{Beijing})}{f(\text{Beijing})} = 0.5 \), half fails
→ \( a(\text{Beijing, China Mobile}) = f(\text{Beijing, China Mobile}) \cdot 0.5 = 5 \)
→ \( a(\text{Beijing, China Unicom}) = f(\text{Beijing, China Unicom}) \cdot 0.5 = 10 \)

\[
GPS = 1 - \frac{\text{avg}(|v(S_1) - a(S_1)|) + \text{avg}(|v(S_2) - f(S_2)|)}{\text{avg}(|v(S_1) - f(S_1)|) + \text{avg}(|v(S_2) - f(S_2)|)}
\] normalization
Overall Architecture

Squeeze

Related System  Anomaly Happens  Timestamp

Measures Database  Forecast

Bottom-Up

Clustering: Sec IV.C

Top-Down

Locating in Each Cluster: Sec IV.D

Root Cause: A Set of Attribute Combinations
Squeeze

Top to Bottom: Search in each cluster

Bottom to Top: clustering for leaf attribute combinations

Root Causes
Clustering

Anomaly Happens

Related System

Timestamp

Measures Database

Forecast

Squeeze

Bottom-Up

Clustering: Sec IV.C

Locating in Each Cluster: Sec IV.D

Root Cause: A Set of Attribute Combinations

Top-Down
Clustering

Find attribute combinations affected by the same root cause

Find attribute combinations have similar deviation scores

local maxima: centroids

local minima: boundaries
Localize in Each Cluster

- Anomaly Happens
- Related System
- Timestamp
- Measures Database
- Forecast
- Clustering: Sec IV.C
- Locating in Each Cluster: Sec IV.D
- Root Cause: A Set of Attribute Combinations
Beijing, Shanghai, ......

Sorted List:
Beijing, Shanghai, ......

Top-K items in this list with highest GPS
Beijing, GPS = 1, Root Cause
Experiment Setup

We use

- real KPI datasets from 2 companies;
- synthetic anomalies => 7 semi-synthetic datasets
- Moving average as the forecasting algorithm.
Effectiveness

Squeeze achieves relatively good F1-score on both fundamental & derived measures.

Two of Fundamental Measure Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Algorithm</th>
<th>(1, 1)</th>
<th>(1, 2)</th>
<th>(1, 3)</th>
<th>(2, 1)</th>
<th>(2, 2)</th>
<th>(2, 3)</th>
<th>(3, 1)</th>
<th>(3, 2)</th>
<th>(3, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squeeze</td>
<td>0.8632</td>
<td>0.7827</td>
<td>0.4932</td>
<td>0.7584</td>
<td>0.6361</td>
<td>0.4097</td>
<td>0.6441</td>
<td>0.5145</td>
<td>0.3618</td>
<td></td>
</tr>
<tr>
<td>HotSpot</td>
<td>0.6856</td>
<td>0.4389</td>
<td>0.2138</td>
<td>0.5085</td>
<td>0.3453</td>
<td>0.2043</td>
<td>0.3988</td>
<td>0.2916</td>
<td>0.1768</td>
<td></td>
</tr>
<tr>
<td>A-Adrubutor</td>
<td>0.3892</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.4010</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.3857</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>R-Adrubutor</td>
<td>0.0180</td>
<td>0.0020</td>
<td>0.0016</td>
<td>0.0075</td>
<td>0.0049</td>
<td>0.0294</td>
<td>0.0081</td>
<td>0.0067</td>
<td>0.0410</td>
<td></td>
</tr>
<tr>
<td>iDice</td>
<td>0.0000</td>
<td>0.0036</td>
<td>0.0425</td>
<td>0.0000</td>
<td>0.0065</td>
<td>0.0437</td>
<td>0.0000</td>
<td>0.0007</td>
<td>0.0172</td>
<td></td>
</tr>
<tr>
<td>Apriori</td>
<td>0.1036</td>
<td>0.0980</td>
<td>0.0001</td>
<td>0.1427</td>
<td>0.0676</td>
<td>0.0019</td>
<td>0.1537</td>
<td>0.0882</td>
<td>0.0062</td>
<td></td>
</tr>
</tbody>
</table>

Derived Measure Dataset
Efficiency

*Squeeze* is fast enough consistently in all cases.

*Squeeze* costs only ten to twenty seconds consistently in all cases.
Various Anomaly Change Magnitude

Squeeze performs well regardless of anomaly change magnitudes

0.4% and 12% are 25 and 75 percentile of change magnitudes
Various Forecasting Residual

Squeeze performs well under various residuals, and always outperforms others.

<table>
<thead>
<tr>
<th>Name</th>
<th>n</th>
<th>d</th>
<th>#AC</th>
<th>Source</th>
<th>Measure</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>400</td>
<td>5</td>
<td>15324</td>
<td>$I_1$</td>
<td>Fundamental</td>
<td>13.0%</td>
</tr>
<tr>
<td>$B_0$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Fundamental</td>
<td>0.80%</td>
</tr>
<tr>
<td>$B_1$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Fundamental</td>
<td>3.19%</td>
</tr>
<tr>
<td>$B_2$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Fundamental</td>
<td>6.37%</td>
</tr>
<tr>
<td>$B_3$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Fundamental</td>
<td>9.54%</td>
</tr>
<tr>
<td>$B_4$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Fundamental</td>
<td>13.0%</td>
</tr>
<tr>
<td>$D$</td>
<td>100</td>
<td>4</td>
<td>21600</td>
<td>$I_2$</td>
<td>Derived</td>
<td>3.99%</td>
</tr>
</tbody>
</table>
Summary

● Bottom-up & Top-down => Squeeze
● Contributions:
  ○ Generalized ripple effect
  ○ Squeeze algorithm.
  ○ Experimental study on real world data and semi-synthetic data show Squeeze is both effective and efficient.
● Future Works
  ○ focus on numerical attributes
  ○ show GRE for more types of derived measures
References


Thank you. Q&A