中山大學 软件工程学院

SUN YAT-SEN UNIVERSITY SCHOOL OF SOFTWARE ENGINEERING

SSE316：云计算技术
Cloud Computing Technology

陈壮彬

软件工程学院

https://zbchern.github.io/sse316.html

云系统智能运维

- 智能运维的概念
- 基于系统指标的智能运维
- 基于系统日志的智能运维
- 基于知识的智能运维

云系统智能运维

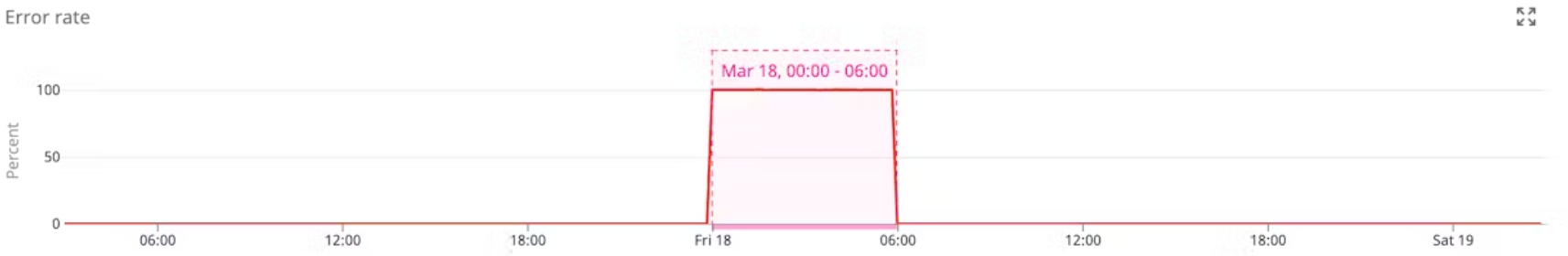
- 智能运维的概念
- 基于系统指标的智能运维
- 基于系统日志的智能运维
- 基于知识的智能运维

云系统运维

- 服务运维人员依靠实时收集、处理和分析分布式系统的信息和状态数据，以维持系统的正常运行并保障系统性能

- 指标（Metric）
- 日志（Log）
- 追踪（Trace）
- 告警（Alert）
- 工单（Ticket）

基于指标的运维



根据人工经验、规则设置阈值

基于日志的运维

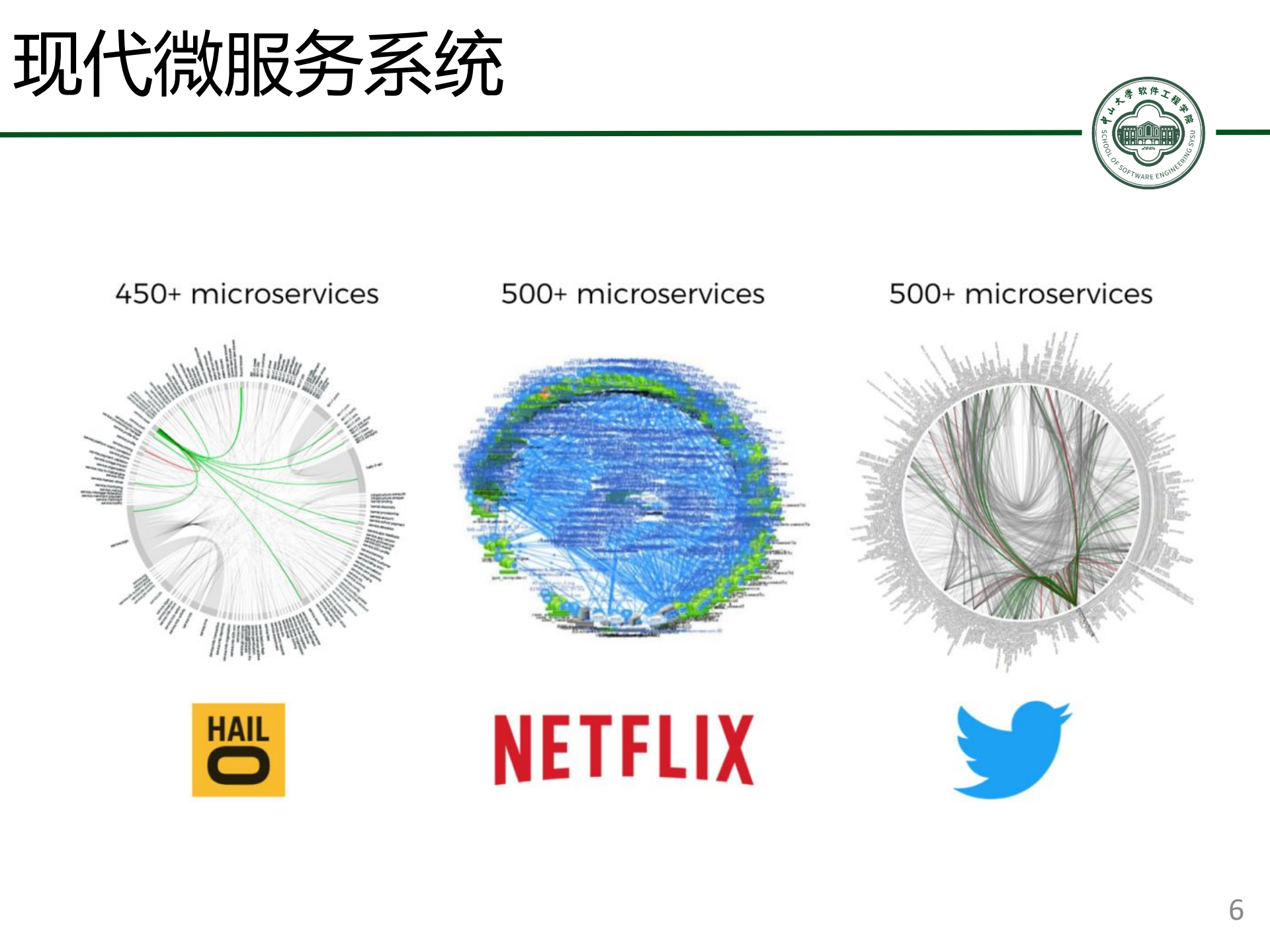
```
    public void failed(Throwable exc, Integer r) {
      LOG.debug("Failed while reading range {} ", r, exc);
      ranges.get(r).getData().completeExceptionally(exc);
    }
```

```
private void error(String category, String message, Object...args) {
  println("ERROR: %s: %s", category, String.format(message, args));
}
```

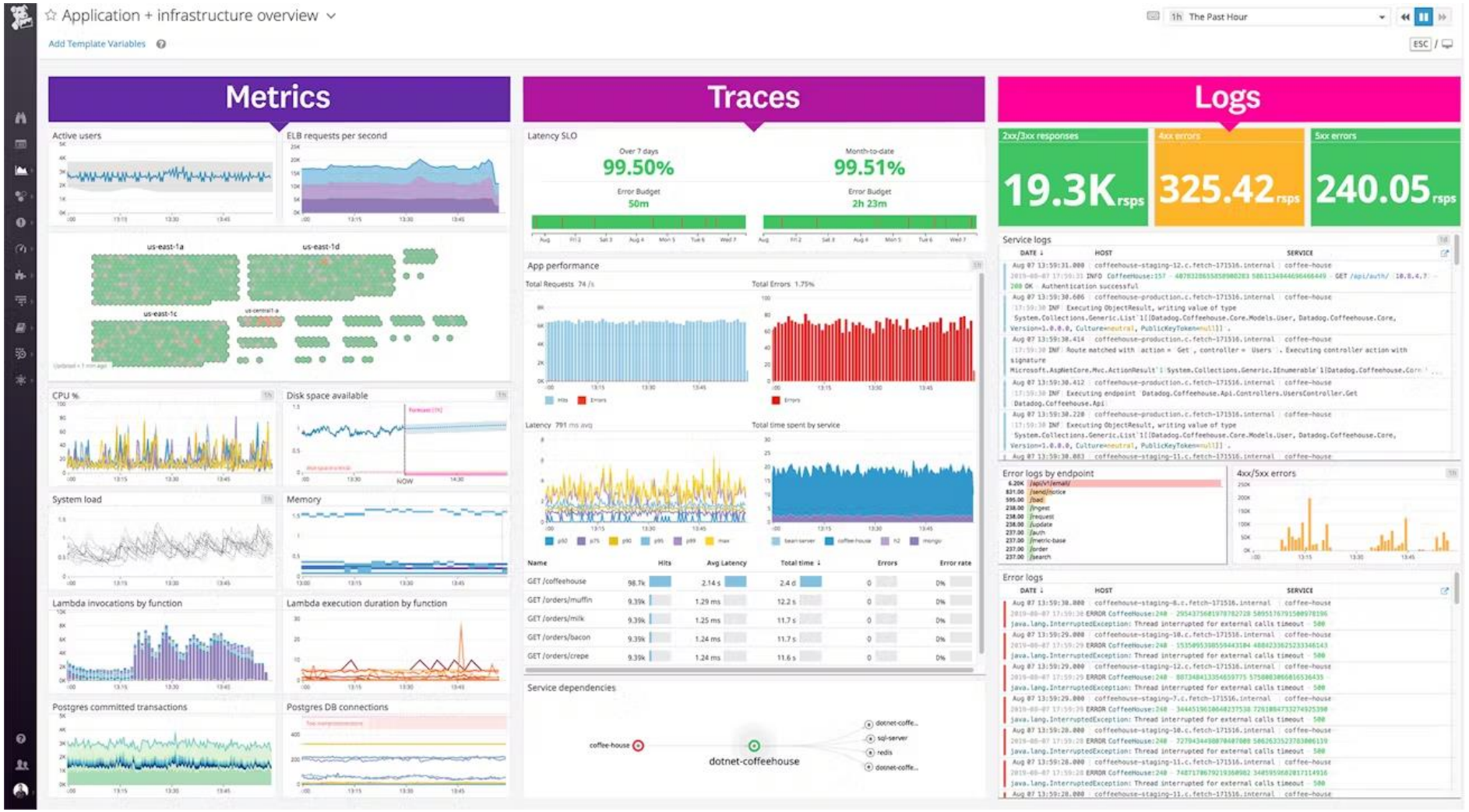
```
        try {
          duShell.startRefresh();
        } catch (IOException ioe) {
          LOG.warn("Could not get disk usage information for path {}",
              getDirPath(), ioe);
        }
      }
```

通过关键字搜索重要日志，比如Error，Exception，Failed

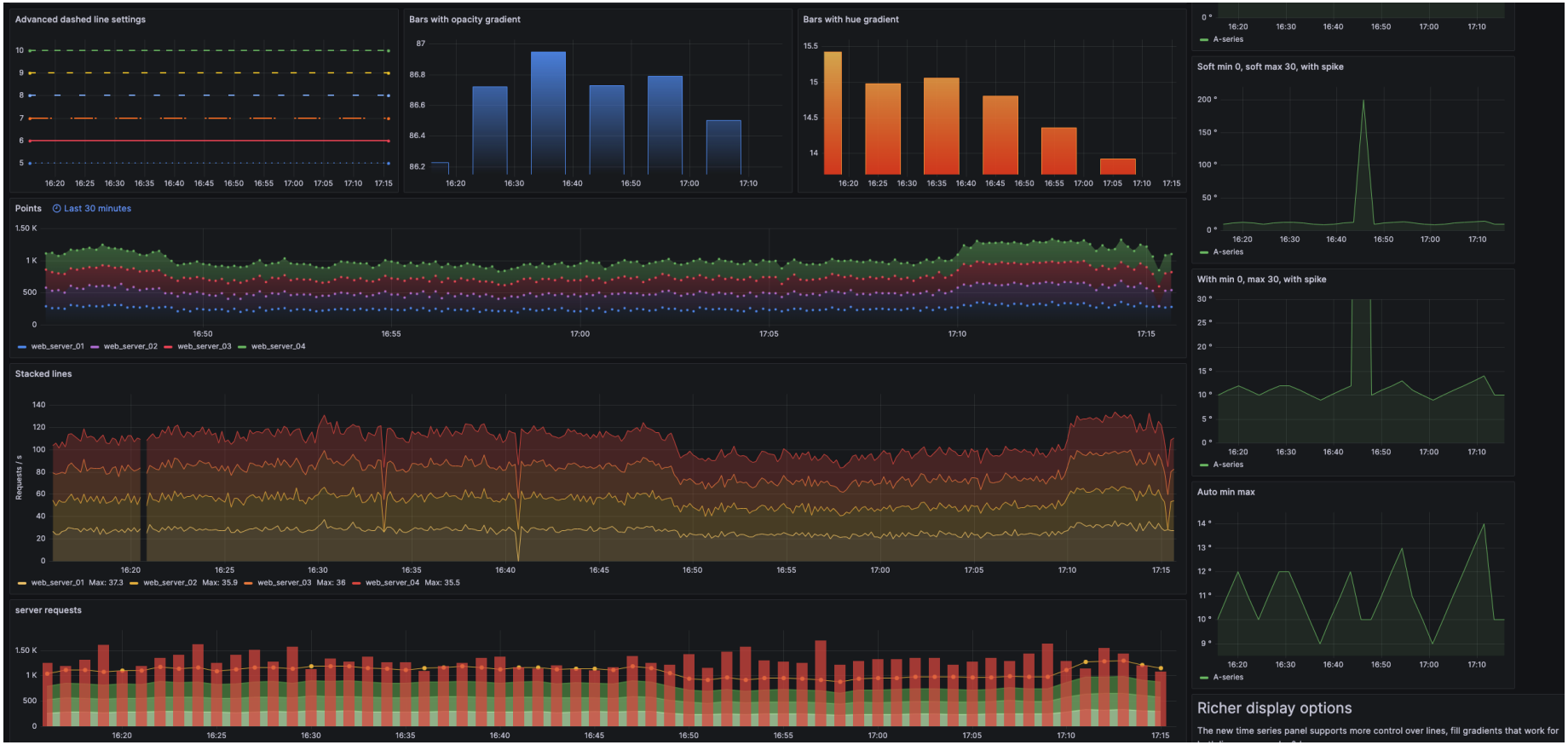
现代微服务系统


450+ microservices
500+ microservices
500+ microservices
HAIL O
NETFLIX

维服务系统监控数据


Application + infrastructure overview
Add Template Variables
1h The Past Hour
Metrics
Traces
Logs
Active users
ELB requests per second
us-east-1a
us-east-1d
us-east-1c
us-central1-a
CPU %
Disk space available
System load
Memory
Lambda invocations by function
Lambda execution duration by function
Postgres committed transactions
Postgres DB connections
Latency SLO
Over 7 days
99.50%
Error Budget
50m
Month-to-date
99.51%
Error Budget
2h 23m
App performance
Total Requests 74 /s
Total Errors 1.75%
Latency 791 ms avg
Total time spent by service
Name
Hits
Avg Latency
Total time
Errors
Error rate
GET /coffeehouse 98.7k 2.14 s 2.4 d 0 0%
GET /orders/muffin 9.39k 1.29 ms 12.2 s 0 0%
GET /orders/milk 9.39k 1.25 ms 11.7 s 0 0%
GET /orders/bacon 9.39k 1.24 ms 11.7 s 0 0%
GET /orders/crepe 9.39k 1.24 ms 11.6 s 0 0%
Service dependencies
coffee-house
dotnet-coffeehouse
dotnet-coffe...
sql-server
redis
dotnet-coffe...
2xx/3xx responses
19.3K rsps
4xx errors
325.42 rsps
5xx errors
240.05 rsps
Service logs
DATE
HOST
SERVICE
Error logs by endpoint
6.20K /api/v1/email/
831.00 /send/notice
595.00 /bad
238.00 /ingest
238.00 /request
238.00 /update
237.00 /auth
237.00 /metric-base
237.00 /order
237.00 /search
4xx/5xx errors
Error logs

百万级监控曲线

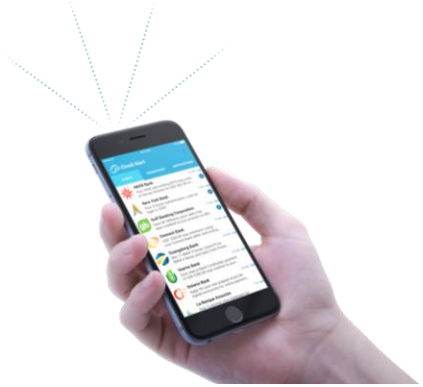

Advanced dashed line settings
Bars with opacity gradient
Bars with hue gradient
Points
Last 30 minutes
web_server_01
web_server_02
web_server_03
web_server_04
Stacked lines
Requests / s
web_server_01 Max: 37.3
web_server_02 Max: 35.9
web_server_03 Max: 36
web_server_04 Max: 35.5
server requests
A-series
Soft min 0, soft max 30, with spike
With min 0, max 30, with spike
Auto min max
Richer display options
The new time series panel supports more control over lines, fill gradients that work for

海量日志

- 搜索 Exception

```
081110 020724 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_2568309208894455676 is added to invalidSet of 10.251.31.160:50010
081110 022226 6281 INFO dfs.DataNode$DataXceiver: 10.250.19.16:50010 Served block blk_1078000656626961731 to /10.250.19.16
081110 023456 6415 WARN dfs.DataNode$DataXceiver: 10.251.67.225:50010:Got exception while serving blk_-6900989714336081087 to /10.251.25.237:
081110 024834 6371 INFO dfs.DataNode$DataXceiver: 10.251.126.227:50010 Served block blk_-8306714721294235181 to /10.251.126.227
081110 030331 6561 WARN dfs.DataNode$DataXceiver: 10.251.42.191:50010:Got exception while serving blk_-8023826090828946372 to /10.251.214.130:
081110 030942 6646 WARN dfs.DataNode$DataXceiver: 10.251.31.5:50010:Got exception while serving blk_-1367876730256254709 to /10.251.67.225:
081110 031019 6604 INFO dfs.DataNode$DataXceiver: 10.251.126.83:50010 Served block blk_-3860894070657427592 to /10.251.126.83
081110 032126 6555 WARN dfs.DataNode$DataXceiver: 10.251.26.8:50010:Got exception while serving blk_-7983508786213002472 to /10.251.38.197:
081110 035357 6606 INFO dfs.DataNode$DataXceiver: 10.251.71.97:50010 Served block blk_5454332143498402824 to /10.250.15.67
081110 040800 6739 INFO dfs.DataNode$DataXceiver: 10.250.6.214:50010 Served block blk_-3384560576963801177 to /10.250.6.214
081110 042826 6827 INFO dfs.DataNode$DataXceiver: 10.251.67.4:50010 Served block blk_-2901225370888235702 to /10.251.91.84
081110 045413 6957 INFO dfs.DataNode$DataXceiver: 10.251.215.50:50010 Served block blk_3963698556744744747 to /10.251.107.196
081110 050300 6683 INFO dfs.DataNode$DataXceiver: 10.251.30.134:50010 Served block blk_2039230511363331616 to /10.251.65.203
081110 051323 6956 INFO dfs.DataNode$DataXceiver: 10.251.71.146:50010 Served block blk_8482111101480751895 to /10.251.71.146
081110 054642 7145 INFO dfs.DataNode$DataXceiver: 10.251.215.16:50010 Served block blk_-5919767990596301121 to /10.251.215.16
081110 060453 7193 INFO dfs.DataNode$DataXceiver: 10.251.199.225:50010 Served block blk_8457344665564381337 to /10.251.199.225
081110 060934 7211 INFO dfs.DataNode$DataXceiver: 10.251.66.102:50010 Served block blk_2986720270598512615 to /10.251.66.102
081110 065635 7324 INFO dfs.DataNode$DataXceiver: 10.251.90.64:50010 Served block blk_-5719934513583495857 to /10.251.199.245
081110 070326 7430 WARN dfs.DataNode$DataXceiver: 10.251.75.228:50010:Got exception while serving blk_-7680599654910200999 to /10.251.75.228:
081110 070334 7327 INFO dfs.DataNode$DataXceiver: 10.251.111.37:50010 Served block blk_-6050976999174805557 to /10.251.111.37
081110 070347 7574 INFO dfs.DataNode$DataXceiver: 10.250.10.6:50010 Served block blk_1598414622053793245 to /10.251.90.239
081110 070614 7513 INFO dfs.DataNode$DataXceiver: 10.250.19.16:50010 Served block blk_-7837339190764609698 to /10.251.197.161
081110 070921 7744 INFO dfs.DataNode$DataXceiver: 10.251.39.144:50010 Served block blk_-8187008844253719581 to /10.251.91.32
081110 071154 7627 INFO dfs.DataNode$DataXceiver: 10.251.214.112:50010 Served block blk_4081177399275502985 to /10.251.110.68
081110 071426 7742 INFO dfs.DataNode$DataXceiver: 10.251.42.191:50010 Served block blk_3515154079719300106 to /10.251.42.191
081110 071657 7700 WARN dfs.DataNode$DataXceiver: 10.251.38.214:50010:Got exception while serving blk_5547569777499890340 to /10.251.195.52:
081110 072121 7820 INFO dfs.DataNode$DataXceiver: 10.251.123.99:50010 Served block blk_5385121219895615240 to /10.251.71.146
081110 072124 7772 WARN dfs.DataNode$DataXceiver: 10.251.42.246:50010:Got exception while serving blk_-7658293778087733436 to /10.251.30.179:
081110 080546 7970 WARN dfs.DataNode$DataXceiver: 10.251.111.130:50010:Got exception while serving blk_3169060243663461885 to /10.251.214.32:
081110 080555 8227 INFO dfs.DataNode$DataXceiver: 10.250.11.194:50010 Served block blk_3087787567144441647 to /10.251.91.84
081110 080718 7850 INFO dfs.DataNode$DataXceiver: 10.250.10.100:50010 Served block blk_-3657665801189425193 to /10.250.10.100
081110 080724 8080 INFO dfs.DataNode$DataXceiver: 10.251.74.79:50010 Served block blk_-3457731723401426942 to /10.251.74.79
081110 080814 8123 INFO dfs.DataNode$DataXceiver: 10.251.42.84:50010 Served block blk_6105506155797750768 to /10.251.42.84
081110 080847 8088 WARN dfs.DataNode$DataXceiver: 10.251.26.8:50010:Got exception while serving blk_-8522942048313632858 to /10.251.31.85:
081110 080922 7633 INFO dfs.DataNode$DataXceiver: 10.251.203.246:50010 Served block blk_3654963980623381141 to /10.251.203.246
081110 080949 7994 INFO dfs.DataNode$DataXceiver: 10.250.19.227:50010 Served block blk_3979872751691718643 to /10.250.19.227
081110 081044 8125 WARN dfs.DataNode$DataXceiver: 10.251.123.1:50010:Got exception while serving blk_-272707591443354058 to /10.251.198.33:
081110 081054 8108 WARN dfs.DataNode$DataXceiver: 10.251.90.239:50010:Got exception while serving blk_-8679916835272129336 to /10.250.15.198:
081110 081337 8145 WARN dfs.DataNode$DataXceiver: 10.251.66.102:50010:Got exception while serving blk_6106884317961925960 to /10.251.66.102:
081110 081515 8312 WARN dfs.DataNode$DataXceiver: 10.251.31.5:50010:Got exception while serving blk_6332892729727950039 to /10.251.123.1:
081110 081643 8095 INFO dfs.DataNode$DataXceiver: 10.251.195.52:50010 Served block blk_6655622109568310643 to /10.251.195.52
081110 081643 8247 WARN dfs.DataNode$DataXceiver: 10.250.17.225:50010:Got exception while serving blk_-5935642747315643391 to /10.251.199.150:
081110 081741 8169 WARN dfs.DataNode$DataXceiver: 10.251.215.70:50010:Got exception while serving blk_-20269367189114433 to /10.251.30.179:
081110 082013 8326 INFO dfs.DataNode$DataXceiver: 10.251.43.115:50010 Served block blk_-7364557883931785608 to /10.251.43.115
081110 082043 8305 INFO dfs.DataNode$DataXceiver: 10.251.215.16:50010 Served block blk_2322432806134104317 to /10.251.215.16
081110 082444 8172 INFO dfs.DataNode$DataXceiver: 10.251.109.209:50010 Served block blk_4848669047361069041 to /10.251.26.177
081110 082702 8223 WARN dfs.DataNode$DataXceiver: 10.250.15.198:50010:Got exception while serving blk_-1851265222873801714 to /10.251.195.52:
081110 082706 8552 WARN dfs.DataNode$DataXceiver: 10.251.198.33:50010:Got exception while serving blk_-8495670552887053546 to /10.250.10.223:
081110 082737 8341 WARN dfs.DataNode$DataXceiver: 10.250.19.227:50010:Got exception while serving blk_-7372087176866857012 to /10.251.110.68:
081110 082954 8543 WARN dfs.DataNode$DataXceiver: 10.251.70.112:50010:Got exception while serving blk_4357276972386184626 to /10.251.74.79:
081110 083045 8495 WARN dfs.DataNode$DataXceiver: 10.251.215.16:50010:Got exception while serving blk_-4590972095204776122 to /10.251.30.6:
081110 083121 8197 INFO dfs.DataNode$DataXceiver: 10.251.106.214:50010 Served block blk_-8277873627721528374 to /10.251.122.79
081110 083231 8530 INFO dfs.DataNode$DataXceiver: 10.250.10.176:50010 Served block blk_3797494971676497497 to /10.251.29.239
081110 083328 8416 INFO dfs.DataNode$DataXceiver: 10.251.126.22:50010 Served block blk_805587860540600864 to /10.251.126.22
081110 083453 13 INFO dfs.DataBlockScanner: Verification succeeded for blk_3141363517520802396
081110 085042 13 INFO dfs.DataBlockScanner: Verification succeeded for blk_-1383276859207001328
081110 085933 13 INFO dfs.DataBlockScanner: Verification succeeded for blk_-4117999745005013424
081110 091216 7593 WARN dfs.DataNode$DataXceiver: 10.251.107.50:50010:Got exception while serving blk_4524198807982839635 to /10.251.122.79:
081110 091550 8683 WARN dfs.DataNode$DataXceiver: 10.251.111.209:50010:Got exception while serving blk_7505828172725463922 to /10.251.111.209:
081110 091657 8720 WARN dfs.DataNode$DataXceiver: 10.251.70.211:50010:Got exception while serving blk_424255210146453297 to /10.251.203.179:
081110 091800 8380 INFO dfs.DataNode$DataXceiver: 10.250.14.224:50010 Served block blk_666713934549639791 to /10.250.14.224
081110 092131 8815 INFO dfs.DataNode$DataXceiver: 10.251.30.179:50010 Served block blk_-2975629975082443857 to /10.251.30.179
081110 092151 8601 WARN dfs.DataNode$DataXceiver: 10.251.126.22:50010:Got exception while serving blk_1686195200514944346 to /10.250.6.223:
081110 092459 8650 INFO dfs.DataNode$DataXceiver: 10.250.9.207:50010 Served block blk_4355450627202483068 to /10.250.9.207
081110 092815 8872 INFO dfs.DataNode$DataXceiver: 10.250.6.191:50010 Served block blk_5952254363678329024 to /10.250.6.191
081110 093020 8827 WARN dfs.DataNode$DataXceiver: 10.251.70.211:50010:Got exception while serving blk_-6679331171485085225 to /10.251.203.246:
081110 093643 13 INFO dfs.DataBlockScanner: Verification succeeded for blk_9188832735514090334
081110 093831 8571 INFO dfs.DataNode$DataXceiver: 10.251.106.10:50010 Served block blk_2273334621242106674 to /10.251.106.10
081110 094019 8808 WARN dfs.DataNode$DataXceiver: 10.251.125.193:50010:Got exception while serving blk_3790492230047189408 to /10.251.199.159:
081110 094657 7835 INFO dfs.DataNode$DataXceiver: 10.251.107.50:50010 Served block blk_-2285729896739318683 to /10.251.70.5
```

告警运维



Wake up at 3am!

云系统人工运维

基于人工的方式无法应对大规模且复杂的运维数据

云系统智能运维

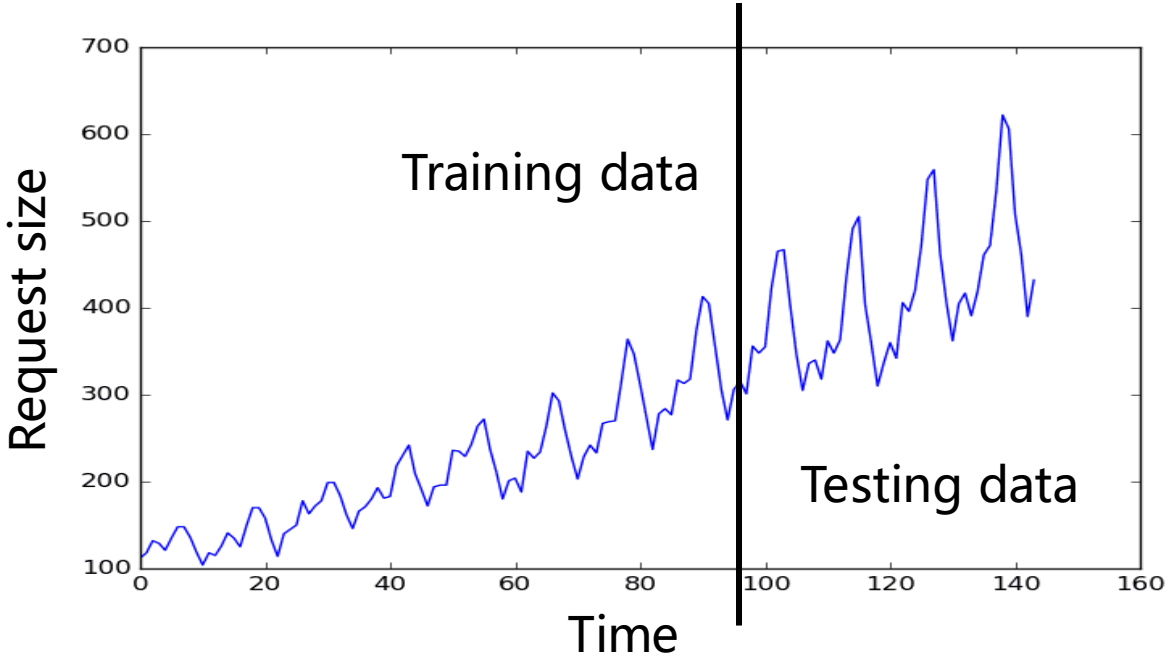
AIOps (Artificial intelligence for IT Operations)：引入机器学习，以数据驱动的方法做云系统运维。

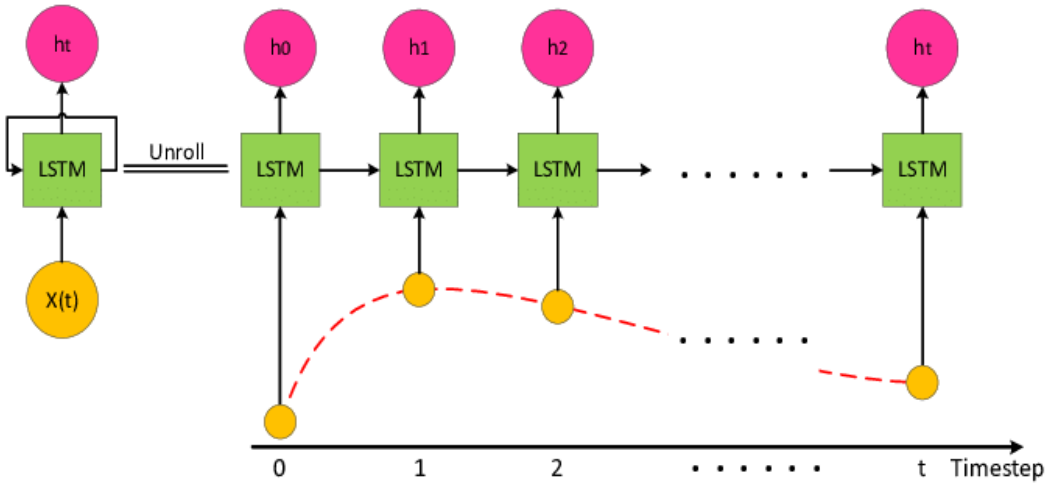
- 典型智能运维任务
  - ✓异常检测（Anomaly detection）
  - ✓故障分析（Failure analysis）
  - ✓故障预测（Failure prediction）
  - ✓根因定位（Root cause localization）
  - ✓软件自动测试（Software testing）
  - ✓自动修复（Auto-healing）
  - ✓…

云系统智能运维

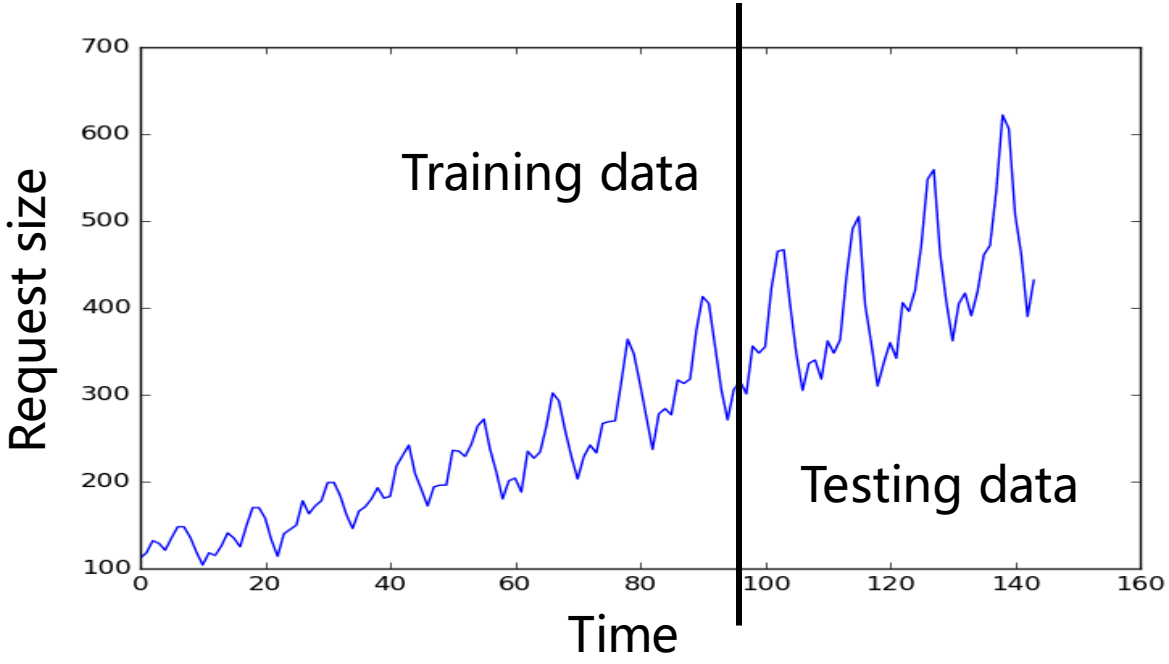
- 智能运维的概念
- 基于系统指标的智能运维
- 基于系统日志的智能运维
- 基于知识的智能运维

基于LSTM的指标异常预测



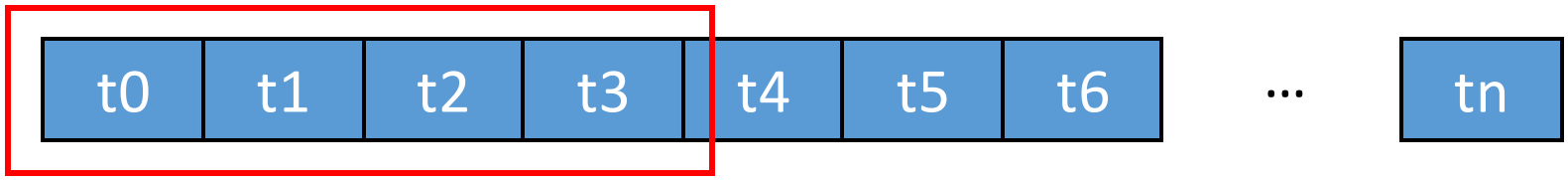


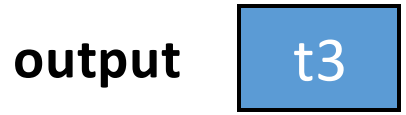
数据处理


Request size
Time
Training data
Testing data
700
600
500
400
300
200
100
0
20
40
60
80
100
120
140
160

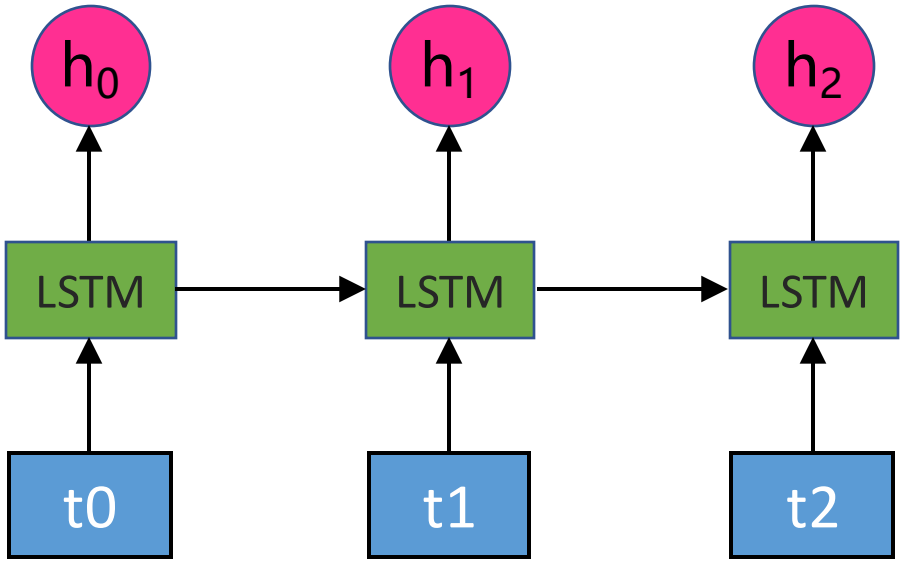

t0
t1
t2
t3
t4
t5
t6
...
tn

模型输入输出



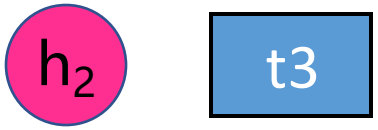


基于LSTM的资源需求预测



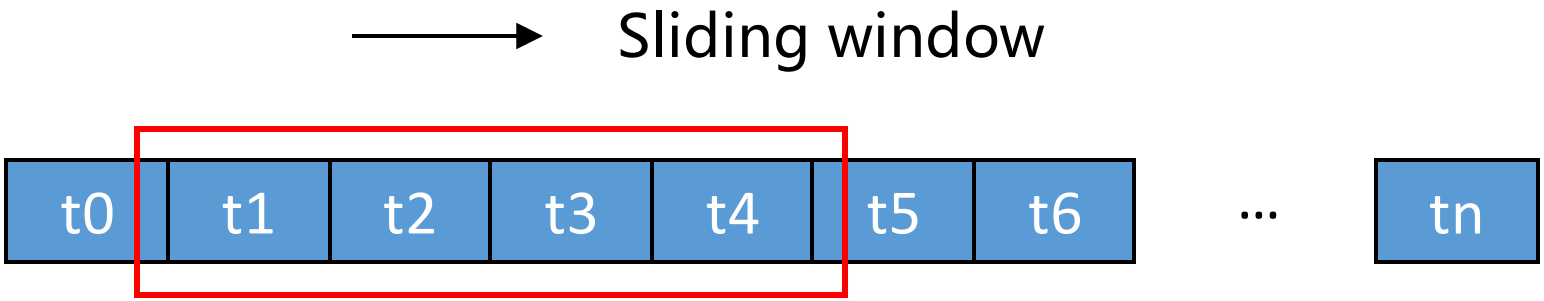
Mean square error (MSE)

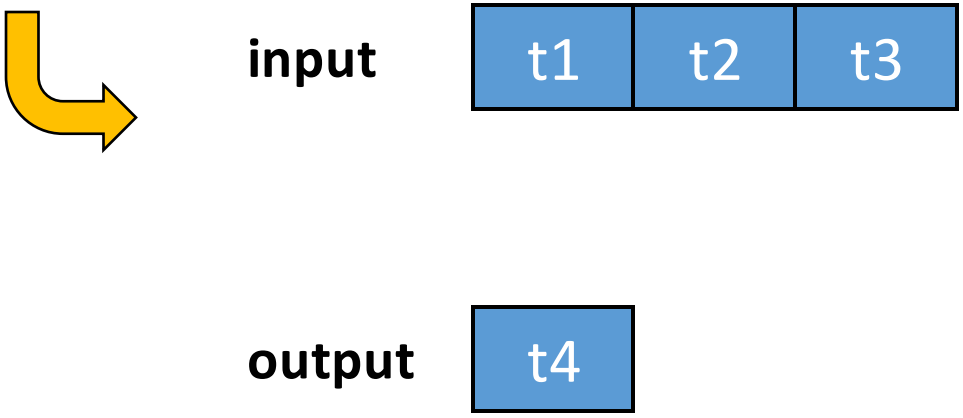
$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$$



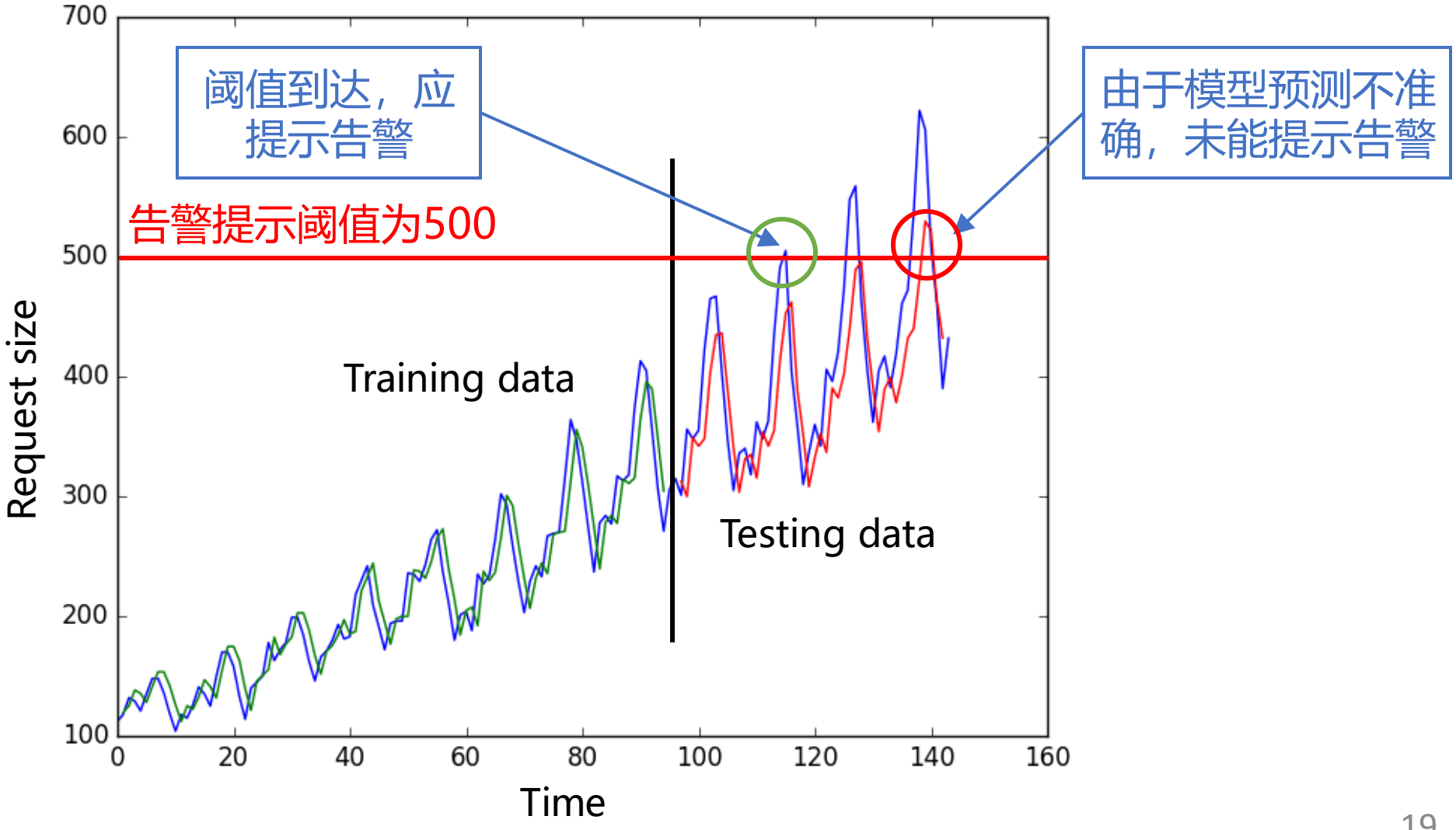
利用前三个数据点预测下一个数据点

滑动窗口





预测结果



Code: https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/

云系统智能运维

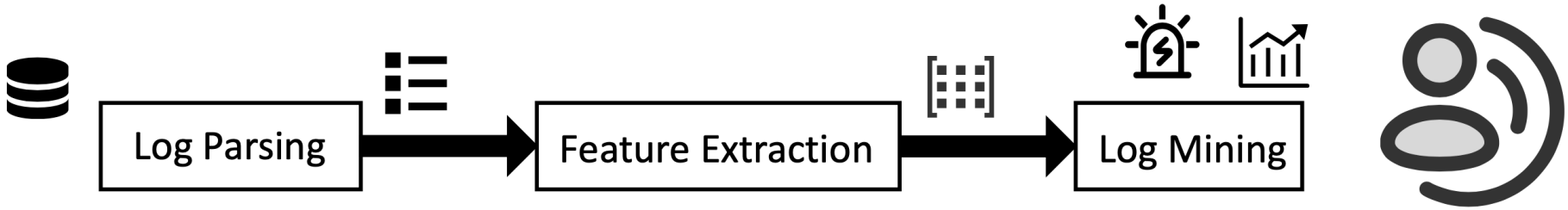
- 智能运维的概念
- 基于系统指标的智能运维
- 基于系统日志的智能运维
- 基于知识的智能运维

自动日志异常检测


```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0014_m_001575_0/part-01575. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:57800 dest: /10.251.121.224:50010
081111 083422 24621 INFO dfs.DataNode$DataXceiver: writeBlock blk_5214640714119373081 received
exception java.io.IOException: Could not read from stream
```

Automatically detected anomaly

日志自动化分析流程



日志解析　　特征提取　　日志挖掘

日志解析


```java
public void setTemperature(Integer temperature) {
  // ...
  logger.debug("Temperature set to {}. Old temperature was {}.", t, oldT);
  if (temperature.intValue() > 50) {
    logger.info("Temperature has risen above 50 degrees.");
  }
}
```

⇓

```
0 [setTemperature] DEBUG Wombat - Temperature set to 61. Old temperature was 42.
0 [setTemperature] INFO Wombat - Temperature has risen above 50 degrees.
```

日志解析

- 日志事件/日志模版

01	Name=Request (GET:http://AAA:1000/BBBB/sitedata.html)	t_41bx0
02	Leaving Monitored Scope (EnsureListItemsData) Execution Time=52.9013	t_51xi4
03	HTTP request URL: /14/Emails/MrX(MrX@mail.com)/1c-48f0-b29.eml	t_23hl3
04	HTTP Request method: GET	t_41bx0
05	HTTP request URL: /55/RST/UVX/ADEG/Lists/Files/docXX.doc	t_01mu1
06	Overridden HTTP request method: GET	t_41bx0
07	HTTP request URL: http://AAA:1000/BBBB/sitedata.html	t_41bx0
08	Leaving Monitored Scope (Request (POST:http://AAA:100/BBBB/ sitedata.html)) Execution Time=334.319268903038	t_41bx0 (Task_ID)

E1	Name=Request (*)
E2	Leaving Monitored Scope (*) Execution Time = *
E3	HTTP Request method: *
E4	HTTP request URL: *
E5	Overridden HTTP request method: *

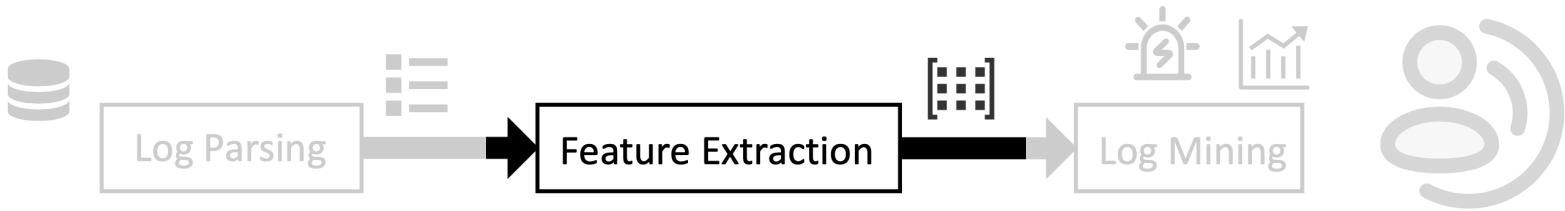
Log Parsing

E1, E2, E4, E3, E4, E5, E4, E2

为什么需要做日志解析?

- 获取系统主要事件，避免干扰信息
- 适应下游算法的特征
- 有些文章也认为变量有用，可以保留

特征提取



- 与具体的日志分析任务相关
  - ✓本例：判断系统的某次任务执行是否发生异常

特征提取

- 日志序列标签

  ✓Job ID, Process ID, etc.

2008-11-11 03:40:58 BLOCK* NameSystem.allocateBlock: /user/root/randtxt4 blk_904791815
2008-11-11 03:40:59 Receiving block blk_904791815 src: /master13 dest: /local22
2008-11-11 03:41:01 Receiving block blk_203948592 src: /master47 dest: /local93
2008-11-11 03:41:48 PacketResponder 0 for block blk_904791815 terminating
2008-11-11 03:41:48 Received block blk_904791815 of size 31864344 from /11.25.18.114
2008-11-11 03:41:48 PacketResponder 1 for block blk_203948592 terminating
2008-11-11 03:41:48 Received block blk_203948592 of size 47394022 from /10.251.43.210
2008-11-11 03:41:48 BLOCK* NameSystem.addStoredBlock added to blk_904791815 size 67108864
2008-11-11 03:41:48 BLOCK* NameSystem.addStoredBlock added to blk_203948592 size 47394022
2008-11-11 08:30:54 Verification succeeded for blk_904791815

特征提取

- Job 1 打印了如下日志事件

  E1, E2, E4, E3, E4, E5, E4, E2　　运行结果：执行正常

- Job 2 打印了如下日志事件

  E1, E2, E4, E3, E4, E5, E6, E2　　运行结果：执行异常

  E1, E2, E4, E3, E4, E4, E4, E4

  E1, E2, E4, E3

特征提取

- 特征向量与标签

	E1, E2, E3, E4, E5, E6
E1, E2, E4, E3, E4, E5, E4, E2	[1, 2, 1, 3, 1, 0], 0
E1, E2, E4, E3, E4, E5, E6, E2	[1, 2, 1, 2, 1, 1], 1
E1, E2, E4, E3, E4, E4, E4, E4	[1, 1, 1, 5, 0, 0], 1
E1, E2, E4, E3	[1, 1, 1, 1, 0, 0], 1

实际系统中负样本（运行正常）远比正样本（运行异常）多

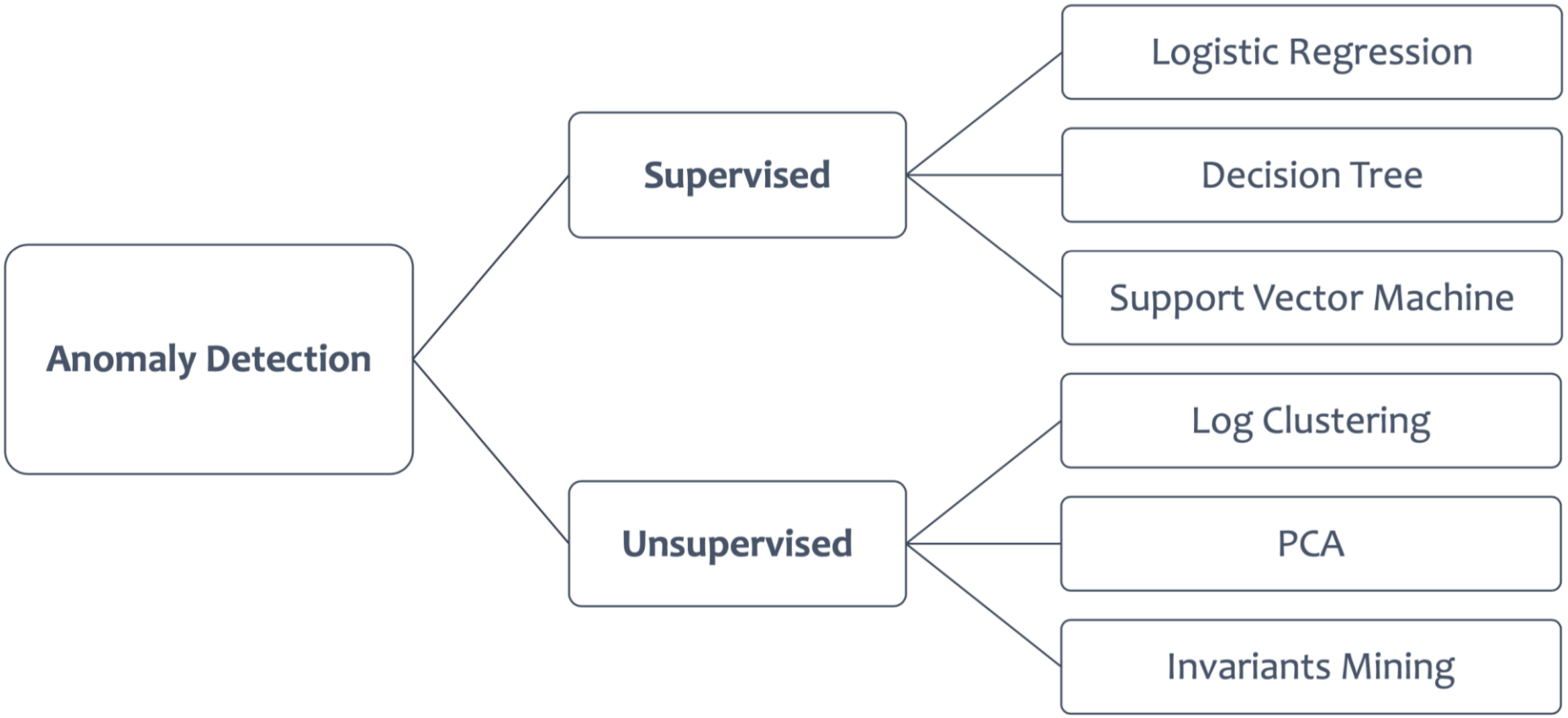
日志挖掘

- 判断系统的某次任务执行是否发生异常

Traditional machine learning

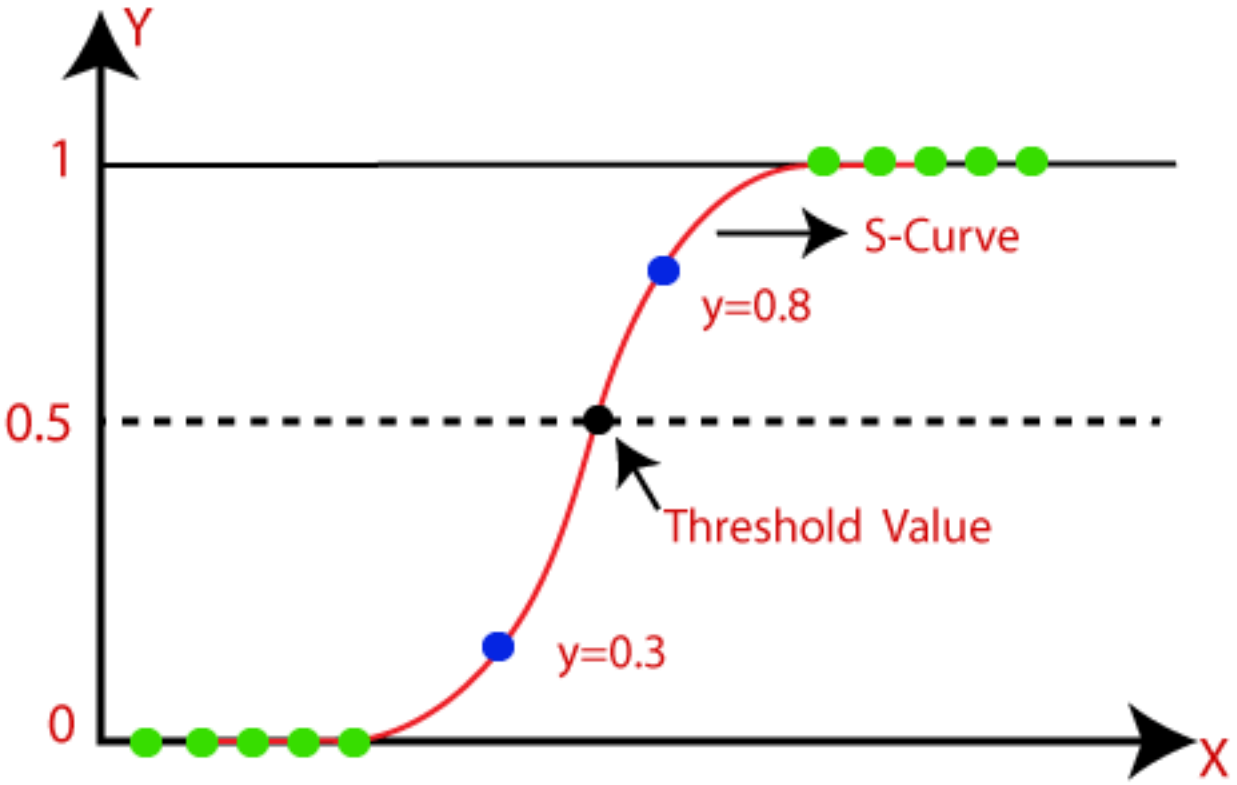
Deep learning models

传统机器学习方法


Anomaly Detection
Supervised
Unsupervised
Logistic Regression
Decision Tree
Support Vector Machine
Log Clustering
PCA
Invariants Mining

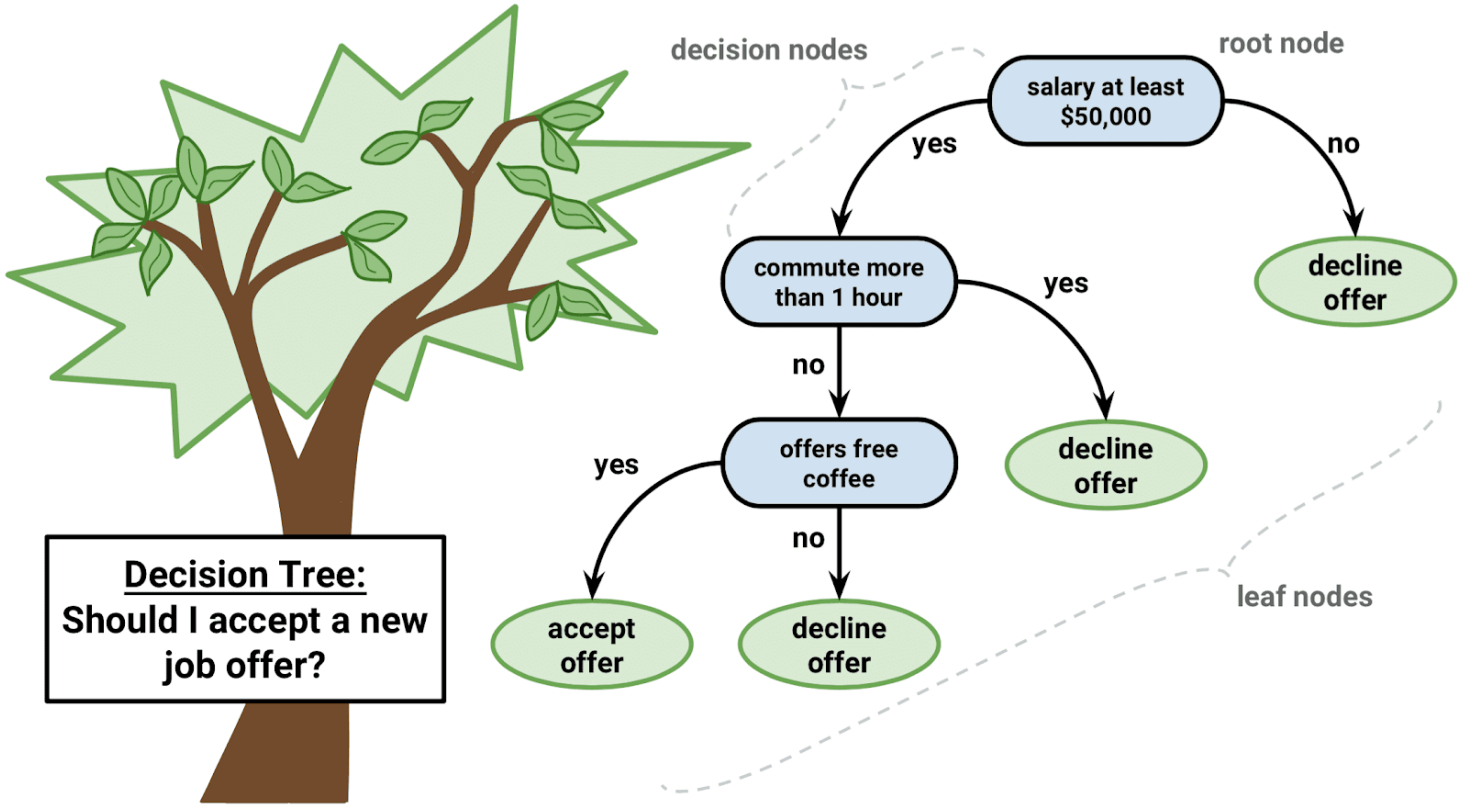
有监督学习算法

- Logistic Regression



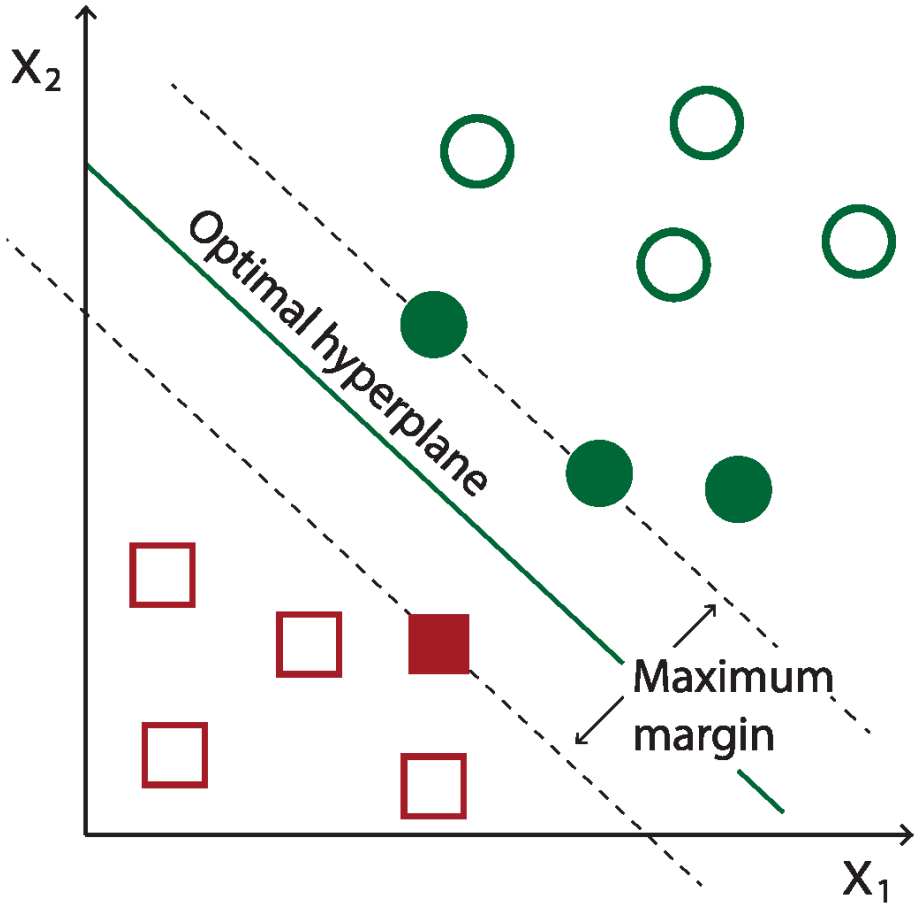
有监督学习算法

- Decision Tree



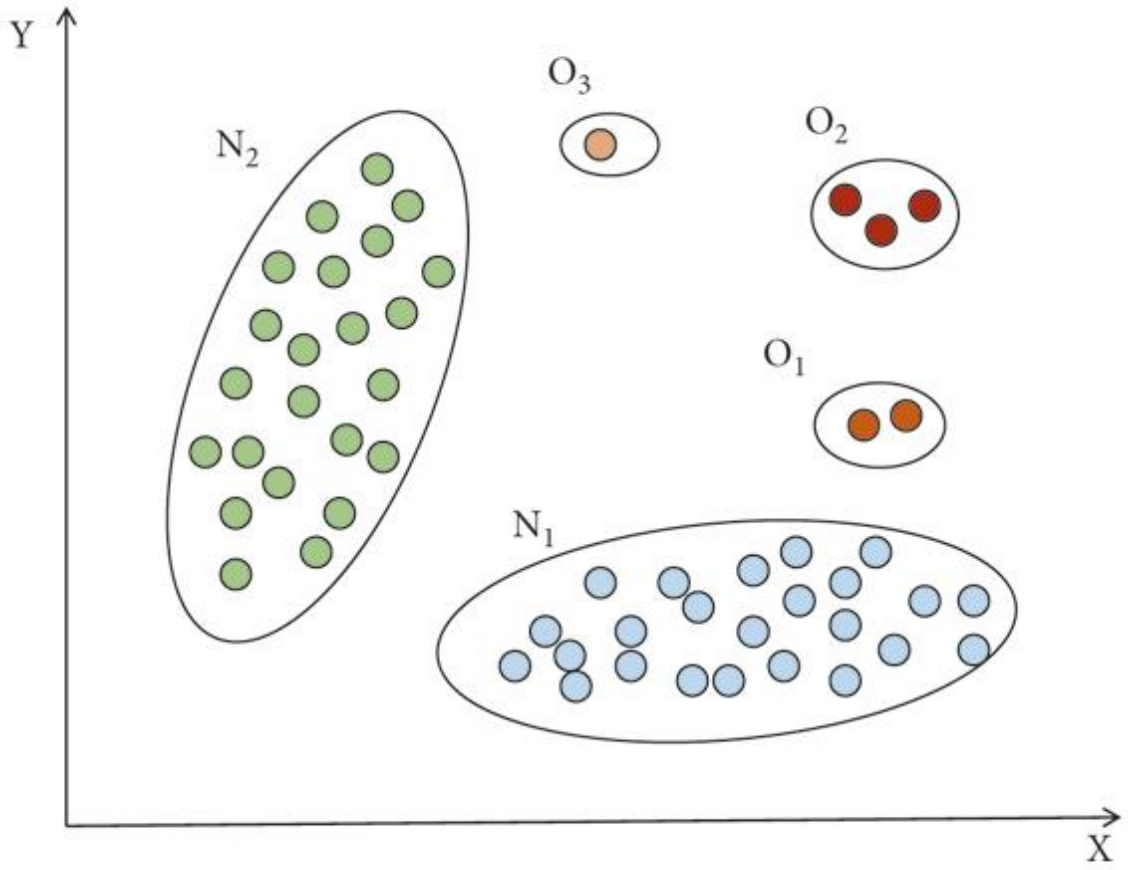
有监督学习算法

- Support Vector Machine



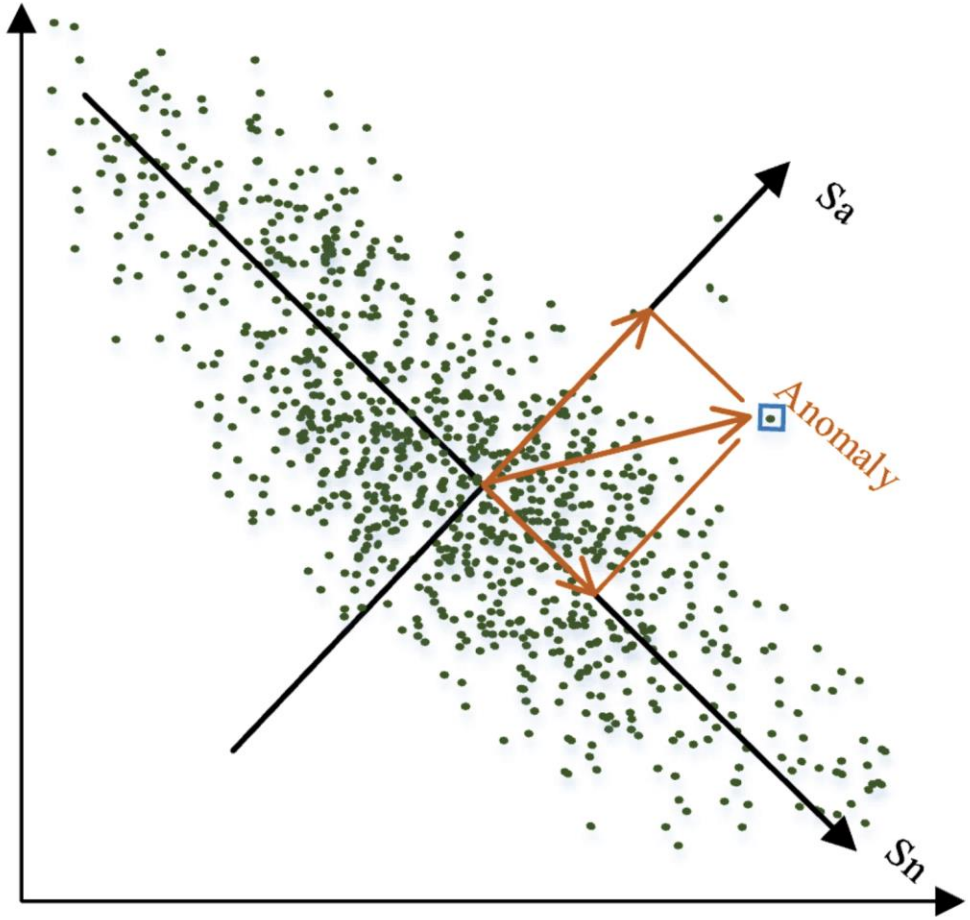
无监督学习算法

- Log Clustering



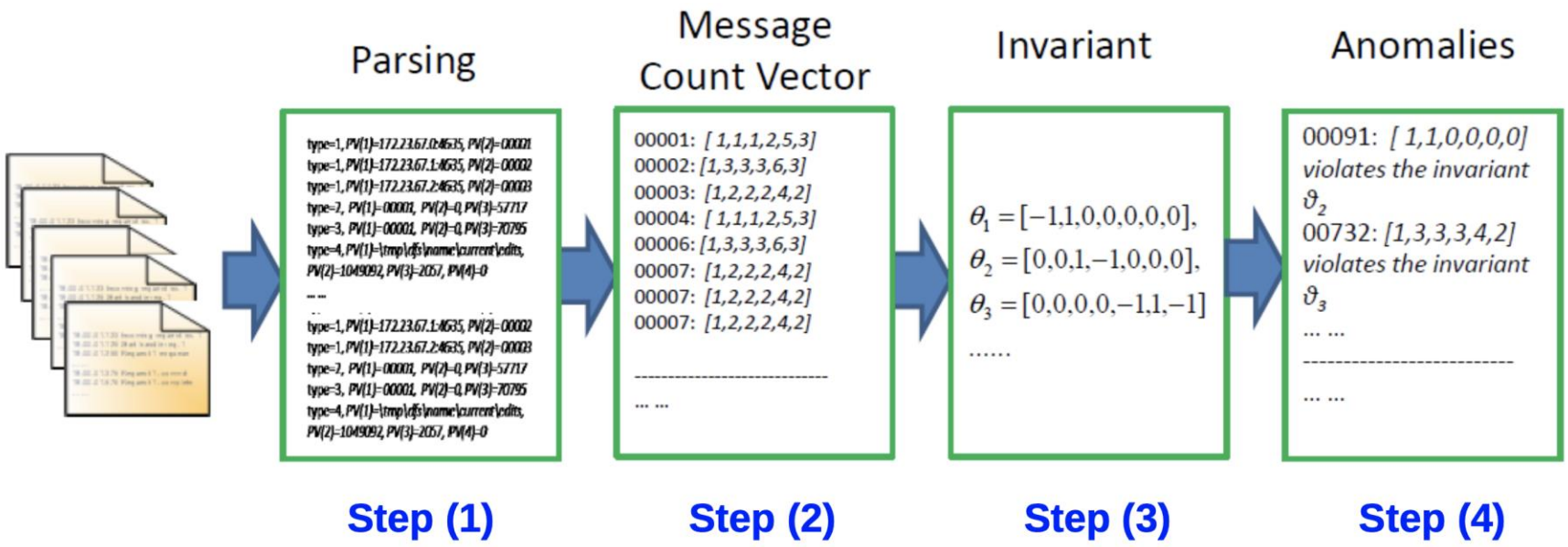
无监督学习算法

- PCA (Principal component analysis)

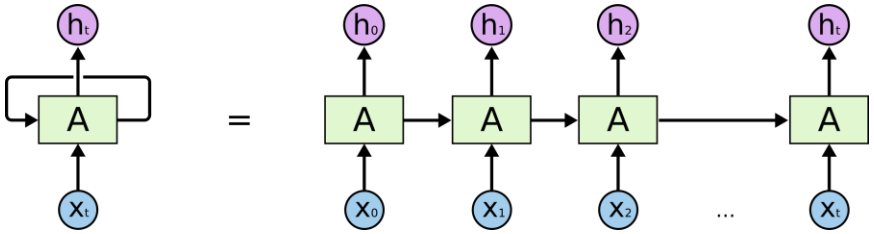


无监督学习算法

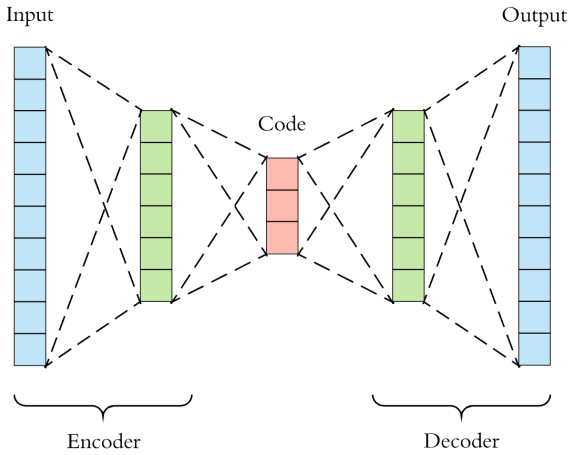
- Invariant mining



深度学习方法

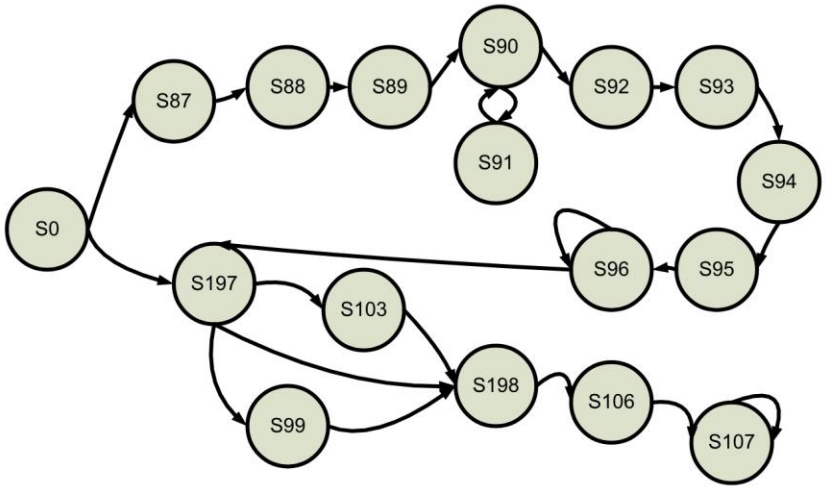


LSTM



Autoencoder

程序执行流程



Hadoop Mapreduce Task

State	Interpretation
S87~ S96	Initialization when a new job submitted
S197	Add a new map/reduce task
S103	Select remote data source
S99	Select local data source
S198	Task complete
S106	Job complete
S107	Clear task resource

The interpretations of states

系统打印出来的日志能反映
程序执行的流程

人工方式?

- 人工梳理程序流程图
  - ✓费时费力
  - ✓容易出错
  - ✓设计异常检测逻辑

- 代码静态分析?
  - ✓第三方服务代码无法获取

LSTM

Session F2: Insights from Log(in)s

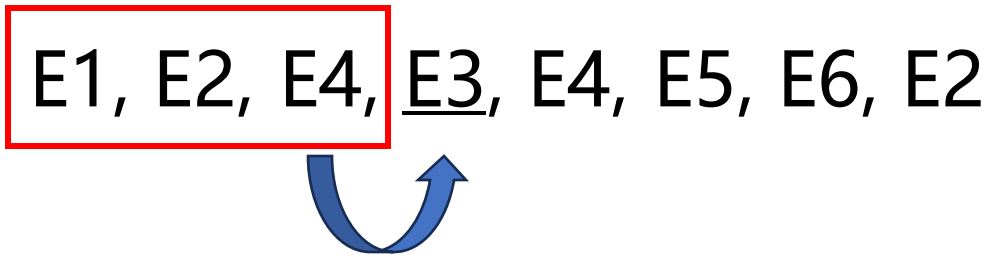
CCS'17, October 30-November 3, 2017, Dallas, TX, USA

DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning

Min Du, Feifei Li, Guineng Zheng, Vivek Srikumar
School of Computing, University of Utah
{mind, lifeifei, guineng, svivek}@cs.utah.edu

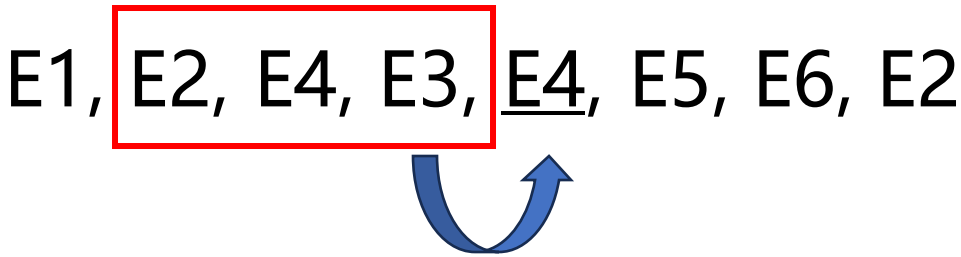
特征提取

- 日志事件序列
  - Window size = 3（超参数）



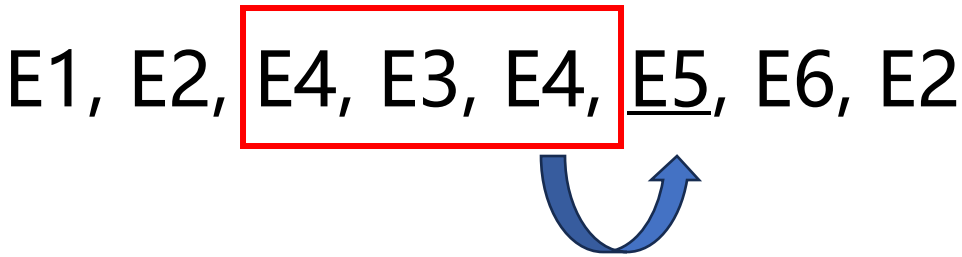
特征提取

- 日志事件序列
  - Window size = 3（超参数）



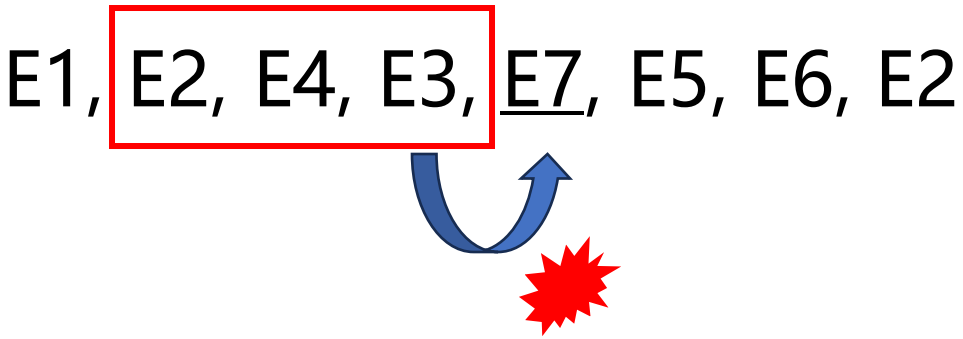
特征提取

- 日志事件序列
  - Window size = 3（超参数）



异常检测逻辑

- 用系统正常运行时产生的日志训练模型
- 模型自动学习程序的执行流程
- 当错误日志事件出现/顺序偏离正常流程，提示异常



Autoencoder

KICS
The Korean Institute of Communications and Information Sciences

Check for updates

Available online at www.sciencedirect.com
ScienceDirect

ICT Express 6 (2020) 229–237

www.elsevier.com/locate/icte

Unsupervised log message anomaly detection

Amir Farzad*, T. Aaron Gulliver

Department of Electrical and Computer Engineering, University of Victoria, PO Box 1700, STN CSC, Victoria, BC, V8W 2Y2, Canada

Received 17 February 2020; received in revised form 11 June 2020; accepted 25 June 2020
Available online 2 July 2020

特征提取

E1, E2, E4, E3, E4, E5, E6, E2

E1, E2, E4

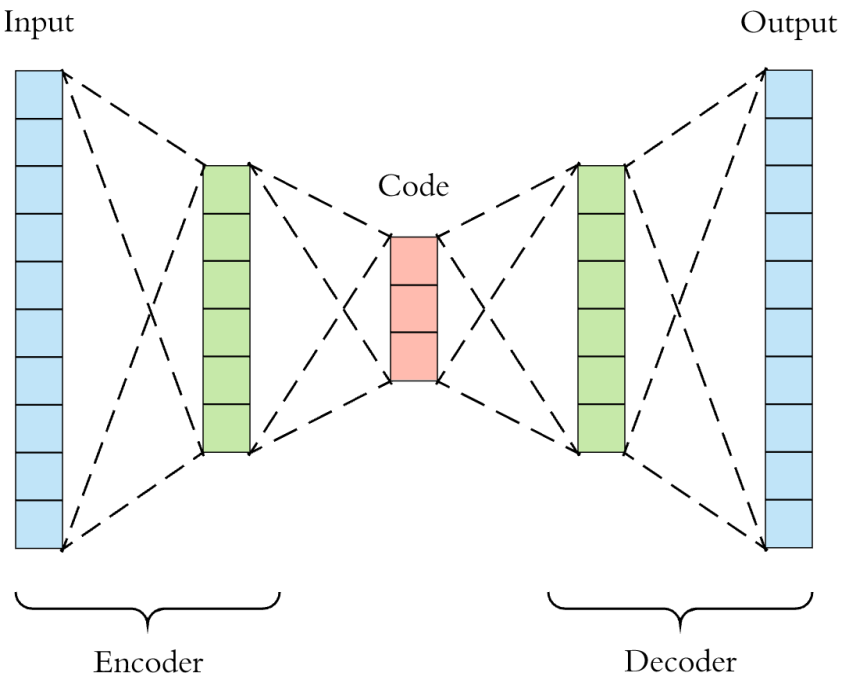
E2, E4, E3

E4, E3, E4

……

E1, E2, E4

输入



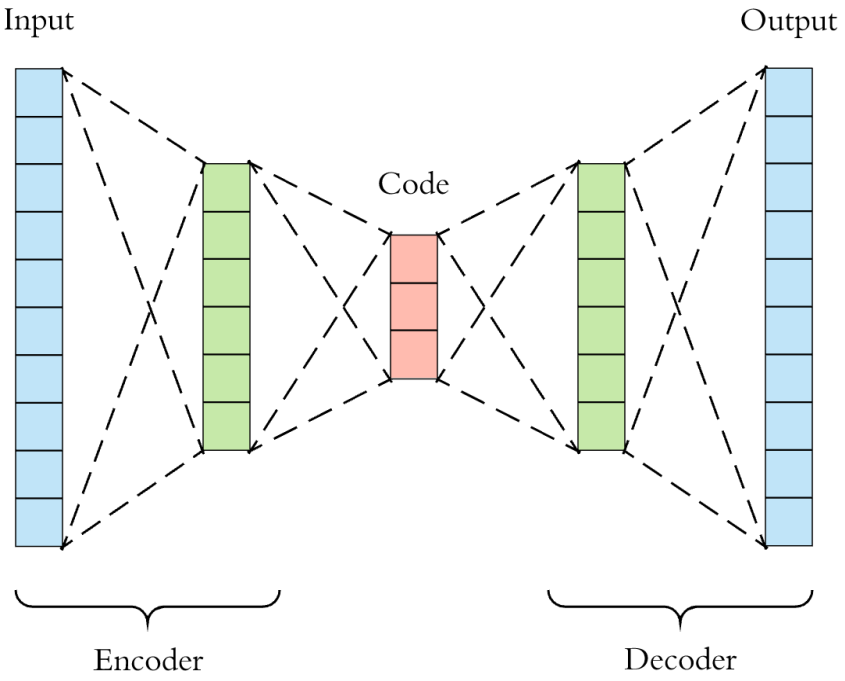
E1, E2, E4

恢复

异常检测

E1, E2, E6

输入



E1, E2, E5

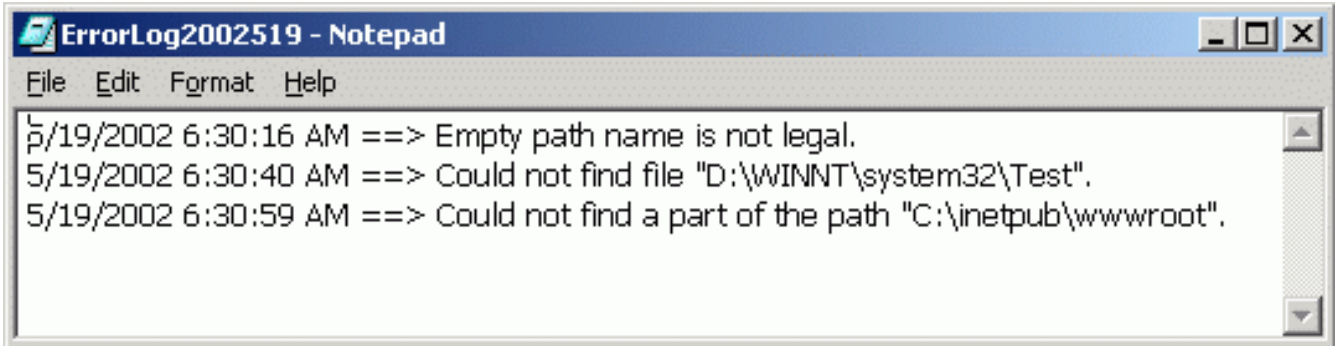
异常!

云系统智能运维

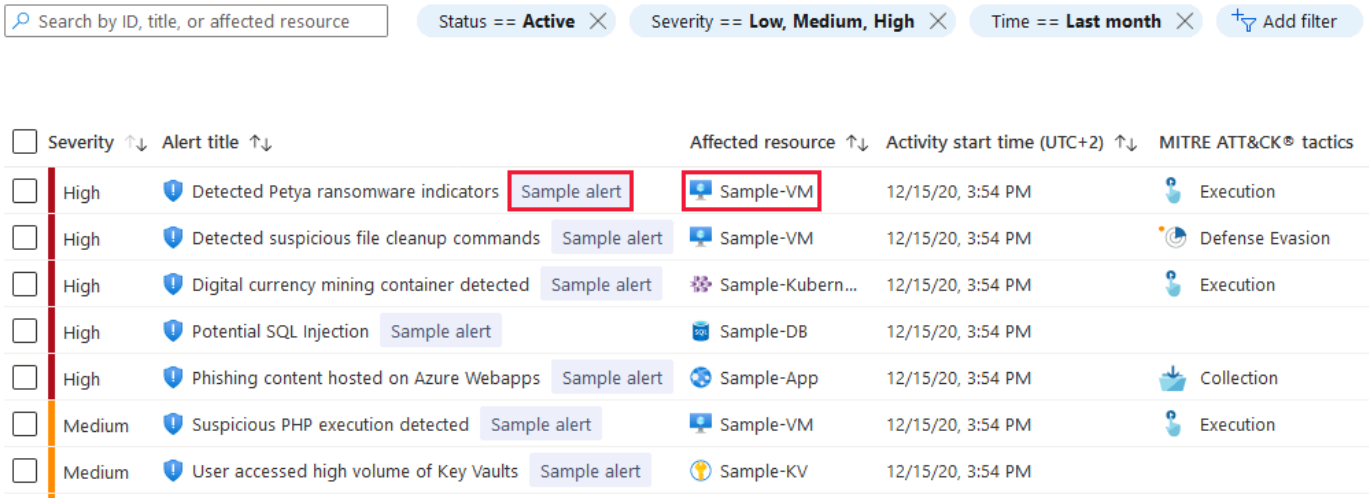
- 智能运维的概念
- 基于系统指标的智能运维
- 基于系统日志的智能运维
- 基于知识的智能运维

语义信息

- 日志


ErrorLog2002519 - Notepad
File Edit Format Help
5/19/2002 6:30:16 AM ==> Empty path name is not legal.
5/19/2002 6:30:40 AM ==> Could not find file "D:\WINNT\system32\Test".
5/19/2002 6:30:59 AM ==> Could not find a part of the path "C:\inetpub\wwwroot".

- 告警

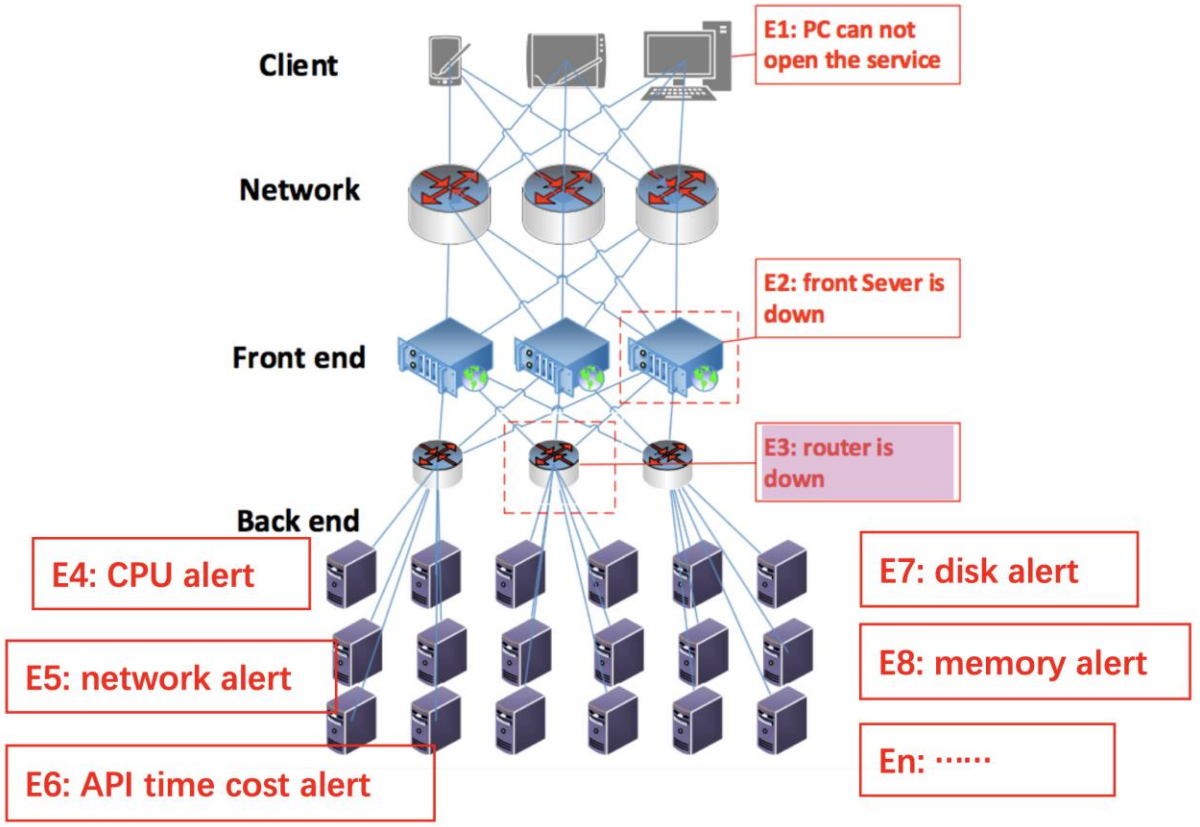

Search by ID, title, or affected resource
Status == Active
Severity == Low, Medium, High
Time == Last month
Add filter
Severity
Alert title
Affected resource
Activity start time (UTC+2)
MITRE ATT&CK® tactics
High Detected Petya ransomware indicators Sample alert Sample-VM 12/15/20, 3:54 PM Execution
High Detected suspicious file cleanup commands Sample alert Sample-VM 12/15/20, 3:54 PM Defense Evasion
High Digital currency mining container detected Sample alert Sample-Kubern... 12/15/20, 3:54 PM Execution
High Potential SQL Injection Sample alert Sample-DB 12/15/20, 3:54 PM
High Phishing content hosted on Azure Webapps Sample alert Sample-App 12/15/20, 3:54 PM Collection
Medium Suspicious PHP execution detected Sample alert Sample-VM 12/15/20, 3:54 PM Execution
Medium User accessed high volume of Key Vaults Sample alert Sample-KV 12/15/20, 3:54 PM

语义信息

- 算法能否像人一样理解文本监控数据的含义?
  - ✓系统故障主题挖掘
  - ✓工单及故障排除指南

故障传播

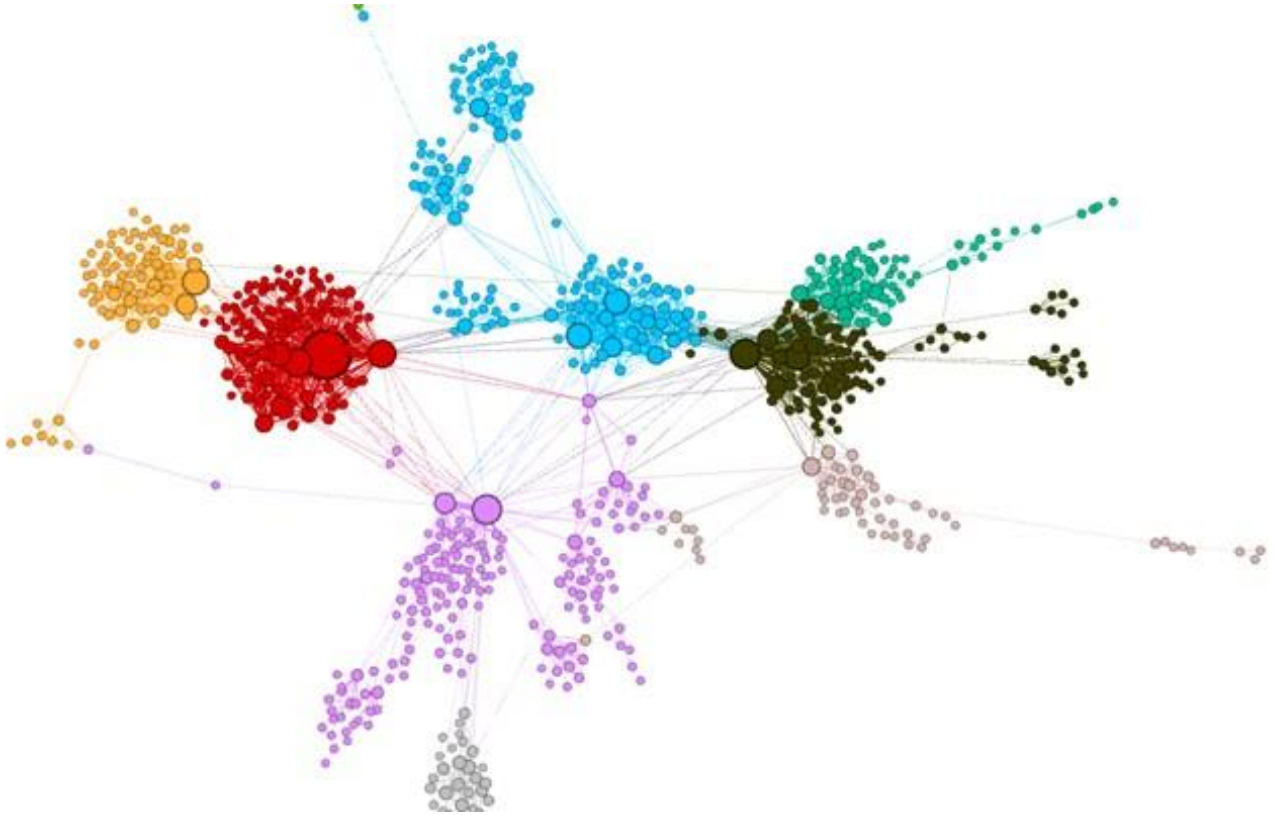
- 当系统出现故障时，通常会出现各种类型的问题



系统故障主题挖掘

- 提取整个系统有哪些类型的问题，类似文章的多个主题
  - ✓帮助运维人员了解故障影响范围
  - ✓帮助运维人员了解故障类型
  - ✓帮助运维人员进行根因定位

词向量（word embedding）



- 可用系统运维领域的词做fine tuning

日志/告警语义向量化

1. 对日志/告警中的每个词作word embedding

告警事件：The number of Oracle sessions is high.

$$L = [\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_N}], \text{where } \boldsymbol{v_i} \in \mathbb{R}^d, i \in [1, N]$$

日志/告警语义向量化

2. 将告警事件表征为一个向量，强调重要的词

TF-IDF (Term Frequency-Inverse Document Frequency)

是一种常用于信息检索和文本挖掘的统计方法，反映一个词对于一个文档集或一个语料库中的一份文件的重要性

TF-IDF

- TF-IDF由两部分组成：词频 (TF) 和逆文档频率 (IDF)
  - ✓比如单词 Block 在一个告警事件里出现多次，说明很重要

假设 Block 出现 2 次，告警事件总字数为 10，则 TF = 2/10 = 0.2

  - ✓但是如果 Block 在多个告警事件里出现，那么其辨识度较低

IDF = log(告警总数/(包含 Block 的告警 + 1))
假如有 100 个告警，20 都包含 Block，则
IDF = log(100/21) = 1.56

最终 Block 的 TF-IDF 为 w= 0.2 * 1.56 = 0.312

日志/告警语义向量化

告警事件：The number of Oracle sessions is high.

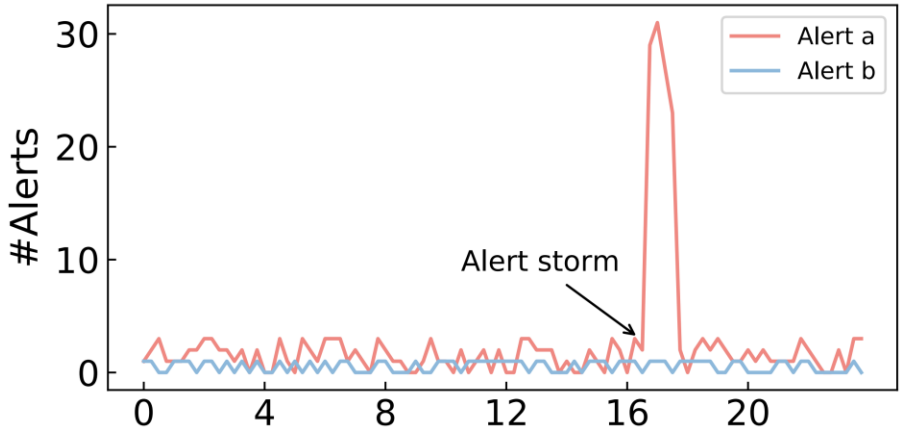
$$L = [\boldsymbol{v_1}, \boldsymbol{v_2}, ..., \boldsymbol{v_N}], \text{where } \boldsymbol{v_i} \in \mathbb{R}^d, i \in [1, N]$$

$$W = [w_1, w_2, \cdots, w_N], where\ w_i \in \mathbb{R}, i \in [1, N]$$

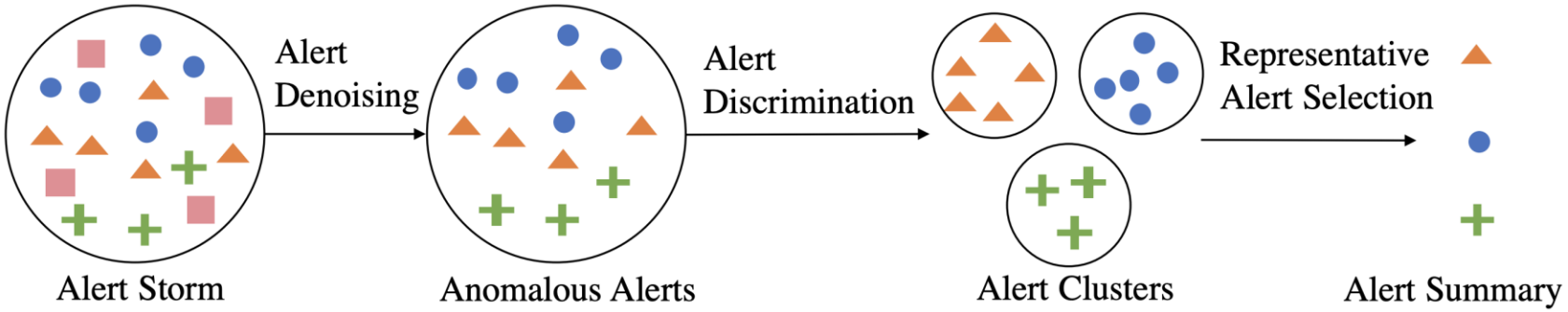
告警事件的语义向量：$$\boldsymbol{V} = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot \boldsymbol{v_i}$$

系统故障主题挖掘

- 当系统出现问题时，会有大量告警报出



- 挖掘大量告警中的主题，即问题类型



工单及故障排除指南

- 用来追踪用户的问题请求或者服务请求的记录，能够追踪**问题解决的全过程**
- 故障排除指南描述缓解问题的操作

Incident description:

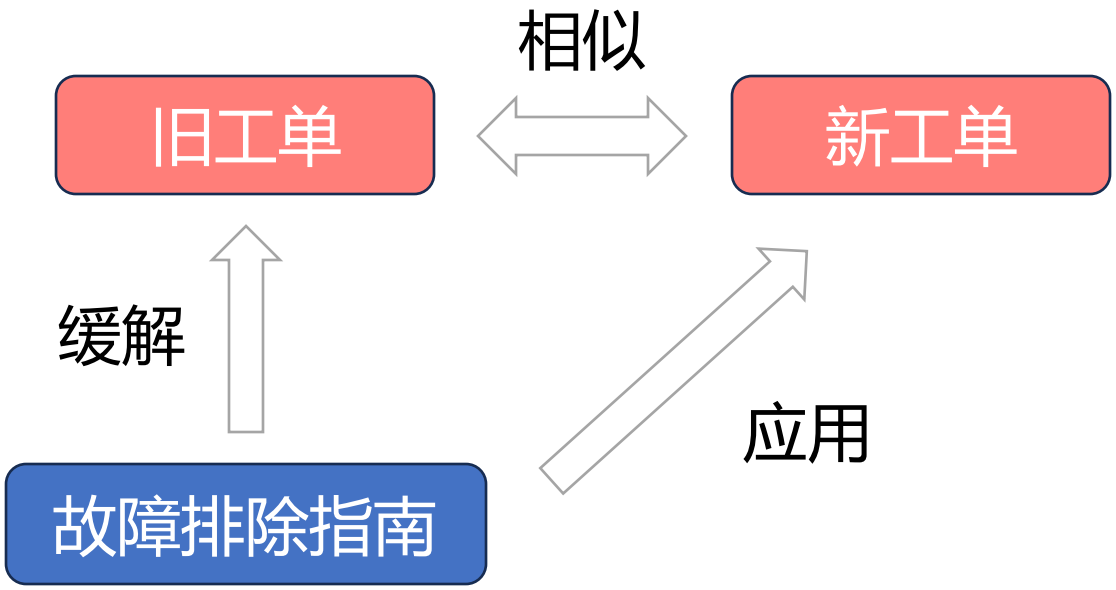
[AX] - Watch Dog RuleName DatabaseSpaceUsedRule 10PercentRemaining for Tenant 9a083aab-e8d6-459d-8407-xxxxx ...

Corresponding TSG:

Watchdog Rule Failure: Database Space Used
Symptoms:
Run the following query to get the current usage and free space details about the database.
*If the used percentage (USEDMB/TOTALMB*100) is greater than xx – it's bad, do the following*
...
SELECT [t].[name] AS [Table], [i].[name] AS [Index], [p].[partition_number] AS [Partition], [p].[data_compression_desc] AS [Compression] FROM [sys]
...

相似工单查找

- 为新的工单查找相似的历史工单有助于问题解决



是否存在相似工单?

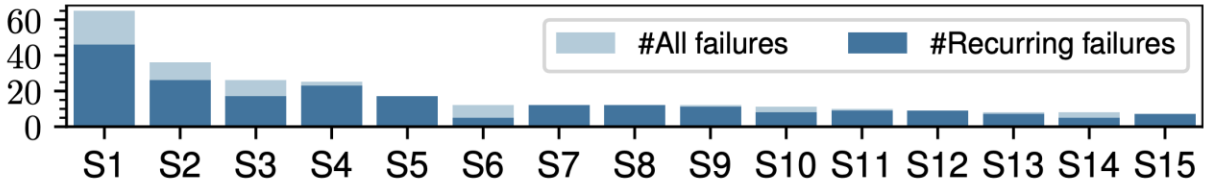
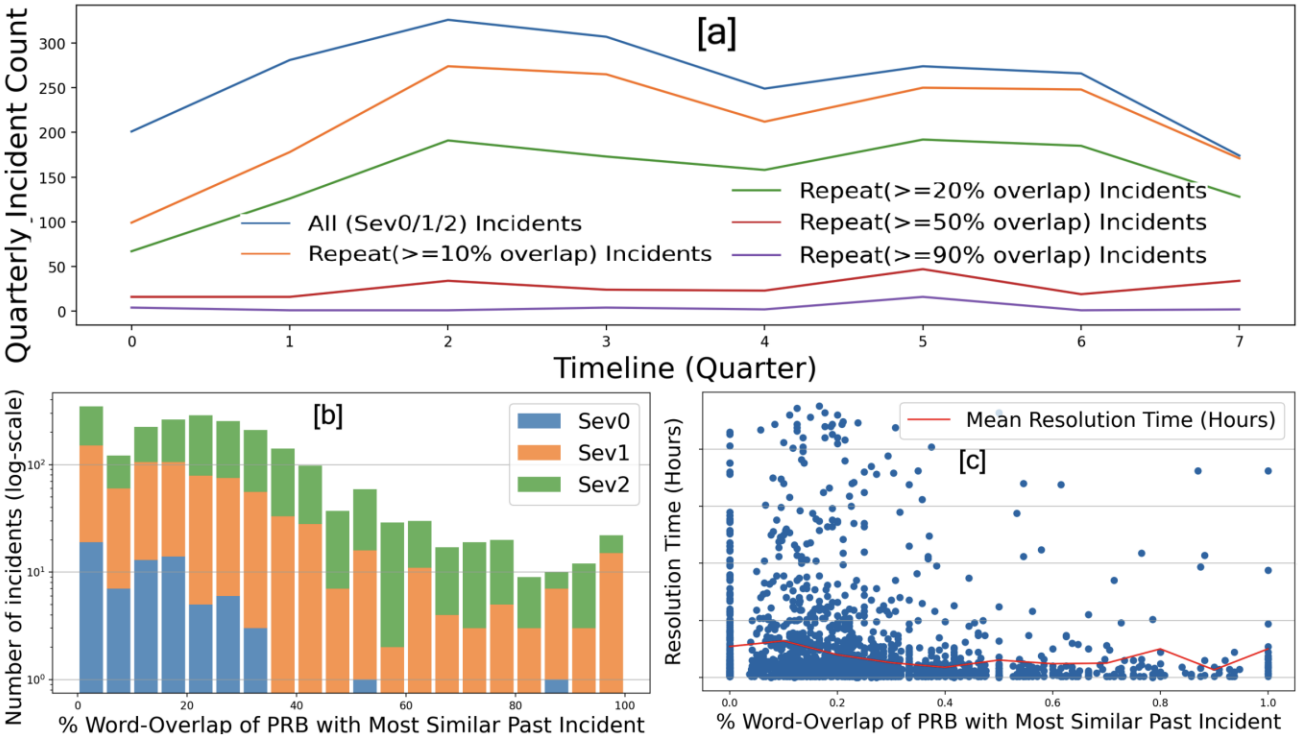


Figure 1: The number of recurring failures at *compA*.



[FSE22] Actionable and Interpretable Fault Localization for Recurring Failures in Online Service Systems
[ICSE-SEIP 22] Mining Root Cause Knowledge from Cloud Service Incident Investigations for AIOps

相似工单查找

① **Incident 1 Title:** Alert: email-api-batchevents-errors-production-allregions-exceeded

② **Description:** The Email Service was experiencing connectivity issues to their replica database in the West US Region. Due to this issue, System-Cloud customers globally were not receiving any type of System-Cloud notifications.

③ **Severity:** 2
Start time: 14:28
Service: SQL

① **Incident 2 Title:** No Success Signal in the last 60 minutes.

② **Description:** Calls to the API-Sub failed with a 5xx HTTP error. Approximately α_1 customers could not upgrade their subscriptions on URL-Cloud-Portal.

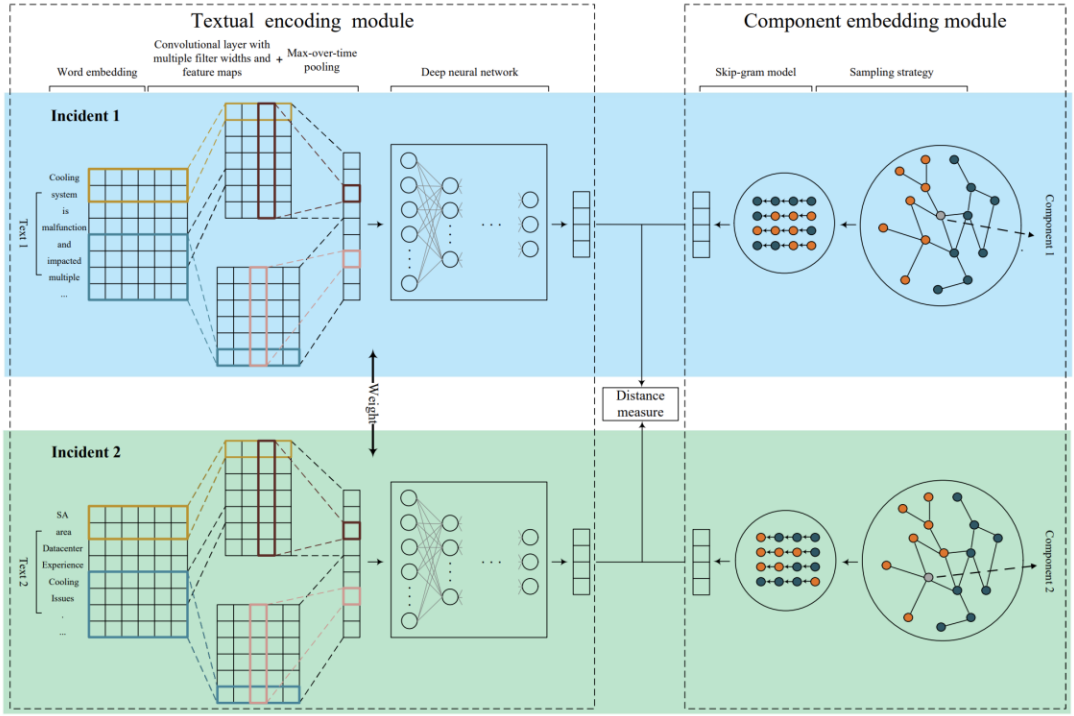
③ **Severity:** 2
Start time: 15:30
Service: Commercial

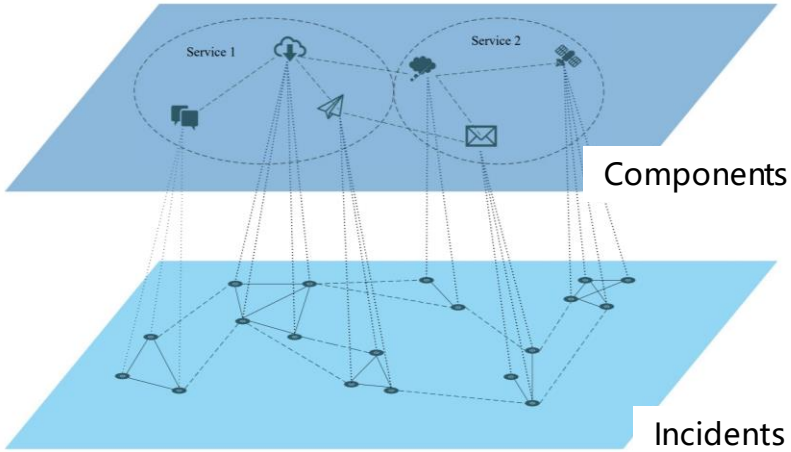
① 可利用前述的语义向量化方法

② 人工规则过滤，比如必须属于同一个服务

③ 更复杂的深度学习模型

工单匹配模型





[FSE20] Identifying linked incidents in large-scale online service systems

自动话程度更高的智能运维

- 一直以来，智能运维领域一直努力提升算法的自动化程度

  ✓知识图谱

  ✓大模型

让模型能真正理解系统内部的逻辑关系，让模型能推导出故障根因并提出修复手段

知识图谱方向

- 故障修复报告

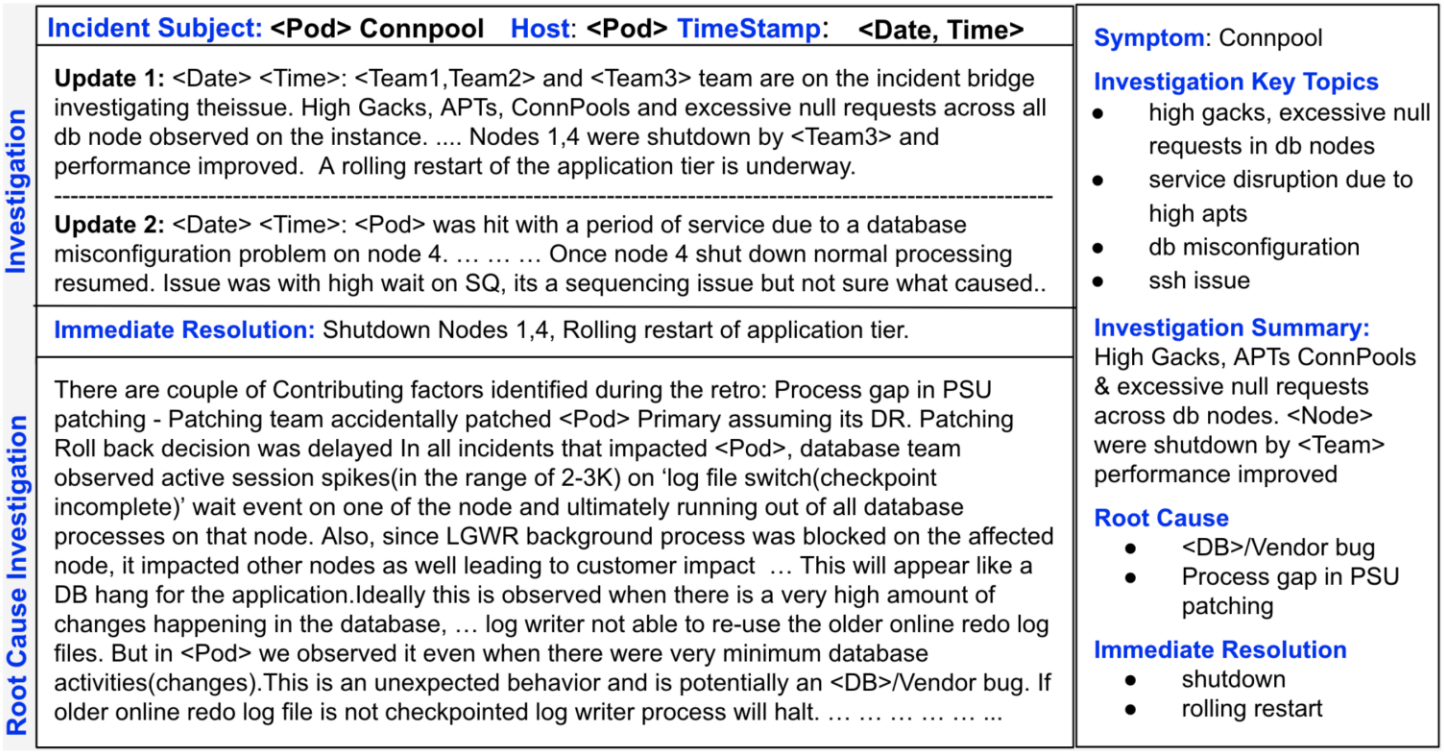


Figure 1: (left) Raw PRB document and (right) Structured form obtained through neural information extraction

[ICSE22] Mining Root Cause Knowledge from Cloud Service Incident Investigations for AIOps

知识图谱方向

关系

- has symptom
- has root cause
- has child

知识图谱问答实例

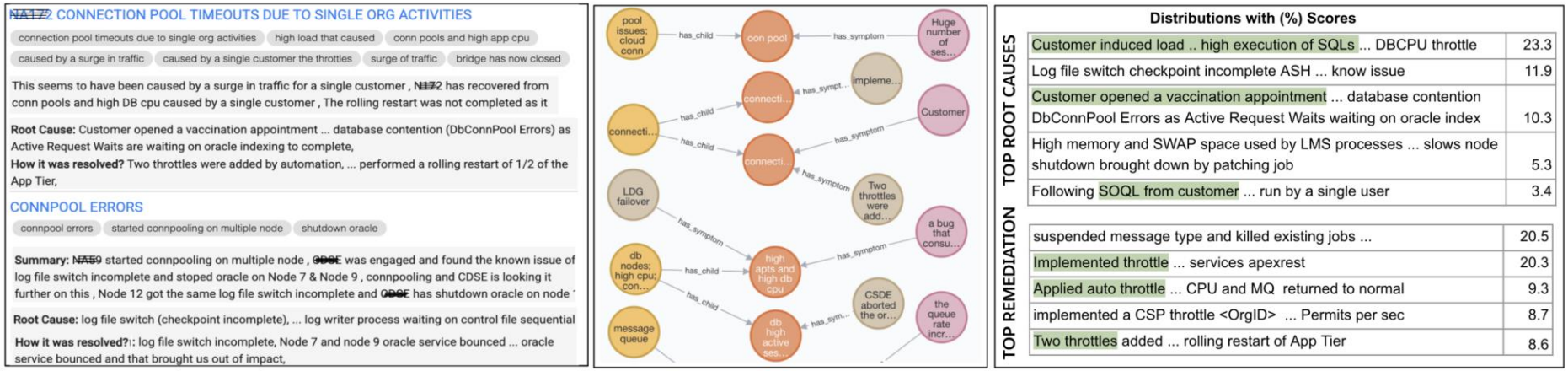


Figure 8: ICA output for Incident with Symptom "Connpool issues" (i) Incident Search Results (ii) Query Specific Causal Knowledge Subgraph iii) Distribution of top detected root causes and resolutions (Highlighted spans show match with true)

知识图谱面临的主要挑战

- 数据质量不高
- 知识获取与更新
- 云系统规模与故障的复杂性
- 知识表示与关联
- 算法本身推理能力有限

大模型方向

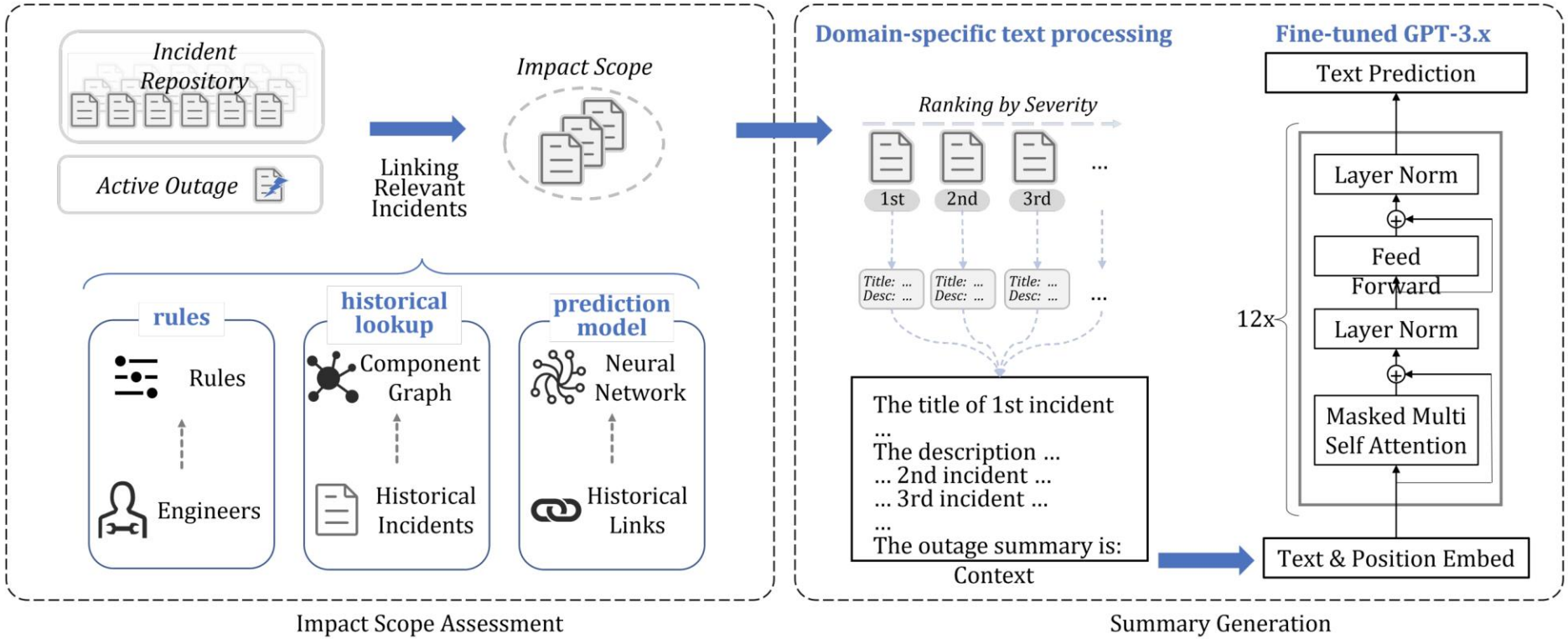
- 总结工单，加快问题理解
- 自动化推荐故障修复步骤

大模型总结工单

Outage Title: Outage for Email Service - Triage
Impact start time: 14:20
Outage declared time: 14:28
OCEs engage time: 14:29
Outage Summary: The Email Service experienced connectivity issues to their replica database in the West US Region. This affected customer email delivery for approximately α_3 internal company services. Due to this issue, System-Cloud customers were not receiving notifications including purchase, renewal, and monitor alert notifications. The Portal team reported that approximately α_1 customers were unable to upgrade their subscriptions on URL-Cloud-Portal.

划线部分为工单主要内容

工单总结模型



[arXiv23] Assess and Summarize: Improve Outage Understanding with Large Language Models

自动化故障修复

Incident description:

[AX] - Watch Dog RuleName DatabaseSpaceUsedRule 10PercentRemaining for Tenant 9a083aab-e8d6-459d-8407-xxxxx ...

Corresponding TSG:

Watchdog Rule Failure: Database Space Used

Symptoms:

Run the following query to get the current usage and free space details about the database.

*If the used percentage (USEDMB/TOTALMB*100) is greater than xx – it's bad, do the following*

...

SELECT [t].[name] AS [Table], [i].[name] AS [Index], [p].[partition_number] AS [Partition], [p].[data_compression_desc] AS [Compression] FROM [sys]

...

大模型能否理解故障及其修复手段?

大模型推荐根因

TABLE I: Effectiveness of fine-tuned GPT-3.x models at finding **root causes** of the incidents

Model	BLEU-4		ROUGE-L		METEOR		BERTScore		BLEURT		NUBIA	
	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5
RoBERTa	4.21	NA	12.83	NA	9.89	NA	85.38	NA	35.66	NA	33.94	NA
CodeBERT	3.38	NA	10.17	NA	6.58	NA	84.88	NA	33.19	NA	39.05	NA
Curie	3.40	6.29	9.04	15.44	7.21	13.65	84.90	86.36	32.62	40.08	33.52	49.76
Codex	3.44	6.25	8.98	15.51	7.33	13.82	84.85	86.33	32.50	40.11	33.64	49.77
Davinci	3.34	5.94	8.53	15.10	6.67	12.95	83.13	84.41	31.06	38.61	**35.28**	50.79
Davinci-002	**4.24**	**7.15**	**11.43**	**17.2**	**10.42**	**16.8**	**85.42**	**86.78**	**36.77**	**42.87**	32.3	**51.34**
%gain for Davinci-002	23.26	13.67	26.44	10.90	42.16	21.56	0.61	0.49	12.72	6.88	-8.45	1.08

大模型推荐修复手段

TABLE II: Effectiveness of fine-tuned GPT-3.x models at finding mitigation plans of the incidents

Model	BLEU-4		ROUGE-L		METEOR		BERTScore		BLEURT		NUBIA	
	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5	Top1	Top5
RoBERTa	4.44	NA	7.10	NA	4.52	NA	86.33	NA	26.80	NA	14.90	NA
CodeBERT	6.02	NA	4.40	NA	3.37	NA	86.83	NA	28.44	NA	27.89	NA
Curie	5.47	10.62	8.03	16.31	6.22	12.75	85.65	87.13	27.20	37.23	15.30	25.46
Codex	5.53	10.62	8.15	16.23	6.19	13.15	85.68	87.35	28.43	37.92	15.77	26.33
Davinci	5.54	10.66	8.10	15.96	6.08	12.49	85.72	87.19	27.15	37.00	15.71	25.61
Davinci-002	**6.76**	**11.66**	**10.22**	**18.14**	**8.23**	**15.13**	**86.17**	**87.65**	**30.19**	**38.96**	**17.58**	**28.81**
%gain for Davinci-002	22.02	9.38	25.40	11.22	32.32	15.06	0.52	0.34	6.19	2.74	11.48	9.42

大模型面临的主要挑战

- 云系统的故障理解与根因推荐需要基于整个系统架构以及组件之间的交互逻辑
- 知识（专家知识，系统知识）表示困难，不知道以何种形式输入模型
- 实时数据的获取与模型更新
- 数据隐私与安全问题

中山大學 软件工程学院

SUN YAT-SEN UNIVERSITY SCHOOL OF SOFTWARE ENGINEERING

谢谢

陈壮彬

软件工程学院

https://zbchern.github.io/sse316.html