AIOps: Artificial Intelligence for IT Operations

Dan Pei

Tsinghua University

Executive Summary

* AIOps is rising: replace manual Ops decisions with AI-based decision aids
  * Improved revenue
  * reduced loss/cost
  * necessary
  * feasible

* Case studies (Collaboration with Baidu, Alibaba, Tencent, Didi, Sogou):
  * Anomaly Detection
  * Anomaly Localization
  * Root Cause Analysis
  * Capacity/Failure Prediction

* AIOps Challenge: Community efforts for widespread adoption of AIOps

My life as an Ops Researcher

My Official Resume

2000-2005 UCLA Ph.D., Best Ph.D. Thesis, working on BGP, OSPF etc.

Summer 2003, Intern at AT&T Research

2005-2011 Senior/Principal Researcher at AT&T Research
ACM, IEEE Senior Member

2012-now Associated Professor at Computer Science Department at Tsinghua University.

My Operator Resume

For five years, chased ISP OPs for data, experiences, and insights.

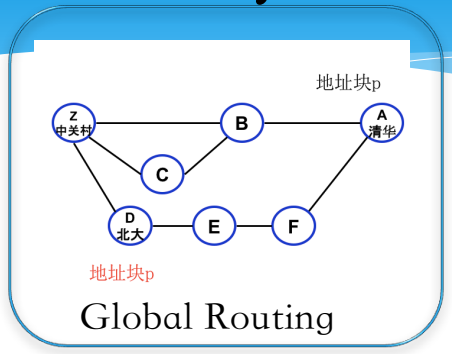
Felt in love with real OP data

Essentially a tier-5 OP

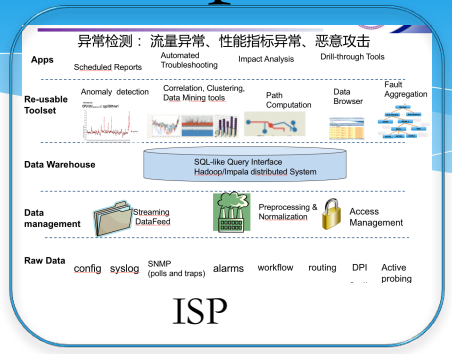
Teaching "Advanced Network Management.

Working with OPs at Baidu, Microsoft Azure,Tecent,Alibaba, DiDi, Sogou

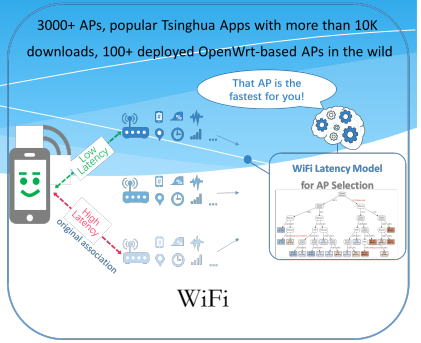
My life as an Ops Researcher



Global Routing



ISP



WiFi

My Official Resume

2000-2005 UCLA Ph.D., Best Ph.D. Thesis, working on BGP, OSPF etc.

Summer 2003, Intern at AT&T Research

2005-2011 Senior/Principal Researcher at AT&T Research
ACM, IEEE Senior Member

2012-now Associated Professor at Computer Science Department at Tsinghua University.

My Operator Resume

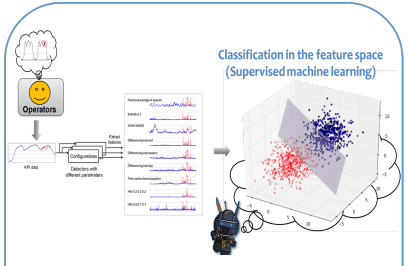
For five years, chased ISP OPs for data, experiences, and insights.

Felt in love with real OP data

Essentially a tier-5 OP

Teaching “Advanced Network Management.
Working with OPs at Baidu, Microsoft Azure,Tecent,Alibaba, DiDi, Sogou

Mobile Apps



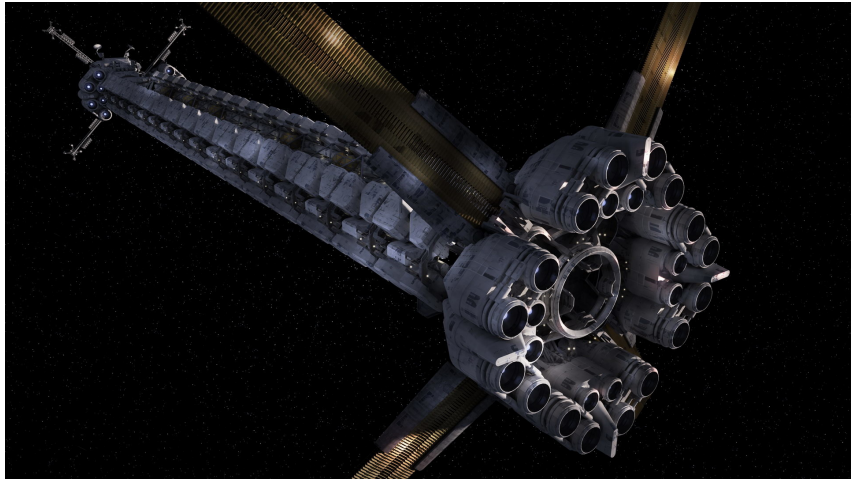
Cloud

Overview

* **What is AIOps**?
* What's the value of AIOps?
* Is AIOps necessary?
* Is AIOps feasible?
* Case studies
* AIOps Challenge
* Summary

Ultimate Goal: Autonomous IT Operations

AIOps: Algorithmic IT Operations -> Artificial Intelligence for IT Operations -> Autonomous IT Operations



Spaceship Covenant: 2000 passengers and 15 crew members all in hibernation. Flying towards Planet Origae-6。Only one awaken android crew.

Spaceship Avalon: 5000 passengers and 258 crew members. Flying towards Planet Homestead II, 120-year trip.

Autonomous IT Operations: Automatically deal with all four causes of changes to IT systems

* Software & hardware failures--> Automatic Healing
* Software changes --> Autonomous software deployment
* Change of user request amount & Pattern --> Elastic Resource Allocation
* Malicious Attacks-->Autonomous Defense

"Most people overestimate what they can do in one year and underestimate what they can do in ten years."

-- Bill Gates

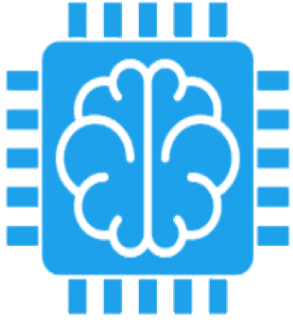
It's the responsibility of the Operations to ensure undisrupted services, despite the inevitable failures of the imperfect underlying hardware and software.

Failure Discovery → Failure Mitigation → Failure Repair → Failure Avoidance

AIOps: the Brain for Ops Decision Aids for reduced loss/cost

Monitoring Data

Automatic, Fast, Accurate, Predictive decision aids: take at most a few minutes



Expert Ops

Mitigation: rollback, HotFix, reboot, reactive traffic switching, reactive capacity upsizing

Repair: replace faulty hardware, fix bugs, refactor code

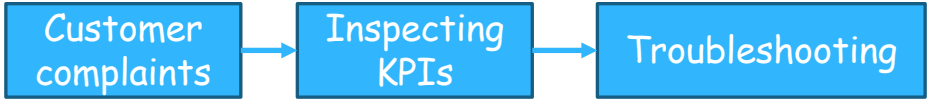
Avoidance: preventive hardware replacement; preventive traffic switching; targeted capacity upsizing

Overview

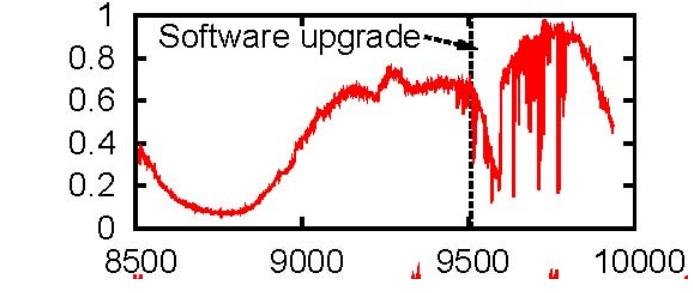
* *What is AIOps?*
* **What's the value of AIOps?**
* Is AIOps necessary?
* Is AIOps feasible?
* Case studies
* AIOps Challenge
* Summary

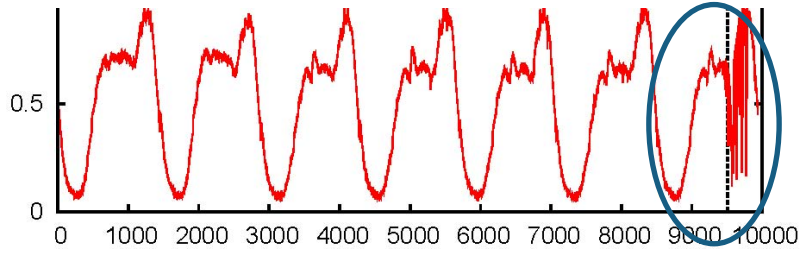
Reduced Business Loss: Rapid Assessment of Software Changes

* A buggy deployment causes significant revenue Loss
* Manual trouble shooting takes 1.5 hours



* AIOps solution takes less than 10 minutes





Joint Work with Baidu

Published in ACM CoNext 2015

Improved Revenue: Reduced Page Response Time

amazon

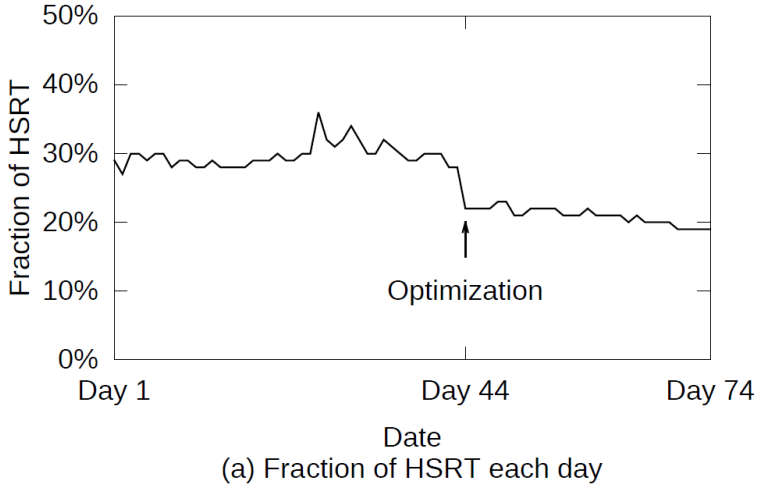
-100ms ->Sales ↑ 1%

[Greg Linden, Amazon]

Google

-100ms~400ms -> Revenue ↑ 0.2%~0.6%

[Jake Brutlag, Google]



(a) Fraction of HSRT each day

After deploying the solutions suggested by AIOps:

Slow responses (>1s) are reduced from 30% to 20%

80th-percentile response time is reduced by 253 ms

Saves 30 man-months (estimated) of manual analysis

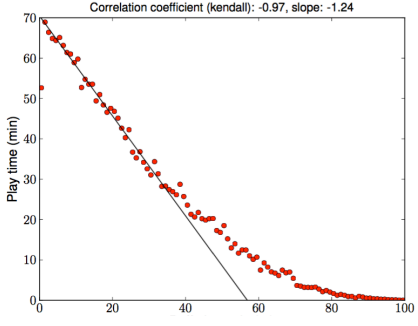
Joint Work with Baidu
Published in IEEE INFOCOM 2016

AIOps Leads to Better User Experience -> Longer Engagement -> More Revenue

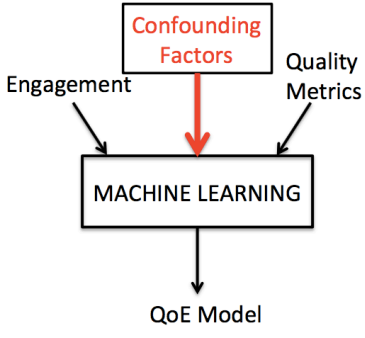
Linear Regression SIGCOMM 2011

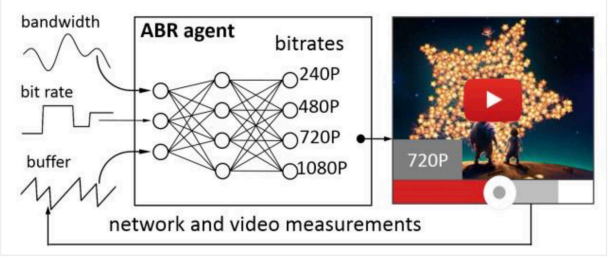
Decision Trees SIGCOMM 2013

Reinforcement Learning SIGCOMM 2017



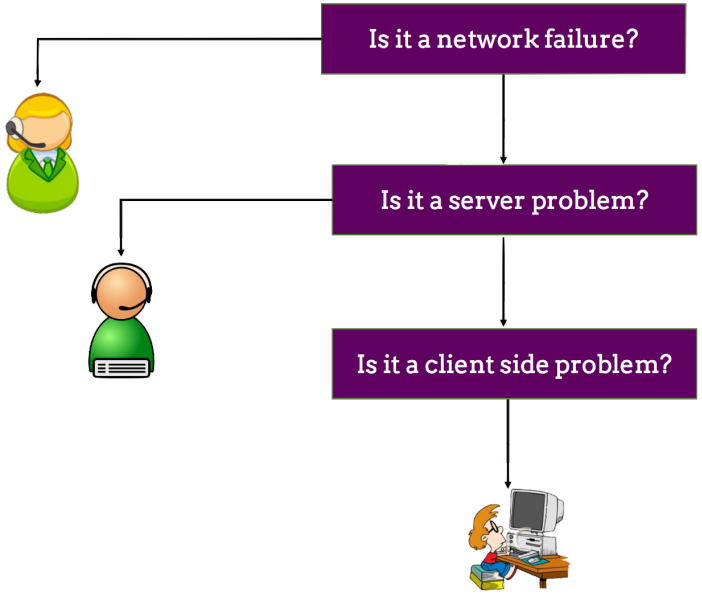
Adding as a feature

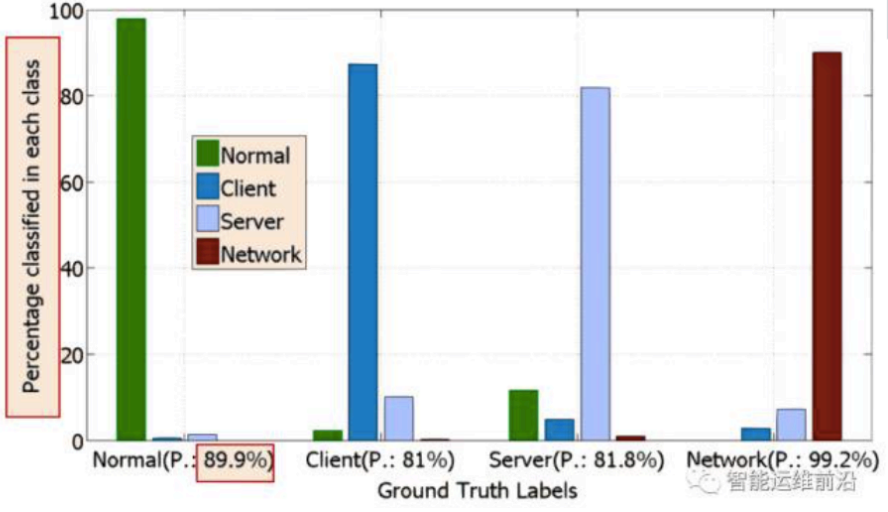




Conviva/CMU/MIT work

AIOps Quickly Decides the Responsibility Boundary: Reduced loss

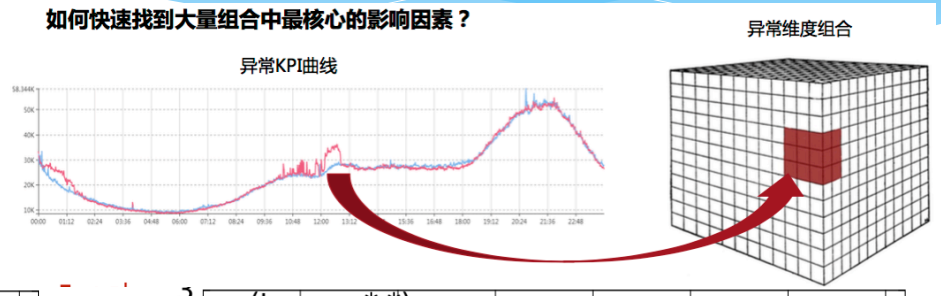


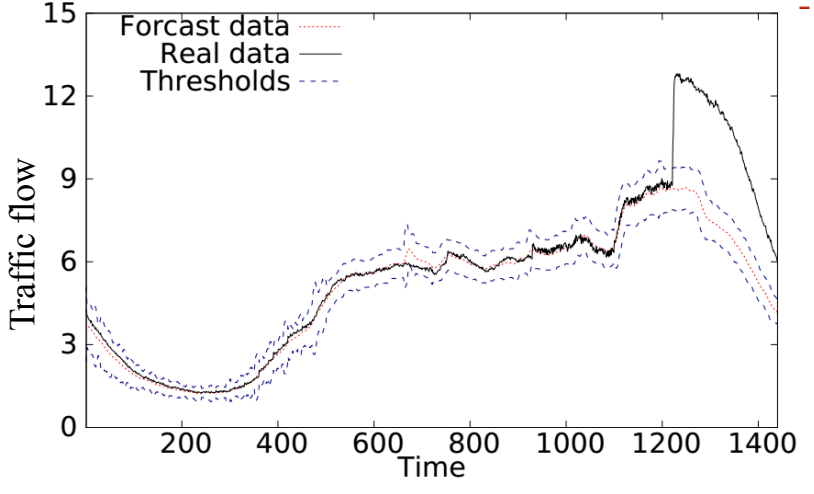


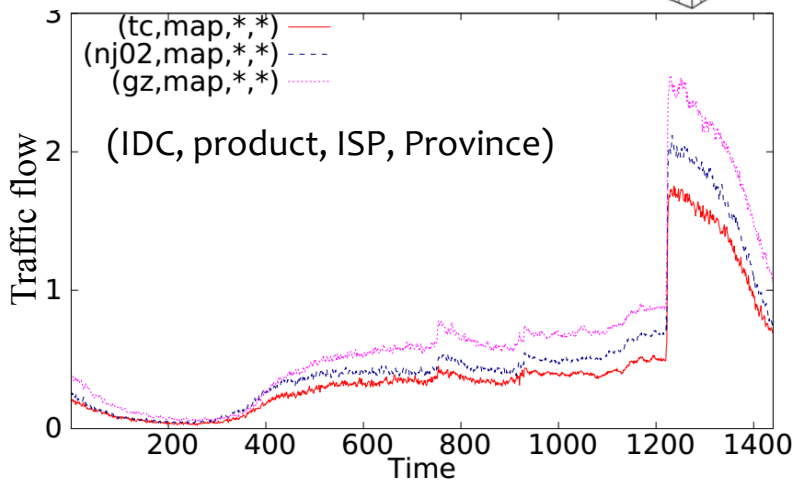
Microsoft Azure Work.
Published in SIGCOMM 2016

Localizing the Anomalous Regions: Reduced Loss

Manual localization: 90 minutes
AIOps: 30s





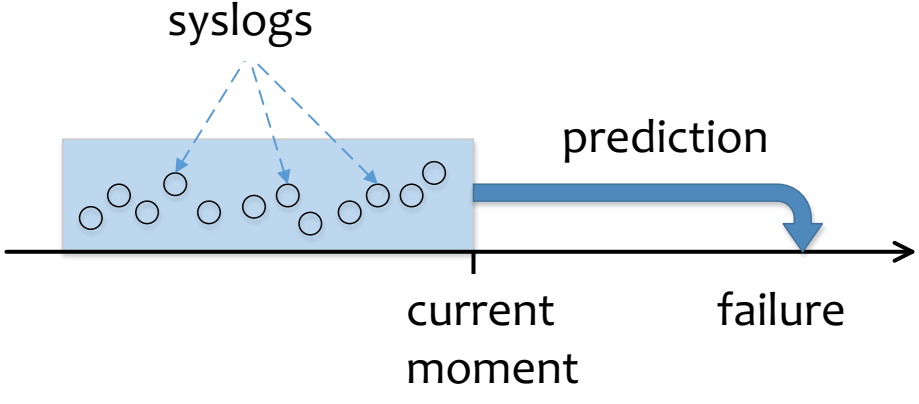


Collaboration with Baidu. Tencent implemented a variant to improve its video streaming service

DC Switch Failure Prediction->Preventive Replacement->Avoided Loss

Problem: Baidu-customized switches intermittently drop/delay packets, causing QoE drop at the application layer.

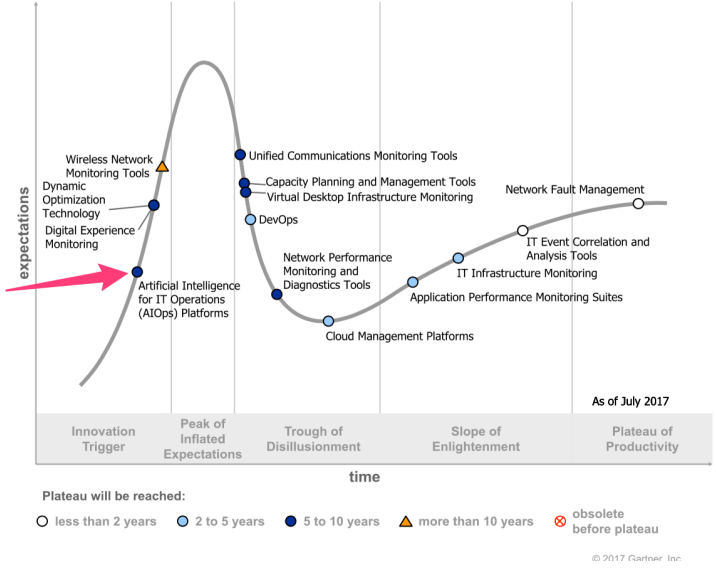
Reboot stops the problem for some while.
Question: Can we predict the this problem 2 hours before it happens again?
Then just switch the traffic away from this switch and reboot it.



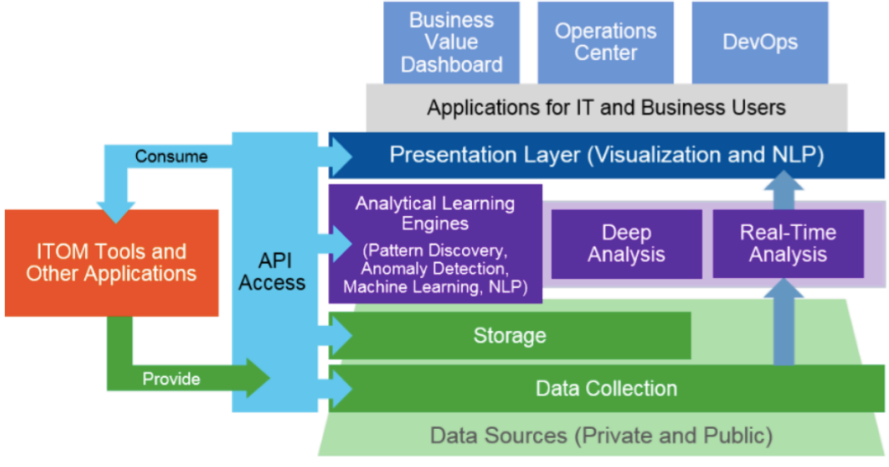
- Precision: 82.15%
- Recall: 74.74%
- FPR: 3.75×10^{-5}

AIOps is rising

- According to Gartner Report：
- AIOps global deployment ratio: 10% （2017）→ 50% （2020）



Source: Gartner (July 2017)

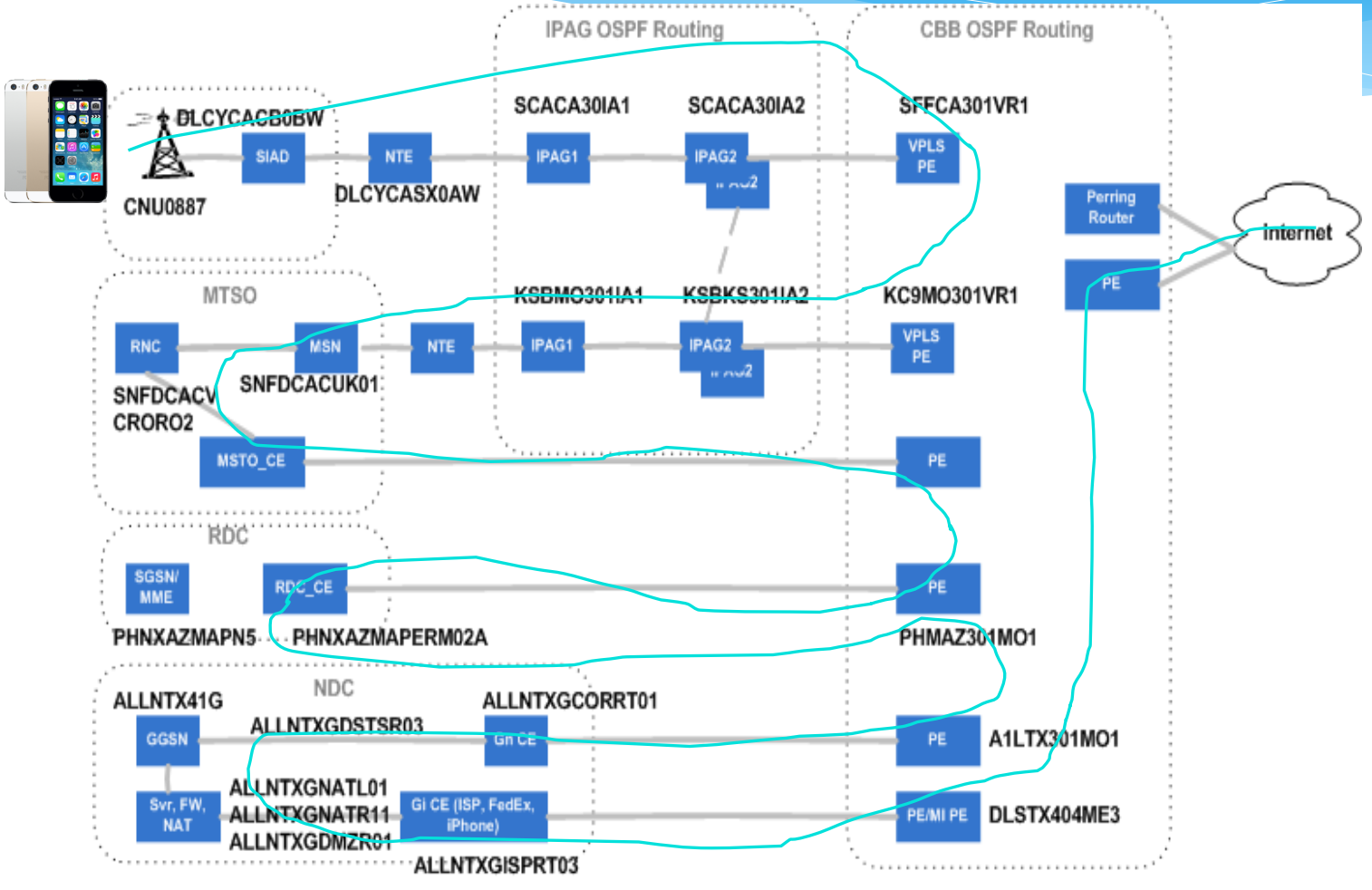


Source: Gartner (March 2016)

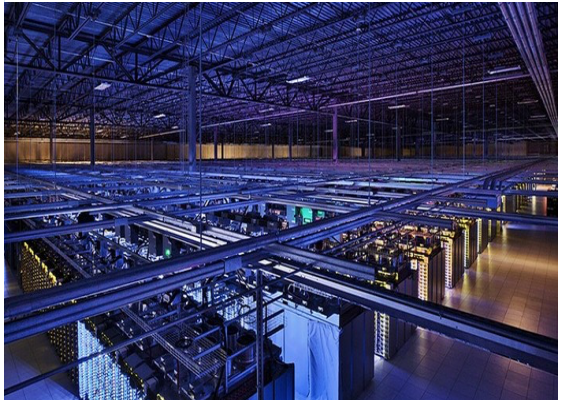
Overview

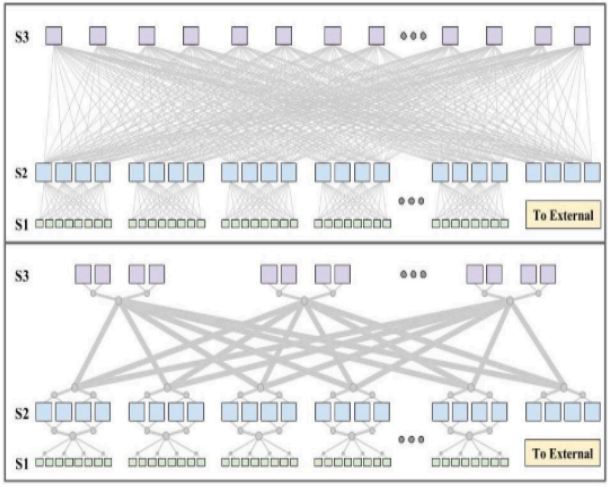
* What is AIOps?
* What's the value of AIOps?
* **Is AIOps necessary?**
* Is AIOps feasible?
* Case studies
* AIOps Challenge
* Summary

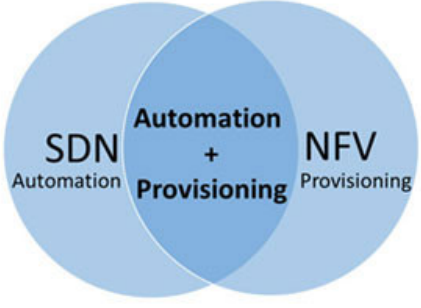
Complex Access Networks

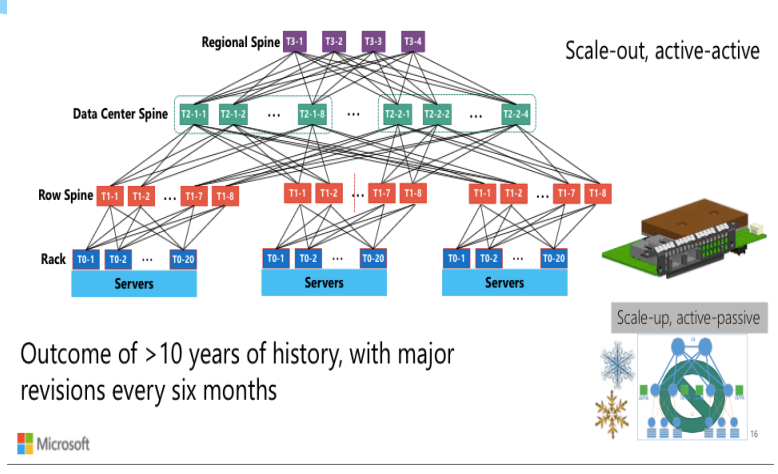


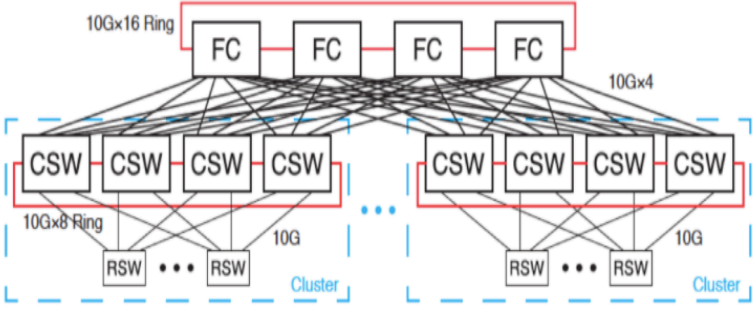
Complex and Evolving Cloud



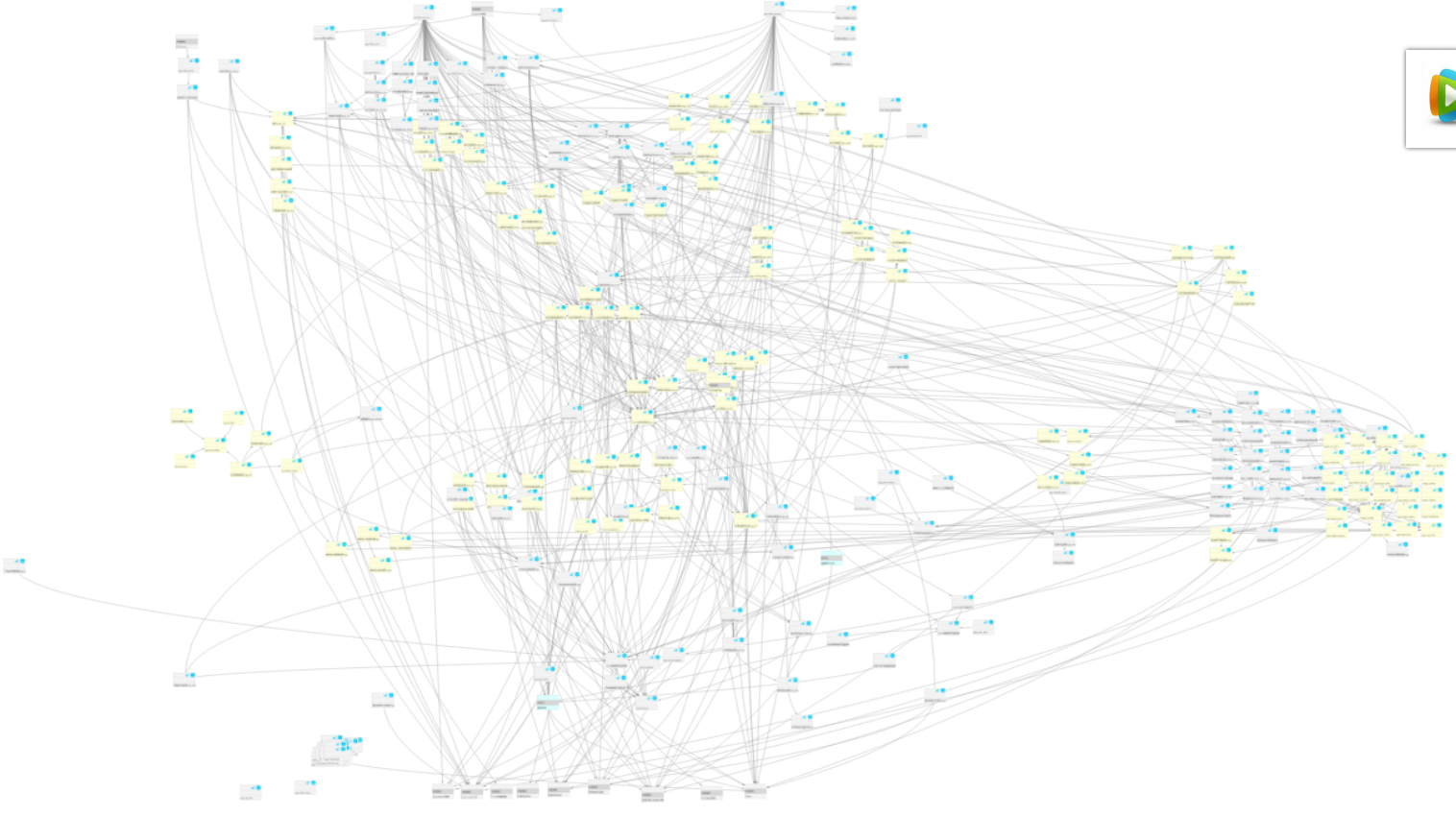




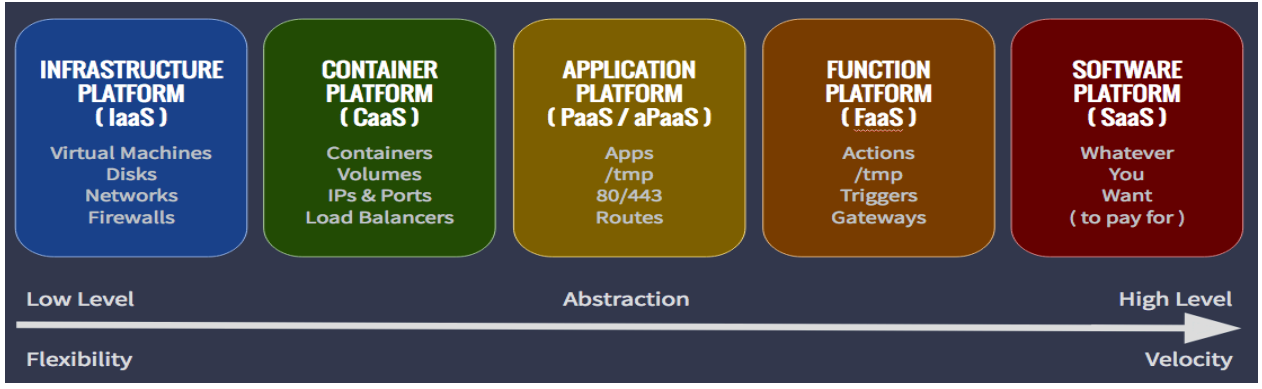


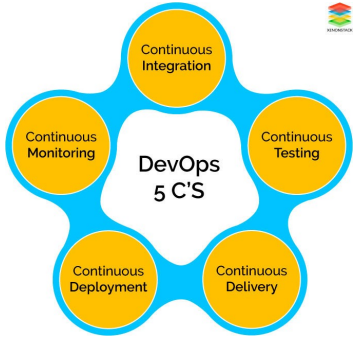


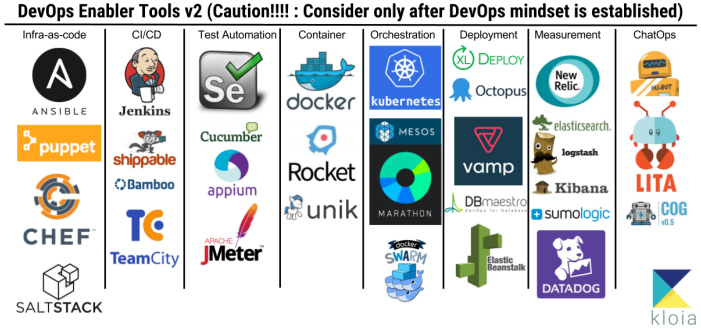
Complex and Evolving Software Module Dependences



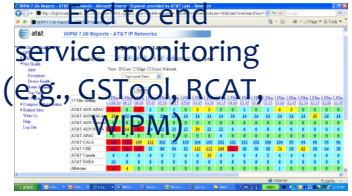
Evolving Techniques Enable Frequent Software Changes



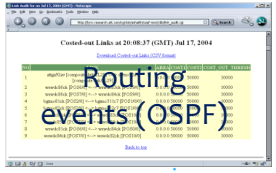




There are a sheer volume of device-generated log data during daily operations



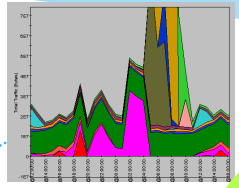
End to end service monitoring (e.g., GSTool, RCAT, WPM)



Routing events (OSPF)

Performance counters (e.g., Compass, Optima)

Logs (workflow, syslogs)



Lower layers (e.g., SONET, CNI)

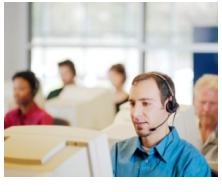
Customer issues (MTS, tickets, tweets)

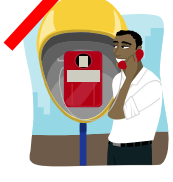
Troubleshooting

Alarms, tickets (e.g., Netcool, AOTS)

Customer trouble tickets

Network alerts





Customers

Network

We have no choice but relying on AI to take advantage of the Big Data from Ops

* Volume
* Velocity
* Variety
* Value

Overview

* What is AIOps?
* What's the value of AIOps?
* Is AIOps necessary?
* **Is AIOps feasible?**
* Case studies
* AIOps Challenge
* Summary

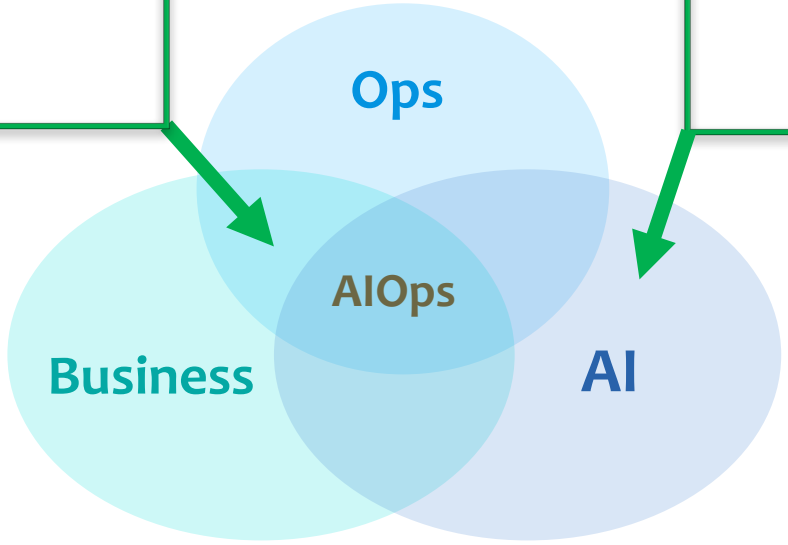
AIOps has *the necessities required for successful ML applications*

* Machine learning tools (algorithms and systems)
* *Applications that show the value*
* *Large amount of data*
* *Labels and the experts who can label*

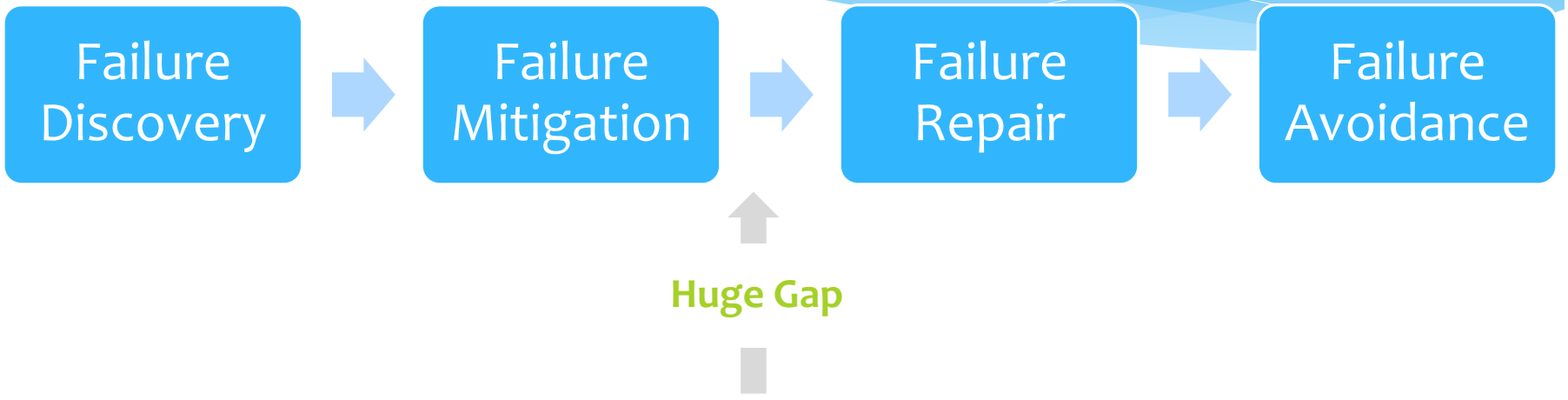
AIOps is still challenging because its the interdisciplinary nature

Ops people familiar with Ops and Business,
but not AI

Algorithm people familiar with general AI, but not Ops and AIOps



Pitfalls: use ML algorithms as Blackbox to tackle Ops challenges



General Machine Learning Algorithms

ARIMA, Time Series Decomposition, Holt-Winters, CUSUM, SST,DiD,DBSCAN, Pearson Correlation, J-Measure, Two-sample test, Apriori, FP-Growth, K-medoids, CLARIONS, Granger Causality, Logistic Regression, Correlation analysis (event-event, event-time series, time series-time series), hierarchical clustering、Decision tree, Random forest, support vector machine, Monte Carlo Tree search, Marcovian Chain, multi-instance learning, transfer learning, CNN, RNN ,VAE, GAN, NLP

AIOps needs architecture

Ops scenarios

Failure Discovery → Failure Mitigation → Failure Repair → Failure Avoidance

Each scenario needs an architecture of various algorithm components

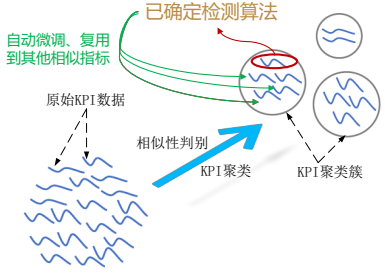
Each component solves a problem that AI is good at solving

(1) Abundant data
(2) Complete information
(3) Well defined
(4) Single domain

Anomaly detection for failure discovery becomes feasible after being divided into multiple problems.

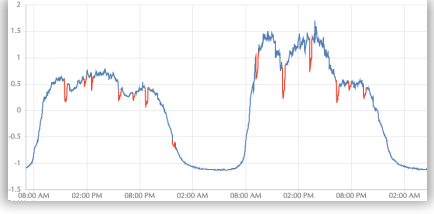
Unsupervised Anomaly Detection

KPI clustering and transfer learning to pick the features for unsupervised anomaly detection



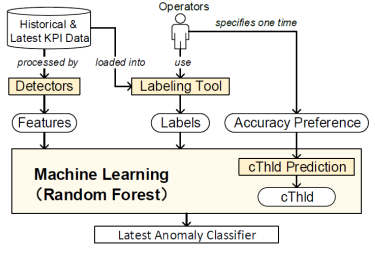
Feedback Tool

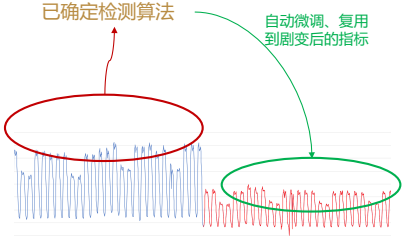
Fully utilize each user feedback's implication by automatically looking up similar anomalous regions.



Automatic algorithm/parameter adaptation

1. Basic on limited feedback, automatically pick algorithm and tune parameters.
2. Automatically adapt to level shift etc.



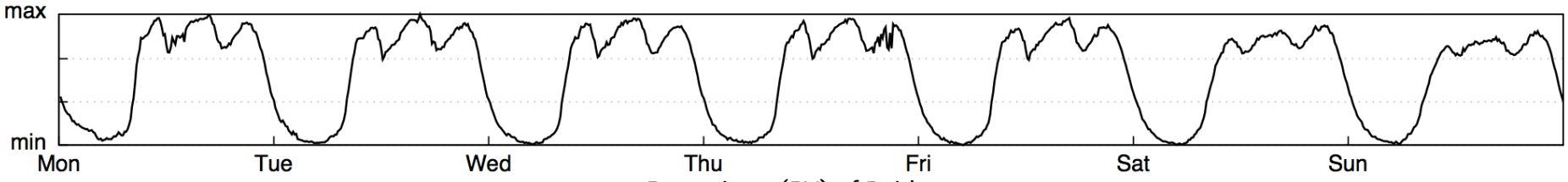


Overview

* What is AIOps?
* What's the value of AIOps?
* Is AIOps necessary?
* Is AIOps feasible?
* **Case studies**
* AIOps Challenge
* Summary

Case 1： KPI Anomaly Detection (Dapeng Liu et al., IMC 2015)

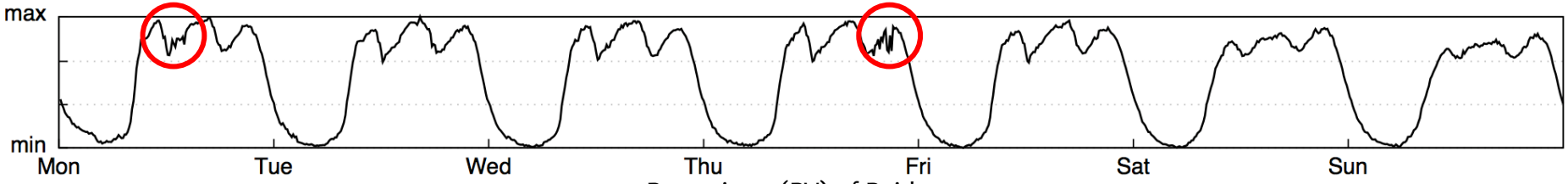
KPIs and Anomaly Detection



Page views (PV) of Baidu

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

KPIs and Anomaly Detection

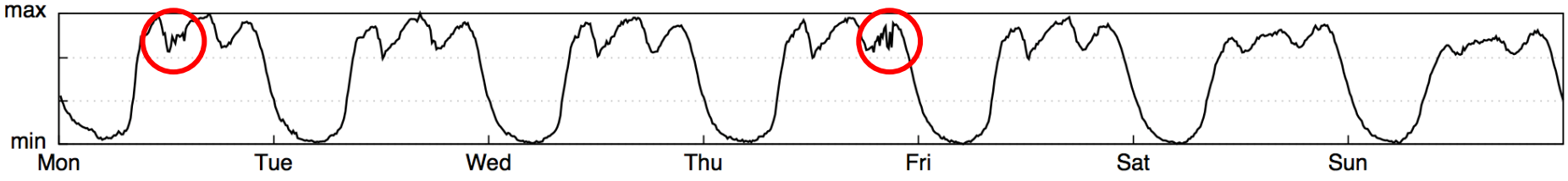


Page views (PV) of Baidu

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



Page views (PV) of Baidu

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

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

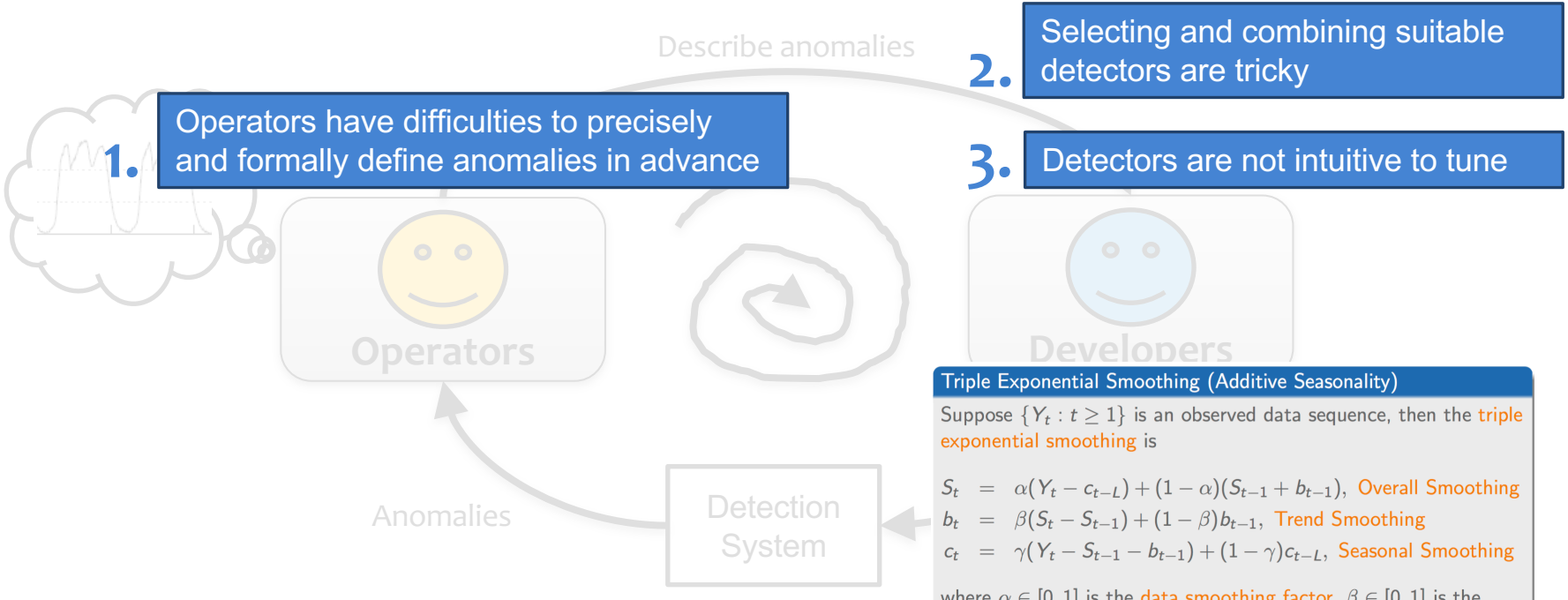
Anomaly detection matters: Find anomalous behaviors of the KPI curve

→ Diagnose and fix it

→ Avoid further influences and revenue losses

How to Build an Anomaly Detection System

Challenges



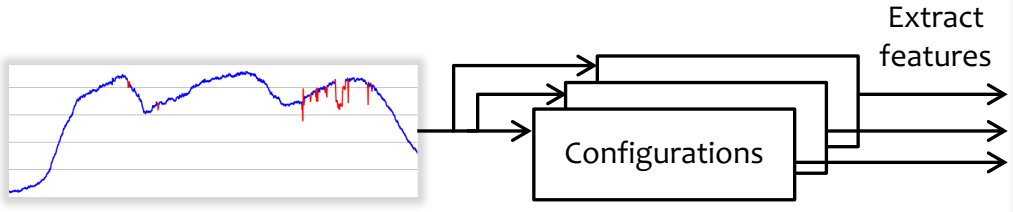
Triple Exponential Smoothing (Additive Seasonality)

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, then the triple exponential smoothing is

$$
\begin{aligned}
S_t &= \alpha(Y_t - c_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1}), \text{ Overall Smoothing} \\
b_t &= \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}, \text{ Trend Smoothing} \\
c_t &= \gamma(Y_t - S_{t-1} - b_{t-1}) + (1-\gamma)c_{t-L}, \text{ Seasonal Smoothing}
\end{aligned}
$$

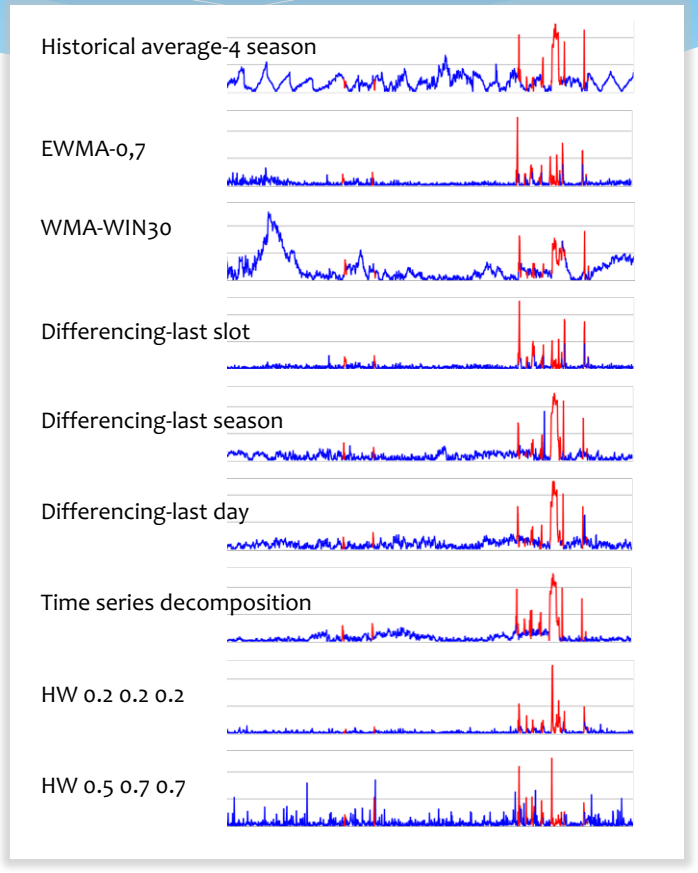
where $\alpha \in [0,1]$ is the data smoothing factor, $\beta \in [0,1]$ is the trend smoothing factor, $\gamma \in [0,1]$ is the seasonal change smoothing factor.

Key Ideas



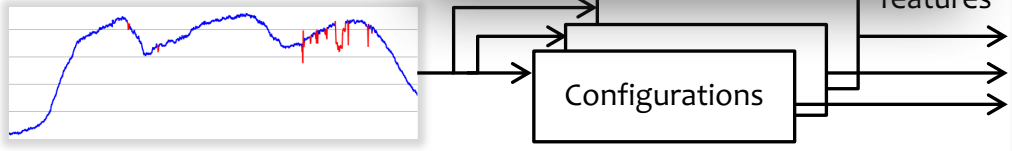
KPI data

(Detectors with different parameters)



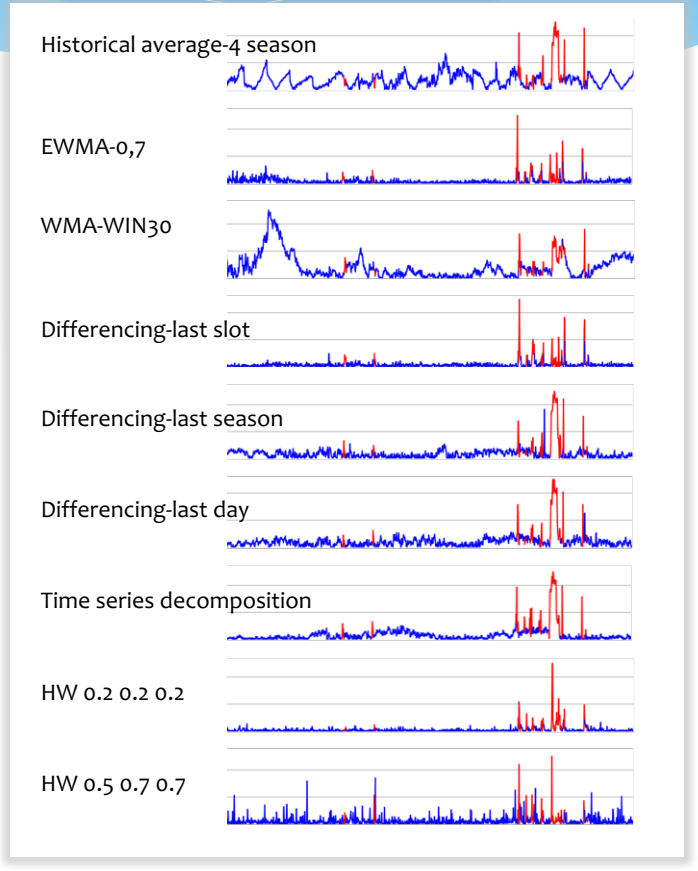
Key Ideas

Detector / # of configurations	Sampled parameters
Simple threshold [24] / 1	none
Diff / 3	last-slot, last-day, last-week
Simple MA [4] / 5	win = 10, 20, 30, 40, 50 points
Weighted MA [11] / 5	
MA of diff / 5	
EWMA [11] / 5	$\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$
TSD [1] / 5	win = 1, 2, 3, 4, 5 week(s)
TSD MAD / 5	
Historical average [5] / 5	
Historical MAD / 5	
Holt-Winters [6] / $4^3 = 64$	$\alpha, \beta, \gamma = 0.2, 0.4, 0.6, 0.8$
SVD [7] / $5 \times 3 = 15$	row =10, 20, 30, 40, 50 points, column =3, 5, 7
Wavelet [12] / $3 \times 3 = 9$	win = 3, 5 ,7 days, freq = low, mid, high
ARIMA [10] / 1	Estimation from data
In total: 14 basic detectors / 133 configurations	



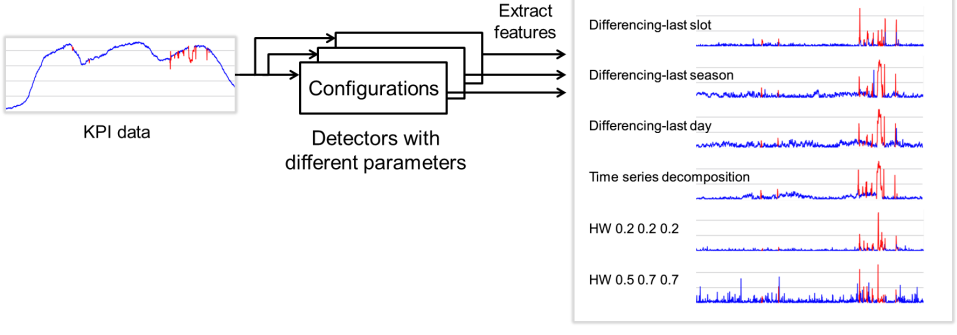
KPI data

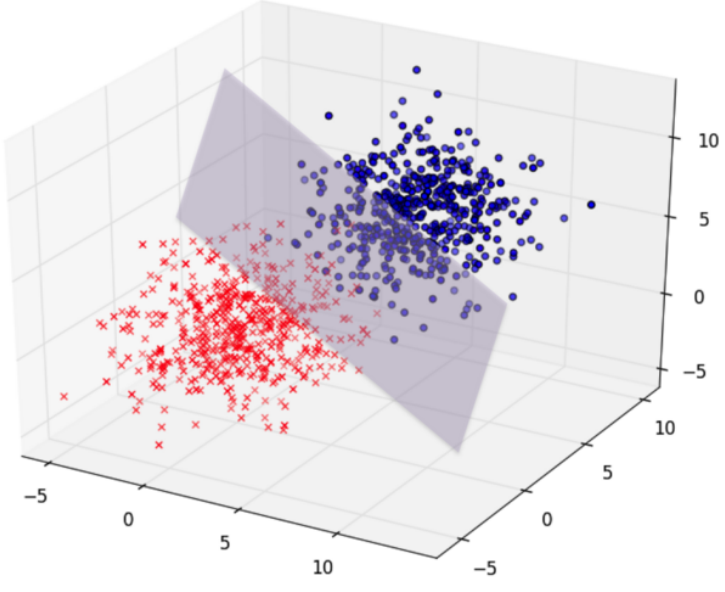
(Detectors with different parameters)



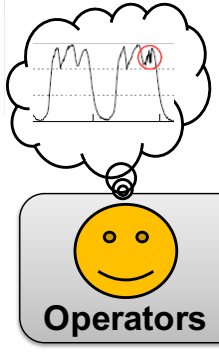
Key Ideas

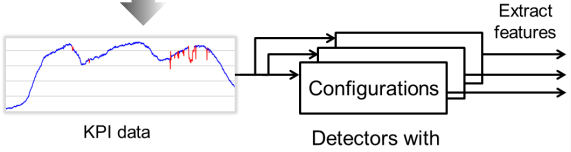
Classification in the feature space (Supervised machine learning)

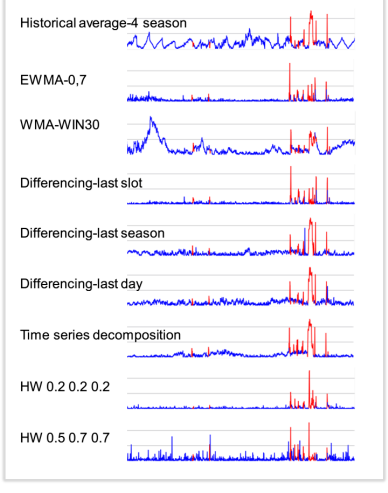




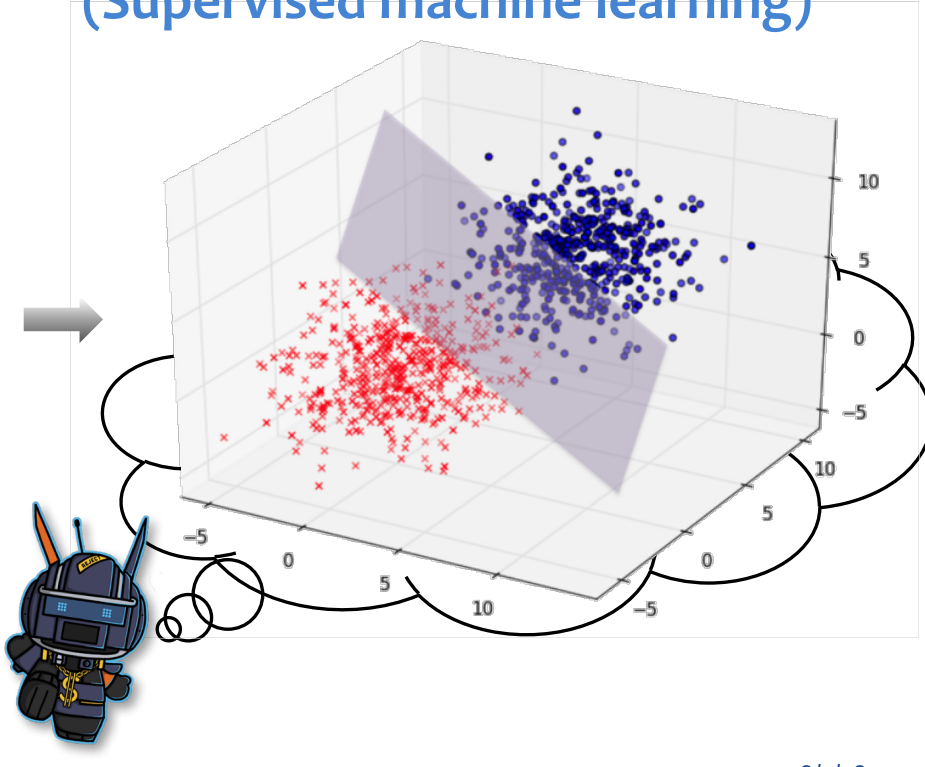
Key Ideas



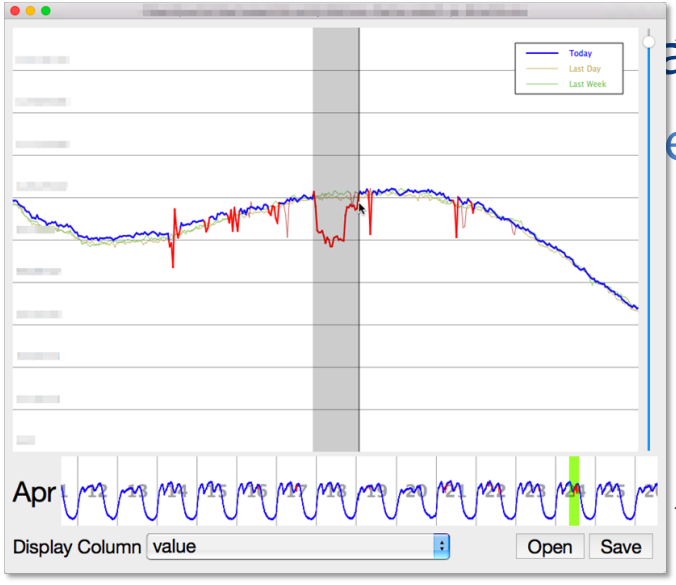




Classification in the feature space (Supervised machine learning)



Address Challenges of Designing Opprentice



Y axis scale slider

Labeling

drag

Label window

drag

Cancel labels

Time series control

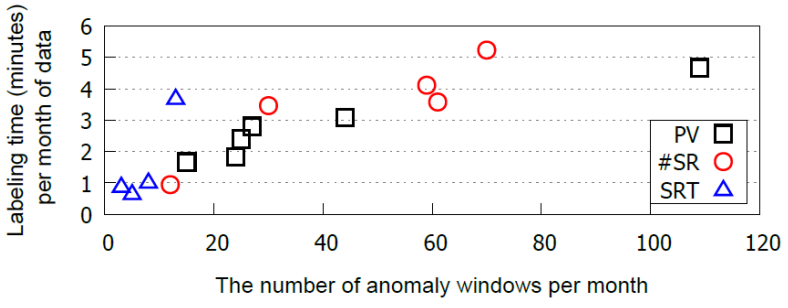
Zoom in

Backward

Forward

Zoom out

Navigator

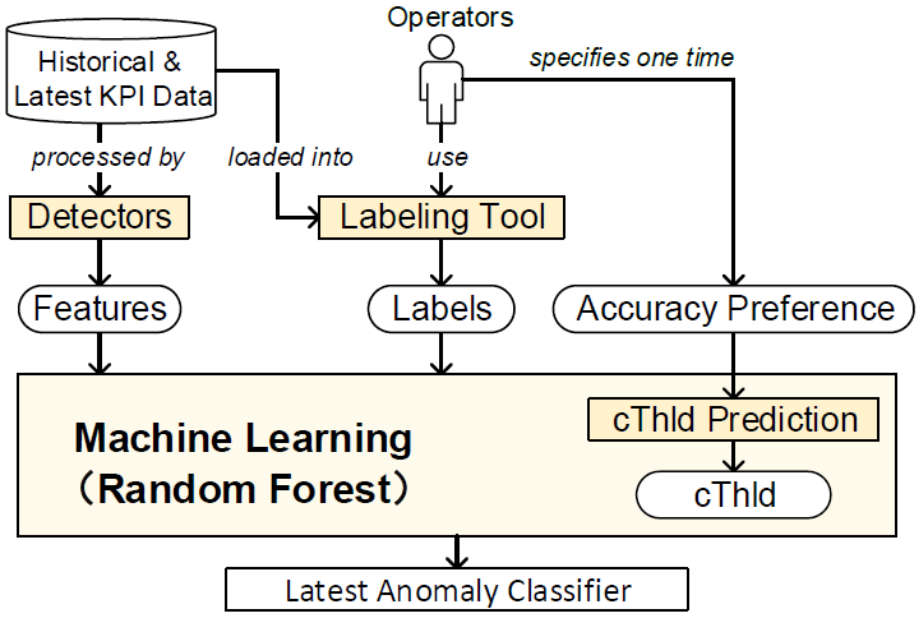


Address Challenges of Designing Opprentice

* 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
* Irrelevant and redundant features
  * Solution: random forests

Design Overview

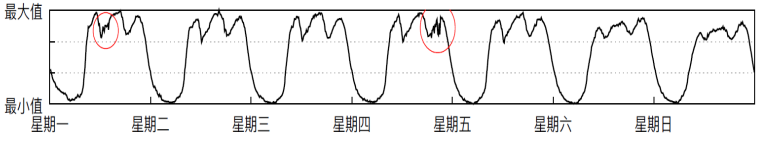
Training a classifier



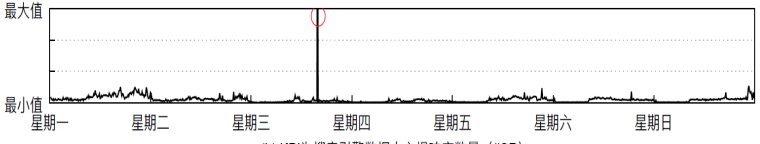
See the paper for full details

Detecting anomalies

Evaluation

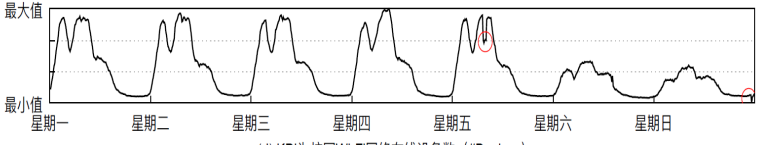


(a) KPI为搜索引擎访问量（PV）。



(b) KPI为搜索引擎数据中心慢响应数量（#SR）。

(c) KPI为搜索响应时间（SRT）。



(d) KPI为校园Wi-Fi网络在线设备数（#Devices）。

Search PV (25 weeks))

#slow queries（19 weeks）

Baidu

Search Response Time（16 weeks）

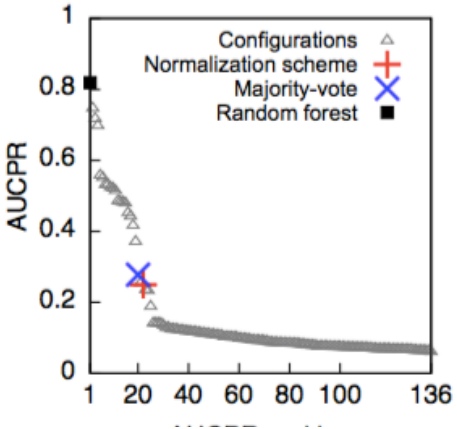
#online devices (15 weeks)

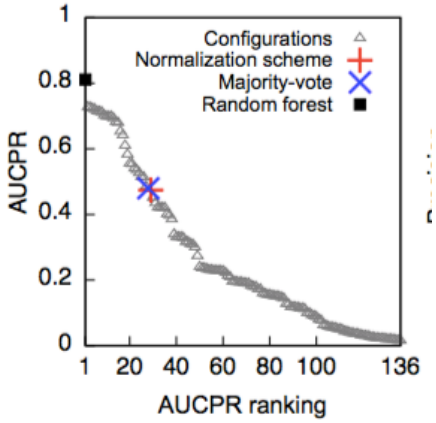
Tsinghua Enterprise WiFi

Evaluation

* Compared with all existing detectors （Four KPIs）

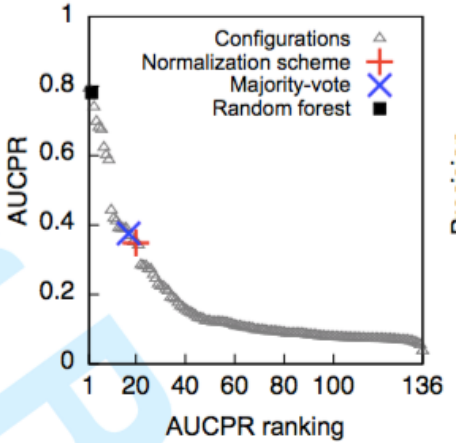
first

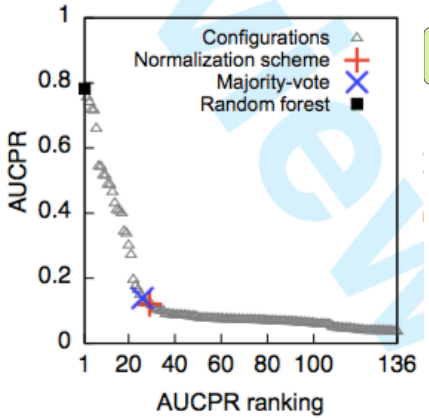




first

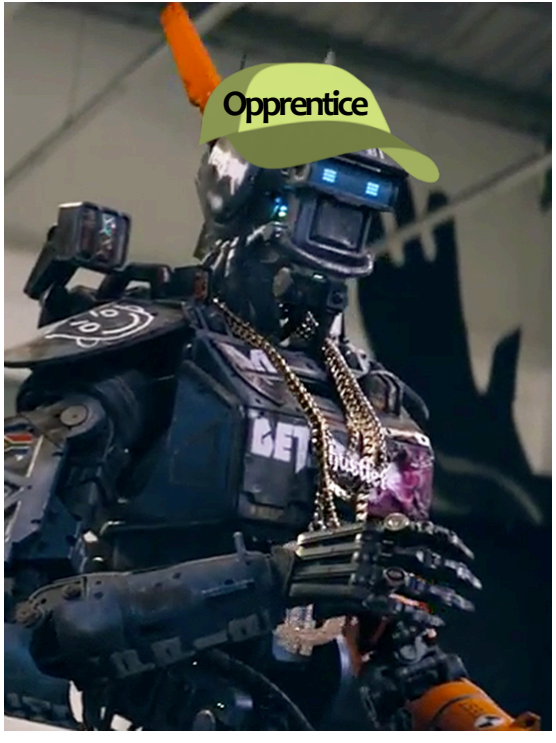
second





first

Case 1 summary



* Opprentice is an **automatic** and **accurate** machine learning framework for KPI anomaly detection

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

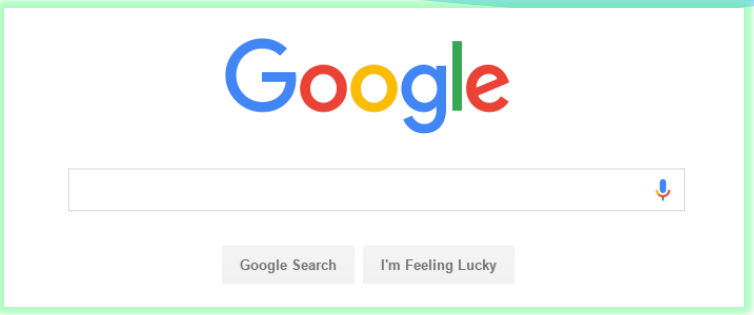
* The idea of Opprentice

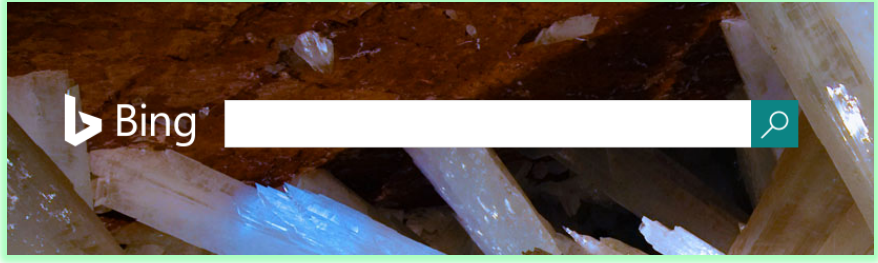
  i.e., **using machine learning to model the domain knowledge**

could be a very promising way to automate other service managements

Case 2 : Bottleneck Identification for Search Response Time (Dapeng Liu et al., INFOCOM 2016)

Web Search Engines



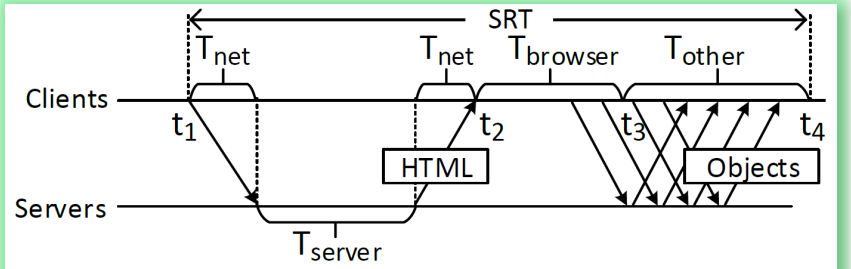


Search Response Time (SRT)

t_1 A search query is submitted

t_4 The result page Is rendered

$$SRT = t_4 - t_1$$



Search Response Time Matters

Bing

+500ms revenue ↓ 1.2%
[Eric Schurman, Bing]

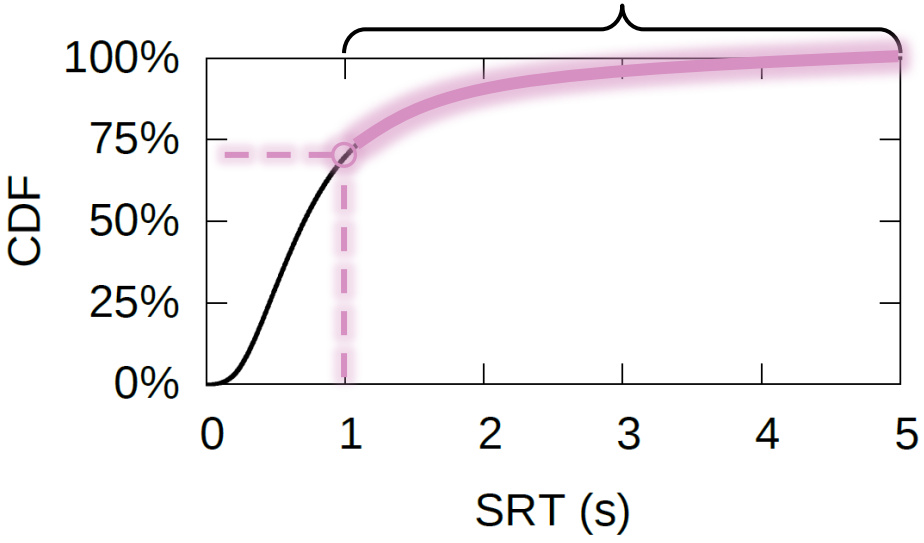
Google

+100ms~400ms queries ↓ 0.2%~0.6%
[Jake Brutlag, Google]

Given two content-wise identical search result pages, users are more likely to perform clicks on the fast page
[SIGIR 2014]

Search Response Time in the Wild

User's flow of thought is interrupted
if pages take **longer than 1s** to load



Why?

https://www.nngroup.com/articles/response-times-3-important-limits/

Monitoring SRT: Search Logs

Measurable attributes that can potentially impact SRT

SRT	User's ISP	Browser engine	# of Images	Ads	Server Load	…
800ms (Low SRT)	China Unicom	WebKit	10	Yes	1000 queries/s	…
1200ms (High SRT)	China Telecom	Trident 5.0	5	No	500 queries/s	…
……						

Goal of FOUCS

Measurable attributes that can potentially impact SRT

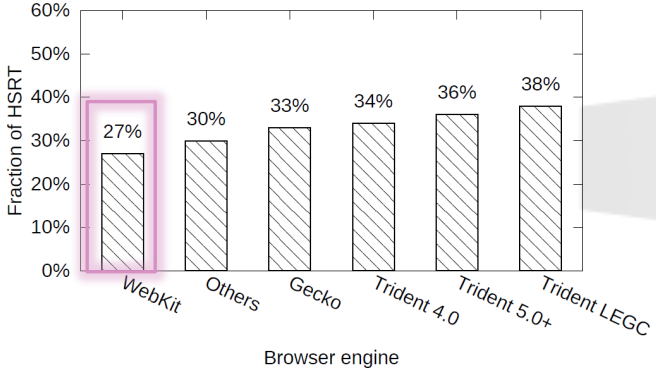
SRT	Client ISP	Browser engine	# of Images	Ads	Server Load	…
800ms (Low SRT)	China Unicom	WebKit	10	Yes	1000 queries/s	…
1200ms (High SRT)	China Telecom	Trident 5.0	5	No	500 queries/s	…
……						

We propose **FOCUS**, a search log analysis system to answer the following questions:

- Under what conditions **HSRT （High SRT）** is more likely to happen?
- Which HSRT conditions are similar (HSRT condition types)?
- How does each attribute affect SRT in HSRT condition types?

Challenges

Limited visibility of naïve single-dimension analysis



What we can see

WebKit is a good condition, where HSRT is **only 27%**

(e.g. used by Chrome and Safari)

What we **cannot** see

HSRT is **more than 38%**
when **"WebKit + #Images >30"**

Challenges

- **Limited visibility of naïve single-dimension analysis**
- **Interdependencies between attributes**
- **Overlapped HSRT conditions**

Key Idea of FOCUS

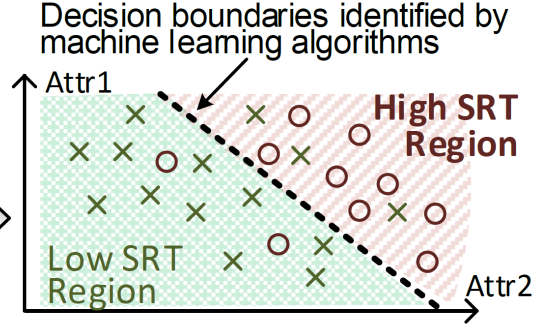
Limited visibility of naïve single-dimension analysis → Multi-dimension analysis

Interdependencies between attributes → Work with interdependencies

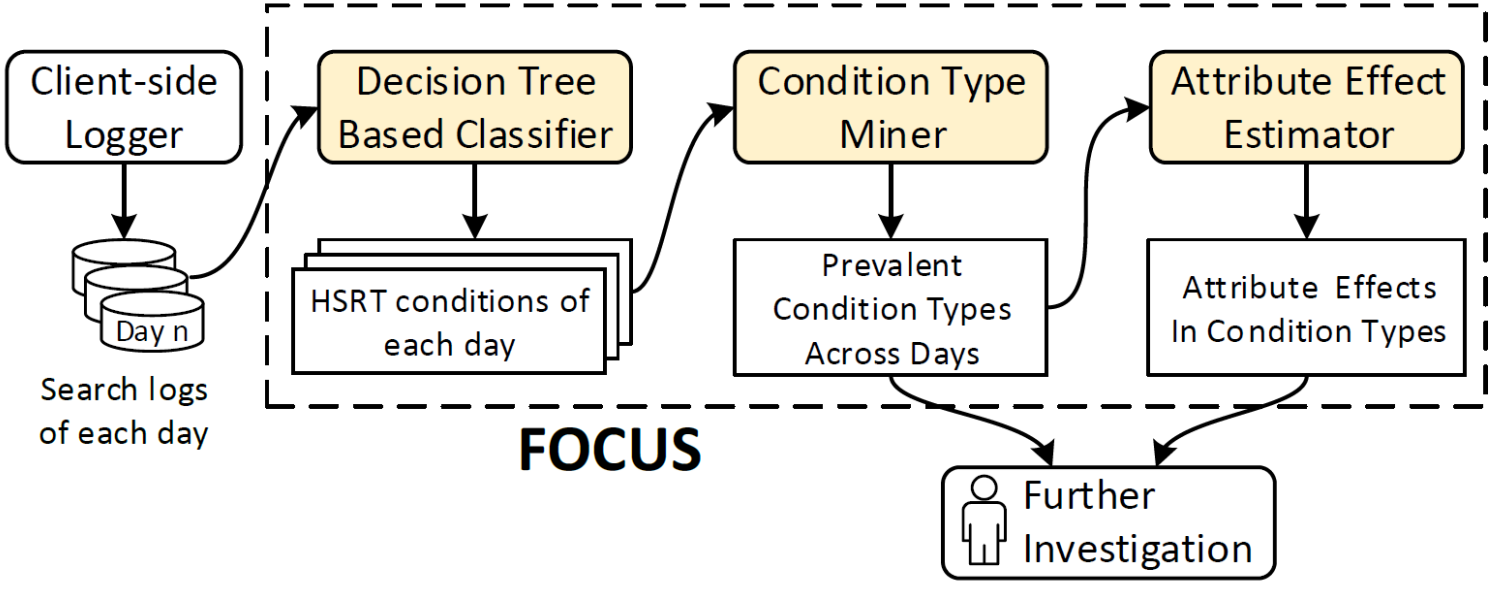
Overlapped HSRT conditions → Classification is non-overlap

- Model it as a classification problem
- Solve it using decision trees

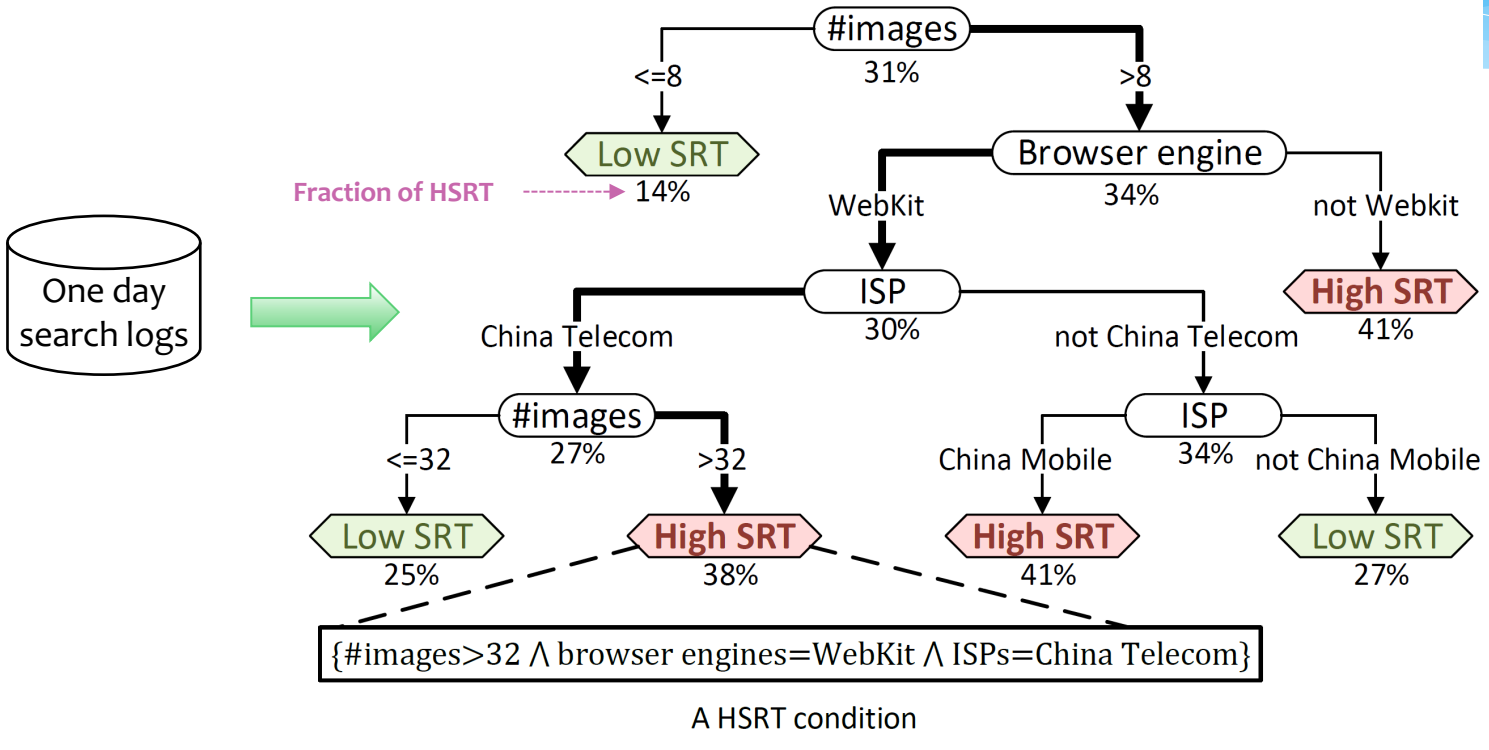
Attr1	Attr2	Label
…	…	High SRT ○
…	…	Low SRT ×
…	…	Low SRT ×
…	…	…



FOCUS Overview

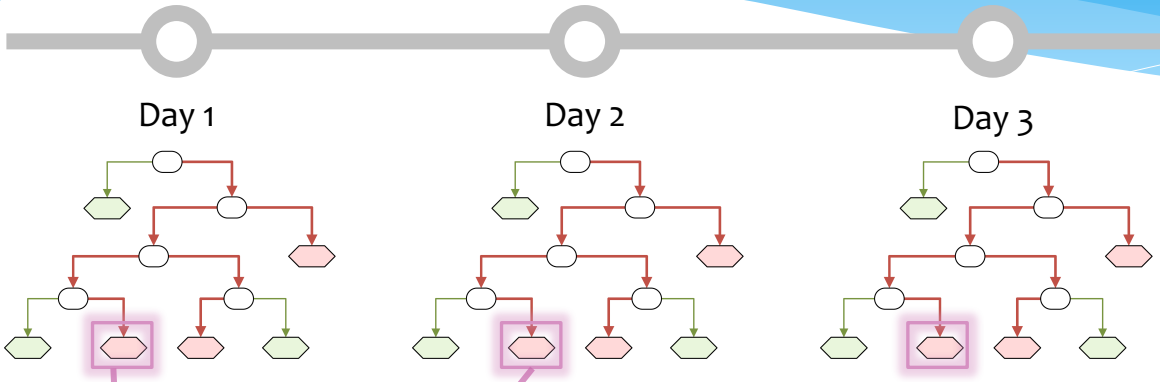


Identify HSRT Conditions Based on a Decision Tree



To build a reasonable tree, we **tailor** the mechanisms of decision trees (Details are in the paper)

Find Similar HSRT Conditions (HSRT Condition Types)



ID	HSRT Conditions		
	#Images	Browser engine	Ads
1	> 9	Not WebKit	no
2	> 10	Not WebKit	no

HSRT Condition Type		
#Images	Browser engine	Ads
$> i, i \in \{9,10\}$	Not WebKit	no

Hierarchical clustering

- Same combination of attributes
- Same value for each categorical attribute
- **Similar** value for each numeric attribute

Estimate the Impact of Each Attribute

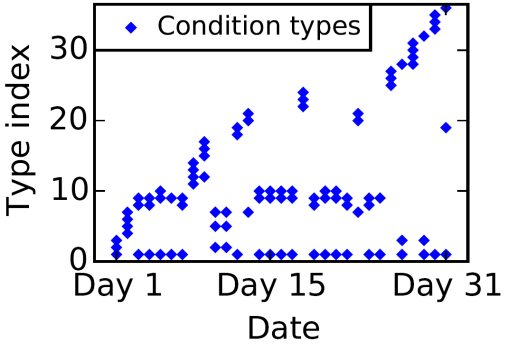
Inspired by controlled experiment

- **Control group:** the original HSRT contrition types
- **Experimental group:** changing one attribute at a time

Compare performance in historical logs

ID	HSRT Condition Type		
	#Images	Browser engine	Ads
C	$> i, i \in \{9,10\}$	Not WebKit	no
C_1	$\leq i, i \in \{9,10\}$	Not WebKit	no
C_2	$> i, i \in \{9,10\}$	WebKit	no
C_3	$> i, i \in \{9,10\}$	Not WebKit	yes

Results of FOCUS: Prevalent HSRT Condition Types



* Find 36 HSRT condition types in one month of search logs
* **Four** of them (11%) appear in more than five days

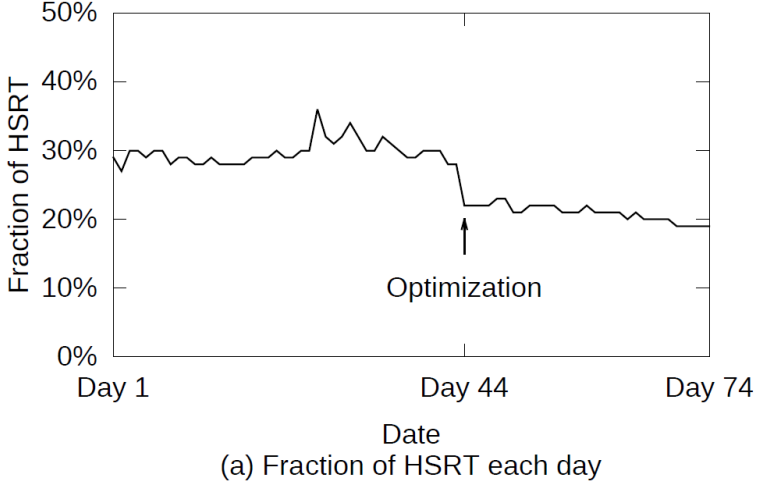
Images are the main bottleneck

(**Attributes** in bold have a bad effect on SRT)

Condition type ID	Prevalent condition type	Prevalence (days)
1	**#images** $> i, i \in \{5,6,7,8,9\}$ $\wedge$ **browser engine = not WebKit**	21
2	**#images** $> i, i \in \{5,6,7,8,9\}$ $\wedge$ **ISP = not China Telecom** $\wedge$ browser engine = WebKit	15
3	**#images** $> i, i \in \{25,26,27\}$ $\wedge$ ISP = China Telecom $\wedge$ browser engine = WebKit	7
4	**#images** $> i, i \in \{5,6,8\}$ $\wedge$ ISP = China Telecom $\wedge$ browser engine = WebKit $\wedge$ **ads = yes**	6

Real-world Optimization

* 1st month results of FOCUS → images are the main bottleneck of SRT
* Deploy “image base64 encoding" to improve the transmission time of images



(a) Fraction of HSRT each day

HSRT percentage is reduced by 30%

SRT 80th-tile is reduced by 253 ms (20%)

The fraction of HSRT is reduced by 30%

Case 2 Summary

* FOCUS can
  * Narrow down the debugging space of High SRT in search logs
  * Analyze the effects of each attribute (potential improvements)

* With the output of FOCUS
  * We make several interesting observations
  * Deploy a solution in practice and greatly improve SRT

* FOCUS is a general method for analyzing multi-attribute logs
  * Web applications other than search engines
  * Performance of mobile apps
  * …

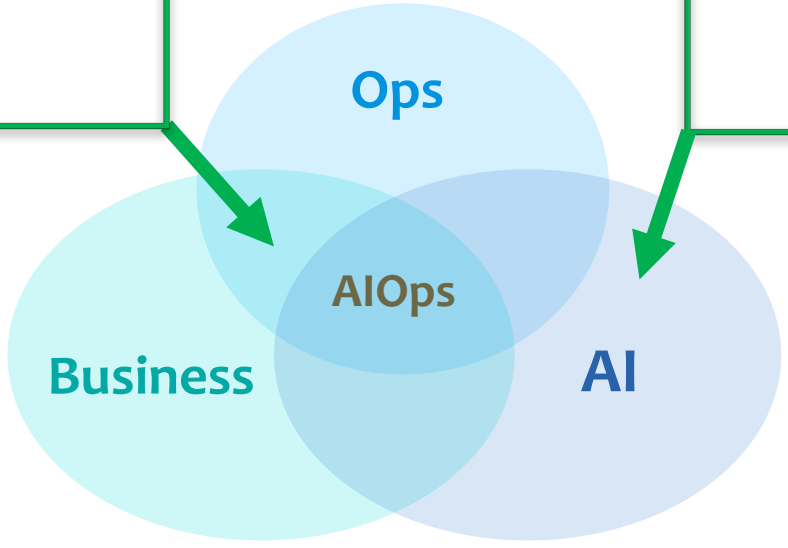
Overview

* What is AIOps?
* What's the value of AIOps?
* Is AIOps necessary?
* Is AIOps feasible?
* Case studies
* **AIOps Challenge**
* Summary

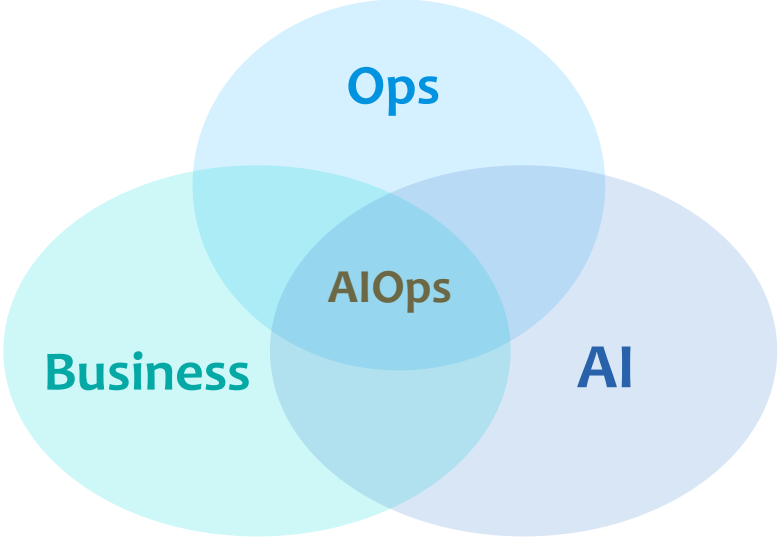
How to speed up the AIOps progress?

Ops people familiar with Ops and Business,
but not AI

Algorithm people familiar with general AI, but not Ops and AIOps



How to speed up the AIOps progress?



AIOps Challenge：Community Efforts to Make AIOps Happen

Ops

Give:

- Sanitized Data
- Competition sponsorship
- Bring up new challenges
- Forum discussion

Get:

- Algorithms
- Collaborators
- Recruiting
- Impact

Problems
Datasets
Competition
Forum

Give：

- Competition
- Forum discussion
- Help define new problems

Get：

- Real Ops problem
- Real data
- Collaborators
- Impact

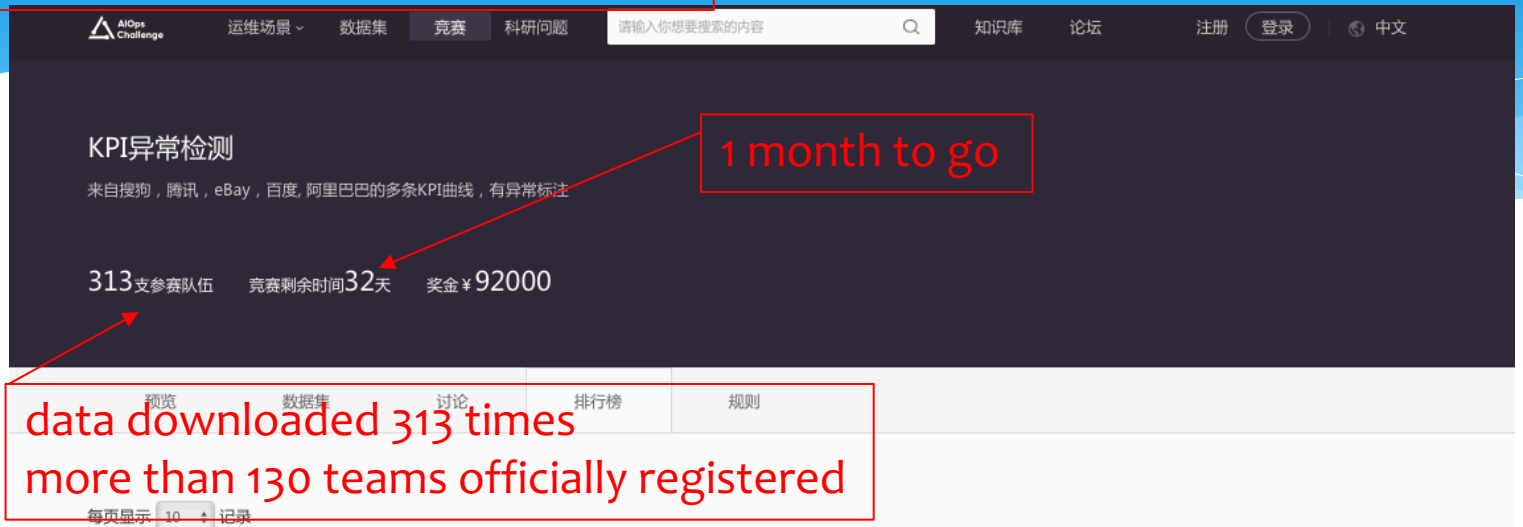
AIOps Challenge: http://iops.ai

亚军
8,000人民币
颁发获奖证书

冠军
80,000人民币
颁发获奖证书

季军
4,000人民币
颁发获奖证书

First competition: anomaly detection



1 month to go

data downloaded 313 times

more than 130 teams officially registered

每页显示 10 记录

队伍排名	队伍名字	队伍分数
1	LogicMonitor-AI	0.783962495189
2	bobochow	0.773771637041
3	wenchi zhang	0.73508053065
4	ComicSans	0.731338698777
5	古文	0.726198143588
6	stanley	0.715077886895
7	AnomalyGo	0.714373679043
8	jack-ma	0.696822275127
9	zhangchuang	0.657866967806
10	freecisco	0.654987434047

current top performer

分页 1 / 5

70

上一页 1 2 3 4 5 下一页

Data Sponsors

Sogou 搜狗

腾讯游戏 Tencent Games

ebay

Baidu 百度

阿里巴巴 Alibaba.com

Website

腾讯游戏 Tencent Games

co-organizer

高效运维社区

GreatOPS Community

Summary

* AIOps is rising: replace manual Ops decisions with AI-based decision aids
  * Improved revenue
  * reduced loss/cost
  * necessary
  * feasible

* Case studies (Collaboration with Baidu, Alibaba, Tencent, Didi, Sogou):
  * Anomaly Detection
  * Anomaly Localization
  * Root Cause Analysis
  * Capacity/Failure Prediction

* AIOps Challenge: Community efforts for widespread adoption of AIOps

THANK YOU