An Integrated Framework for Optimizing Automatic Monitoring Systems in Large IT Infrastructures

Liang Tang, Tao Li
{ltang002,taoli}@cs.fiu.edu

Larisa Shwartz, Florian Pinel
{lshwart,pinel}@us.ibm.com

Genady Ya. Grabarnik
grabarng@stjohns.edu
In large IT infrastructures, the system admin cannot manually monitor so many machines.
Monitoring system: monitor those servers, notify the system admin only when a problem happens.
Configurations of Monitoring Systems are Complicated

- In Large IT infrastructures, there are different machines, different software products...
- IBM Tivoli monitoring defines a lot of monitoring situations for monitoring different alerts
  - High CPU utilization
  - Low disk space
  - Process offline
  - ...

IBM Tivoli Monitoring
What is Misconfiguration? (1)

- False Positive:
  - Too Conservative threshold (CPU utilization < 50%).
  - Transient Alert (Automatically disappear in a short time).

Where is the alert???

System Administrator

Capture

Alert: High CPU Utilization in host:xxx

Notify

Log on, find CPU utilization is normal. (Alert disappeared)
What is Misconfiguration? (2)

- **False Negative:**
  - Installed a new database server, but *forget* to add it into the monitoring situation. If this server has a problem, it would not be captured by the monitoring system.
Our solution: Develop a system module in *Event & Ticket Analysis Portal* to analyze the monitoring events with tickets and to help system admin to correct misconfigurations.
How to Detect False negative and False positive?

- **Ticket data** is the ground truth (labeled data) and created by the human.

Human labor cost is very high!!!

Can we use their knowledge to improve the monitoring?
Eliminating False Positive (1)

- A straightforward solution: **Binary classifier**
  - label “1” means a real alert, “0” means a false alert.
  - features are system event attributes
    - process name
    - CPU time
    - number of threads.

- Limitations:
  - We can NOT miss any real alert (would cause system crash or data loss).
  - No classification algorithm can guarantee 100% accuracy.
Eliminating False Positive (2)

- Gather incident information and create event
- Predict it is “False”? Yes → Wait
- Predict it is “False”? No → Create ticket
- Is this alert cleared?
  - Yes → Remove this event
  - No

Most false positive alerts are transient alerts (automatically disappear in a short time).
Eliminating False Positive (3)

• The rules generated by a classifier can be directly translated into monitoring situations:
  • If $\text{PROC\_CPU\_TIME} > 50\%$ and $\text{PROC\_NAME} = \text{Rtvscan}$, then it is false.

• *Waiting time* is the polling interval of a monitoring situation in IBM Tivoli Monitoring.

We do NOT have to build another system to deploy our classifier
Offline Evaluation on Testing Data

Ratio of the testing data size and training data size
Online Evaluation

A large financial company.

An internal account in IBM.
Eliminating False Negative (1)

- How to eliminate false negatives (missed alerts)?
  - False negative are quite few (less than 20-40 tickets for a situation). No need an automatic approach to correct it.

- False negatives are missed alerts. Where can we track them?
  - Manual Tickets (captured by human).
  - However, manual tickets contain other kinds of tickets, such as customer request.
Eliminating False Negative (2)

- **Problem Definition:** Find missed alerts from manual tickets

- **Challenges:**
  - Not enough labeled data.
  - Highly imbalanced data: few false negative alerts, large amount of other manual tickets.

We cannot hire an expert to label the ticket every day...

We randomly select a subset of tickets to label. Maybe none of them is false negative alert. It is bad for training.
Selective Labeling in Highly Imbalanced Data

- Use some domain words to narrow down the training ticket scope.

<table>
<thead>
<tr>
<th>Situation Issue</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB2 tablespace Utilization</td>
<td>DB2, tablespace</td>
</tr>
<tr>
<td>File System Space Utilization</td>
<td>space, file</td>
</tr>
<tr>
<td>Disk Space Capacity</td>
<td>space, drive</td>
</tr>
<tr>
<td>Service Not Available</td>
<td>service, down</td>
</tr>
<tr>
<td>Router/Switch Down</td>
<td>router</td>
</tr>
</tbody>
</table>

- Build a binary classifier (SVM) on selected tickets.
  - Given a ticket, label “1” means this ticket is a false negative. Label “0” means it is not.
Easy to learn!!! Not many variations of discriminative words
## A Case Study

### Discovered False Negatives (Missed alerts)

<table>
<thead>
<tr>
<th>Situation</th>
<th>Ticket</th>
</tr>
</thead>
<tbody>
<tr>
<td>dsp_3ntc_std</td>
<td>Please clear space from E drive xxxx-fa-ntfwwfdb Please clear space from E drive xxxx-fa-ntfwwfdb. It is having 2 MB free...</td>
</tr>
<tr>
<td>fss_rlzc_std</td>
<td>/opt file system is almost full on xxx Hi Team@/opt file system is almost full. Please clear some space /home/dbasso&gt;df -h /optFilesystem...</td>
</tr>
<tr>
<td>svc_3ntc_std</td>
<td>RFS101681 E2 Frontier all RecAdmin services are down Frontier RecAdmin services are not running on the batch server Kindly logon to the server : xxx.xxx.155.183/xxx ...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

---

I will add these devices into Tivoli monitoring configuration.
Summary

- Analyzed the main types of misconfiguration of monitoring systems in large IT infrastructures.
- Proposed a framework to integrate system events and tickets for improving the configurations of monitoring systems (IBM Tivoli monitoring).
- Conduct offline and online experiments for the proposed framework.
- Develop and deployed the module in Event and Ticket Analysis Portal in IBM IT service platform.
End

• Thank you!

• Any question?