



AI在软件测试领域应用探索

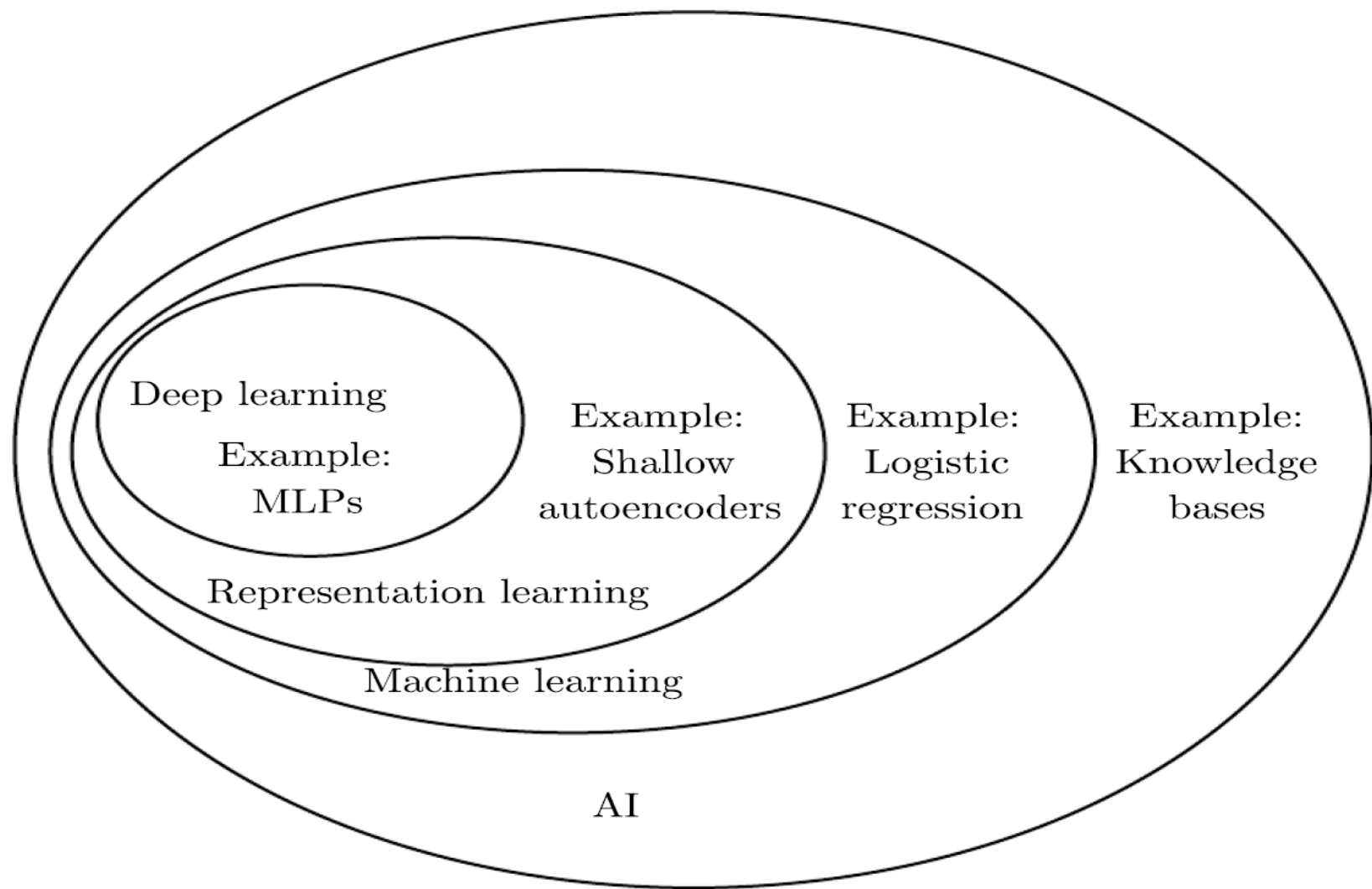
徐小明@华为DevCloud

LEADING NEW ICT

内容

- AI基本介绍
- AI在测中的应用
- 深度学习做日志分析案例分享

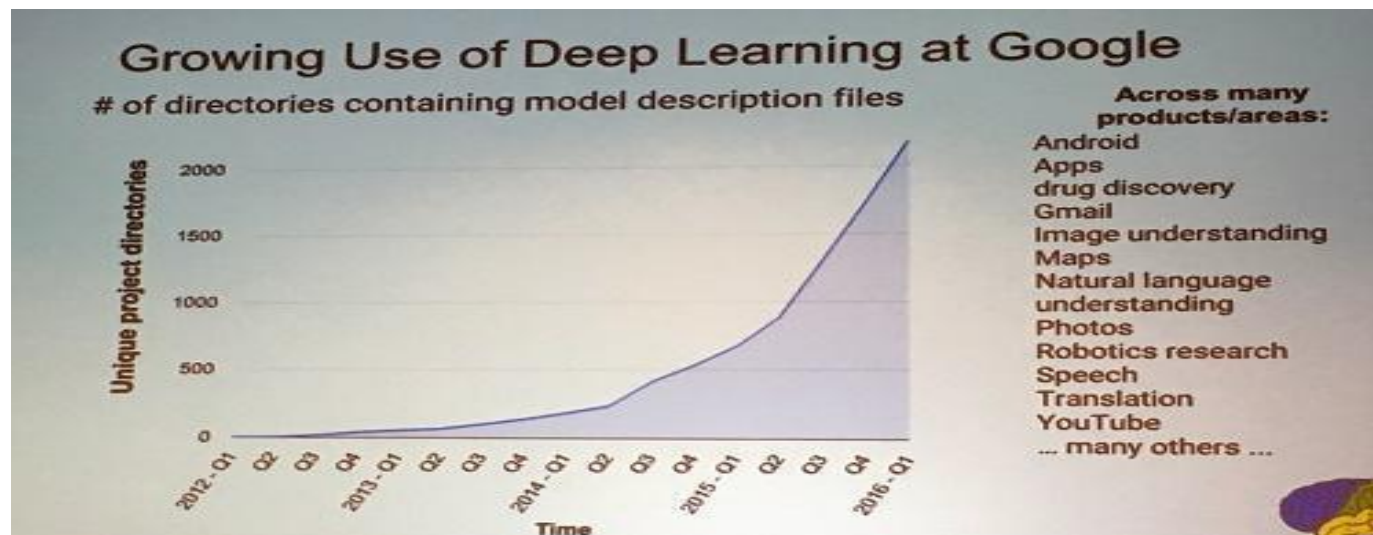
- 人工智能(AI)是对人的智能的模拟
 - 结果论
 - 一个拥有非常详尽的规则系统也可以是 AI
- 机器学习是达成 AI 的一种方法
- 深度学习是机器学习的一种新趋势
 - 从数据当中自动学习出各种规则
 - 训练一个够好的函数来解决特定的问题

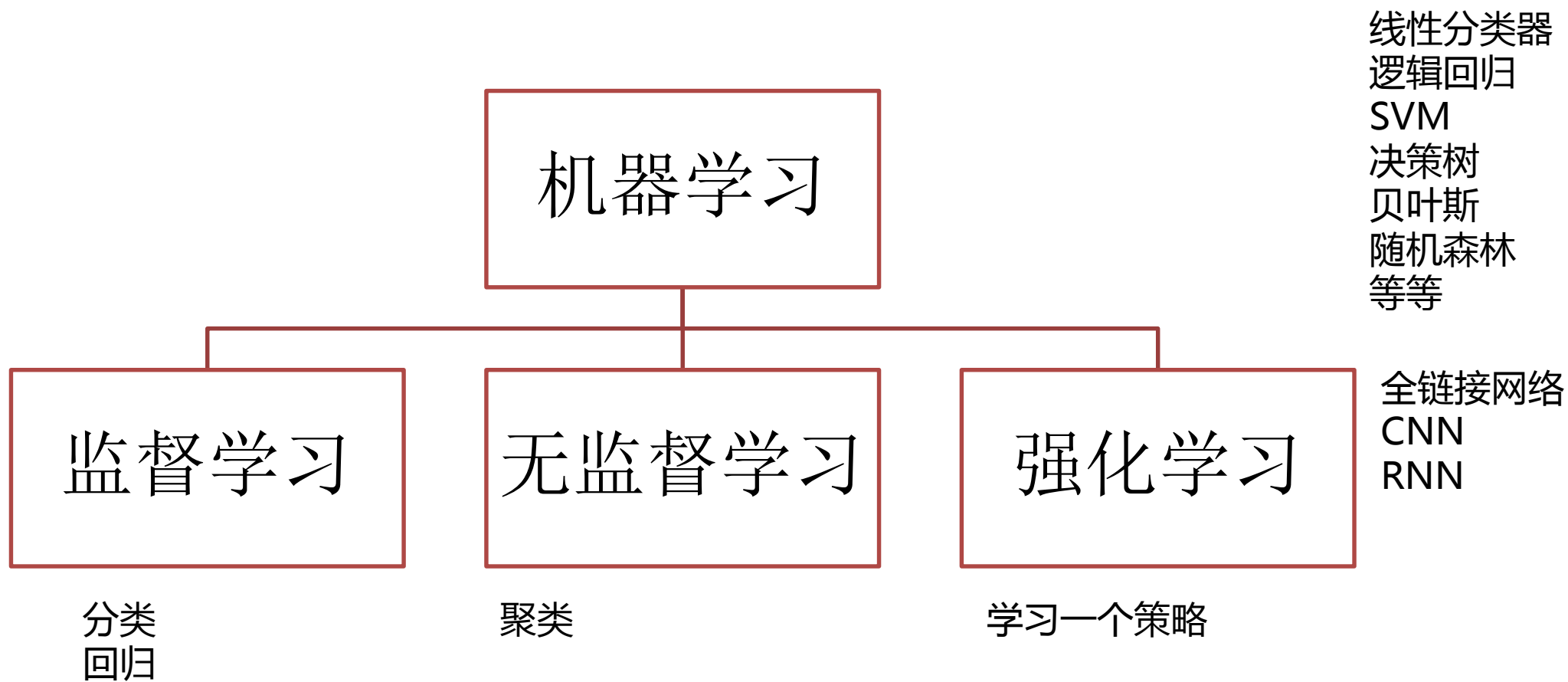


□ 人工智能研究中的核心挑战：

- 适合解决人们很容易实施但不容易表述的任务(proved to be solving the tasks that are easy for people to perform but hard for people to describe formally);
- 比如语音识别和图像识别等等;

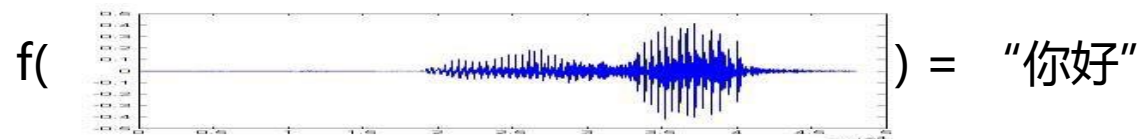
□ 深度学习迅速占领舞台





机器学习 \approx 寻找一个函数

□ 语音识别



□ 图像识别



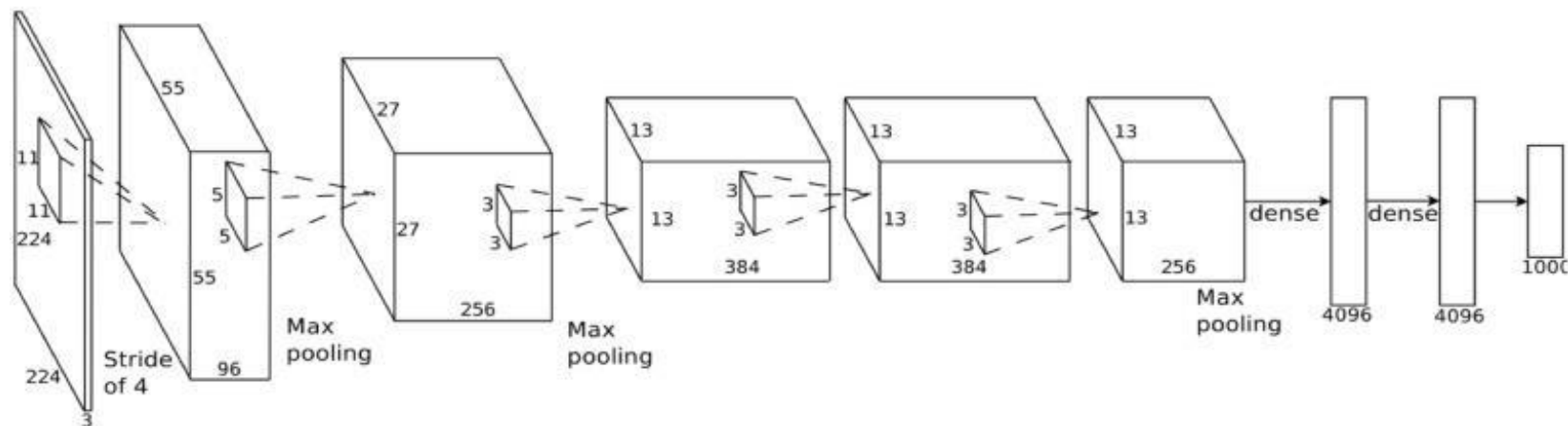
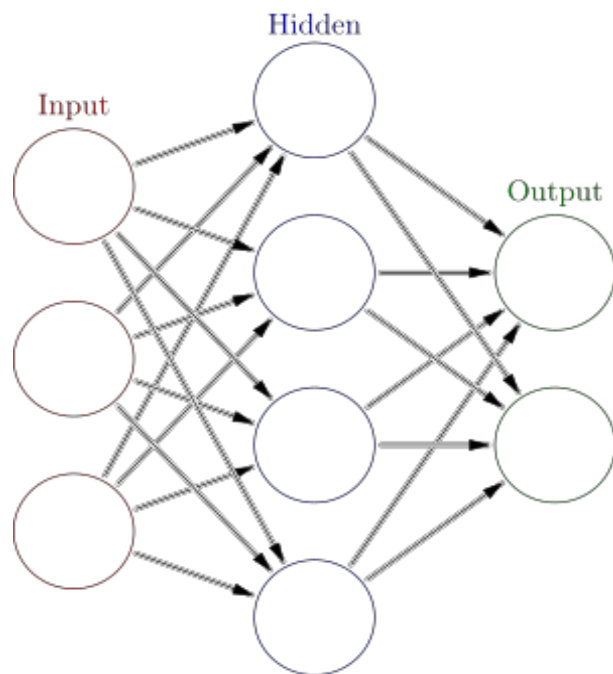
□ AlphaGo



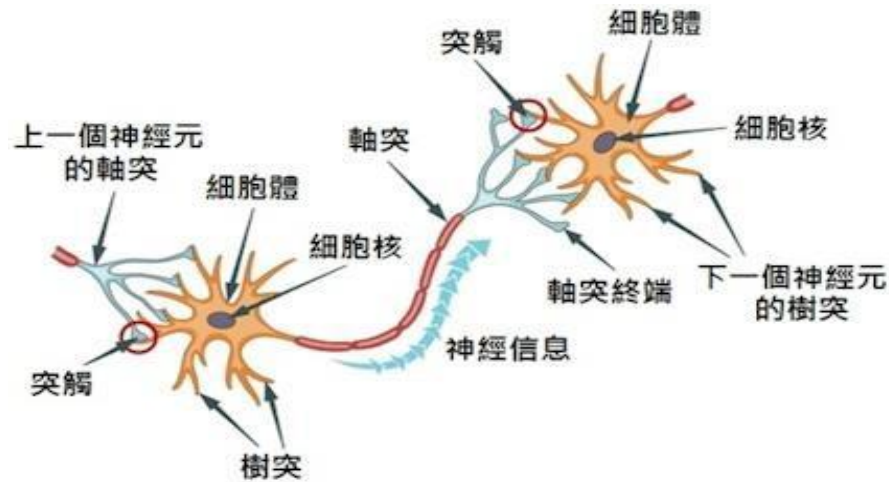
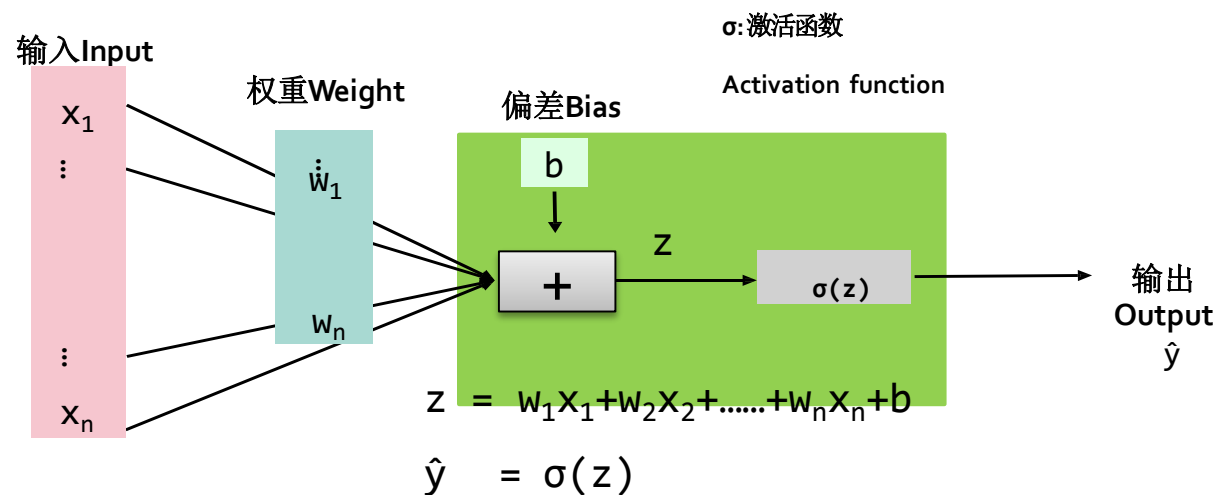
□ 对话系统

$f(\text{"Hi"}) = \text{"Hello"}$

- 人工神经网络 (Artificial neural network , ANN, 1943)
- 模拟人类神经传导机制的设计
 - 由许多层的神经元 互相连结而形成 神经网络
- 当 隐藏层层数够多(一般而言大于三层)就称为深度神经网络



神经元和激活函数



□ Sigmoid

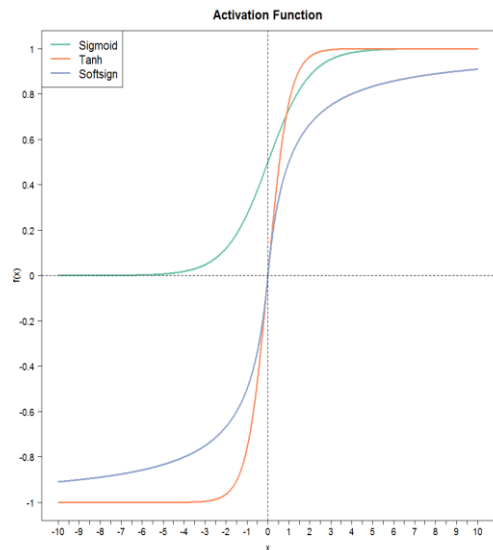
$$f(x) = \frac{1}{1+e^{-x}}$$

□ Tanh

$$f(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$$

□ Softsign

$$f(x) = \frac{x}{1+|x|}$$

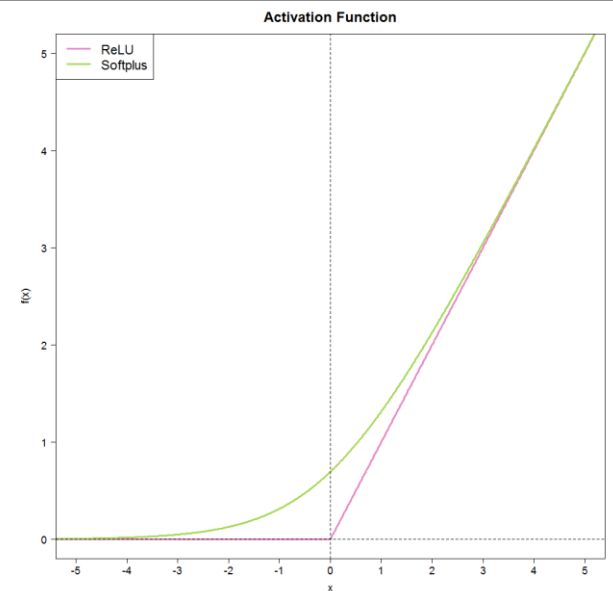


□ ReLU

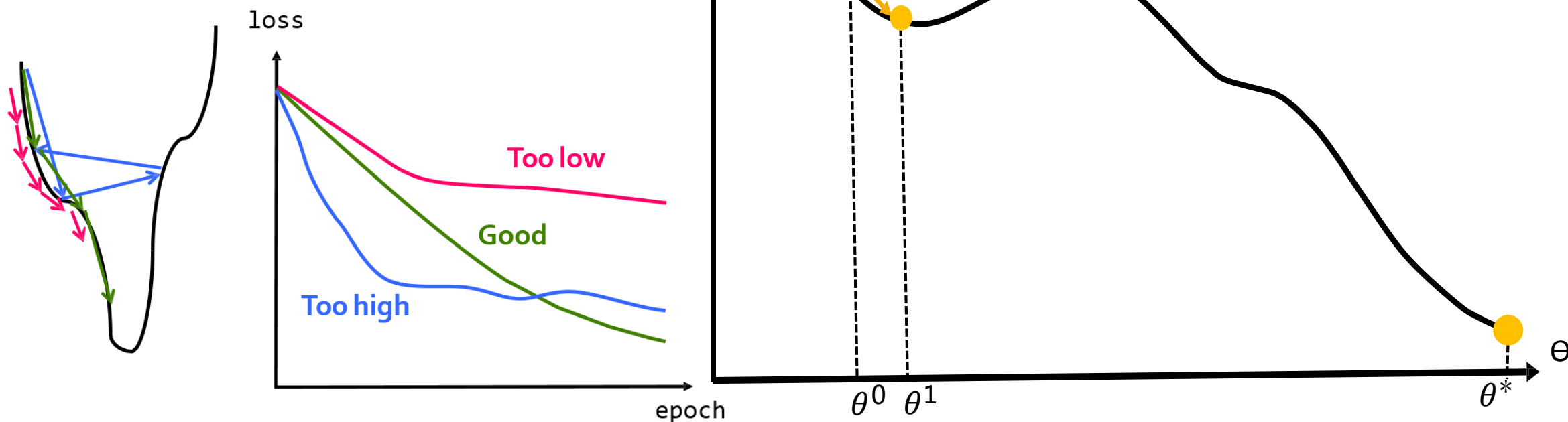
$$f(x) = \max(0, x)$$
$$\frac{df}{dx} = 1 \text{ if } x > 0, \\ 0 \text{ otherwise.}$$

□ Softplus

$$f(x) = \ln(1+e^x)$$
$$\frac{df}{dx} = \frac{e^x}{1+e^x}$$

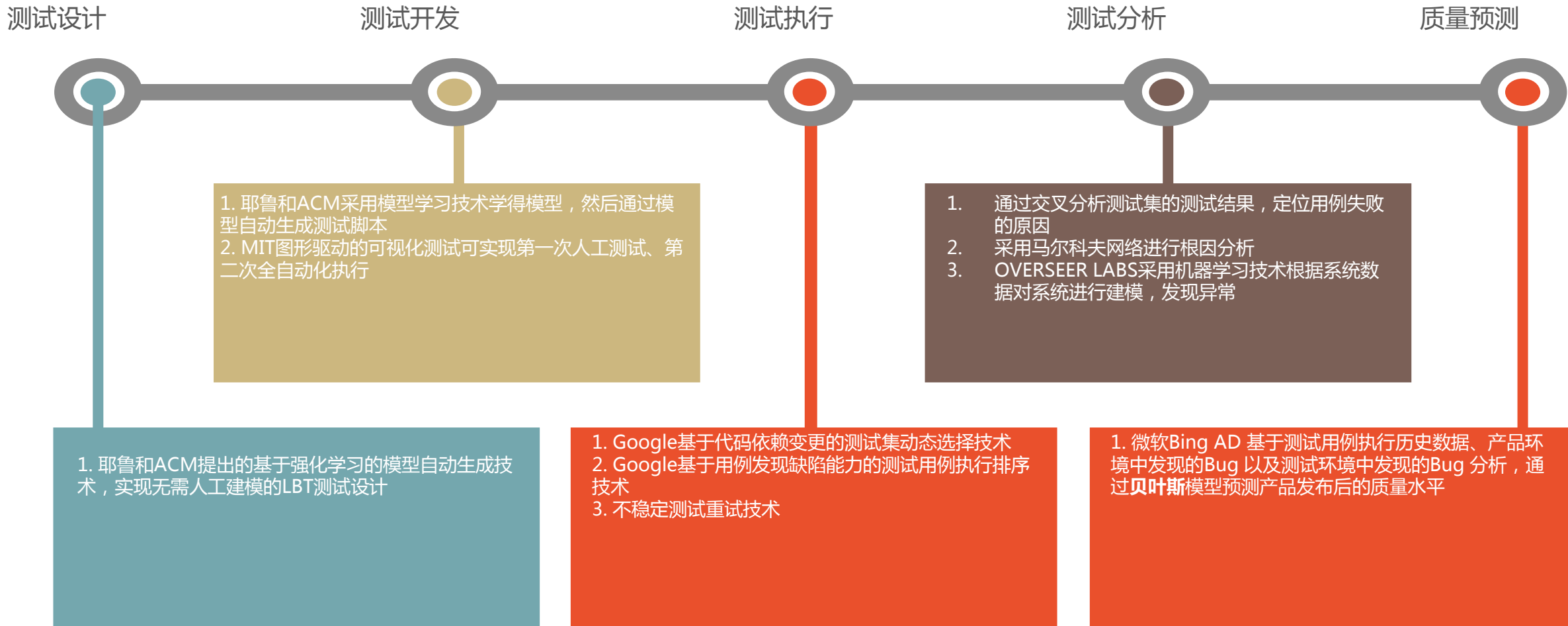


- $\theta^* = \operatorname{argmin}_{\theta} L(\theta)$
- 求解 θ^* : 梯度下降法 (Gradient Descent)
 - 每一步都朝着进步的方向, 直到没办法再进步
 - 学习速率 η 是第一个影响参数



内容

- AI基本介绍
- AI在测中的应用
- 深度学习做日志分析案例分享



学习系统



Techniques for Improving Regression Testing in Continuous Integration Development Environments by Sebastian Elbaum, etc.

Table 1: Google Dataset: Distribution of Test Suite Executions per Testing Phase and Test Suite Size

Phase	Size	% Total	% Failing
pre-submit	Small	26	0.02
pre-submit	Medium	10	0.13
pre-submit	Large	6	0.27
post-submit	Small	35	0.01
post-submit	Medium	13	0.32
post-submit	Large	10	0.65

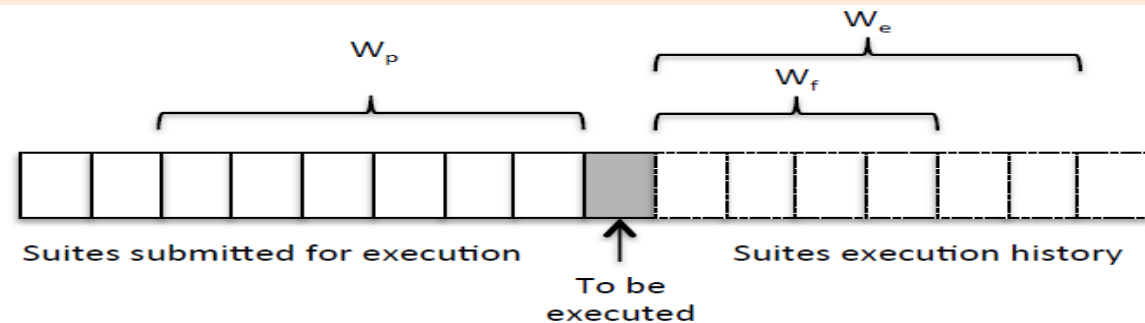
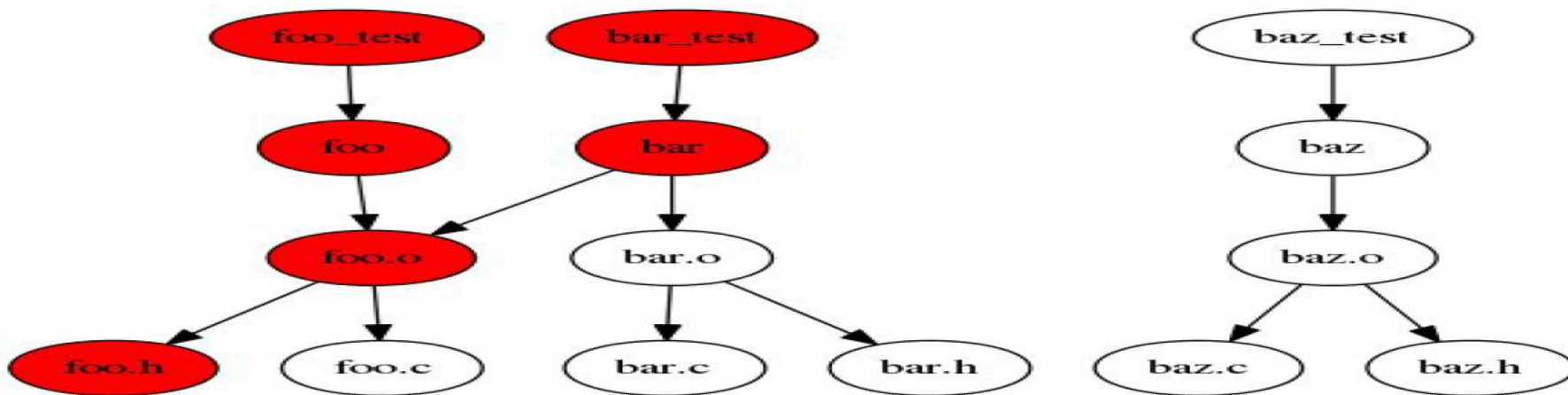


Figure 2: Test Suite Execution Windows

Regression Test Selection (RTS)



MS通过机器学习预测产品发布后的质量水平

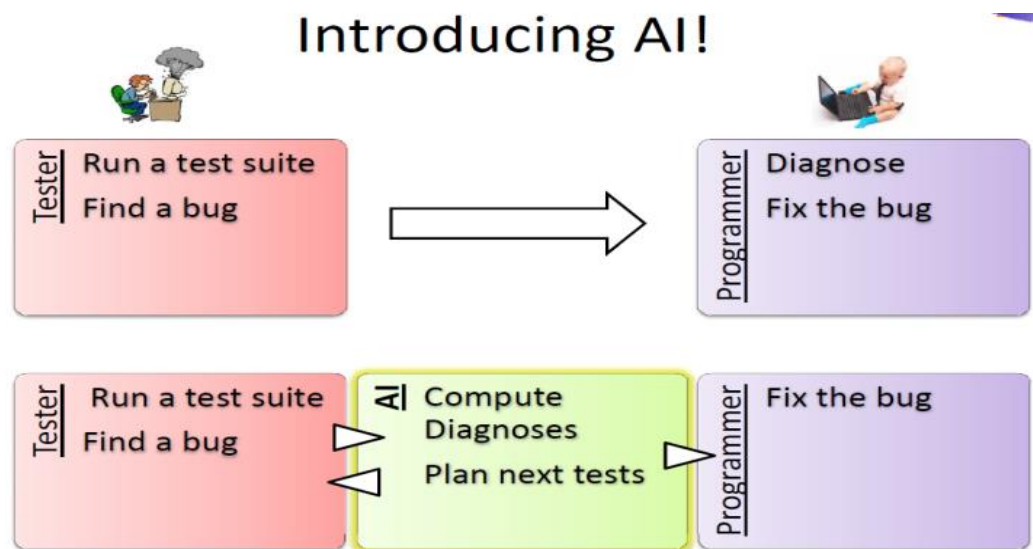
		Product	
		Success	Failure
Test	Pass	# PassingTest Cases	# of Product bugs
		+ # Fixed Bugs: State->closed, Resolved Reason->Fixed	+ # Live Site Records due to service bug in last 6 months
	Fail	# bugs resolved as: By Design, No Repro	# Bugs still not closed: Active or Resolved
		+ Issue: Test Defect	+ Bugs Resolved as: No Fix

执行步骤：

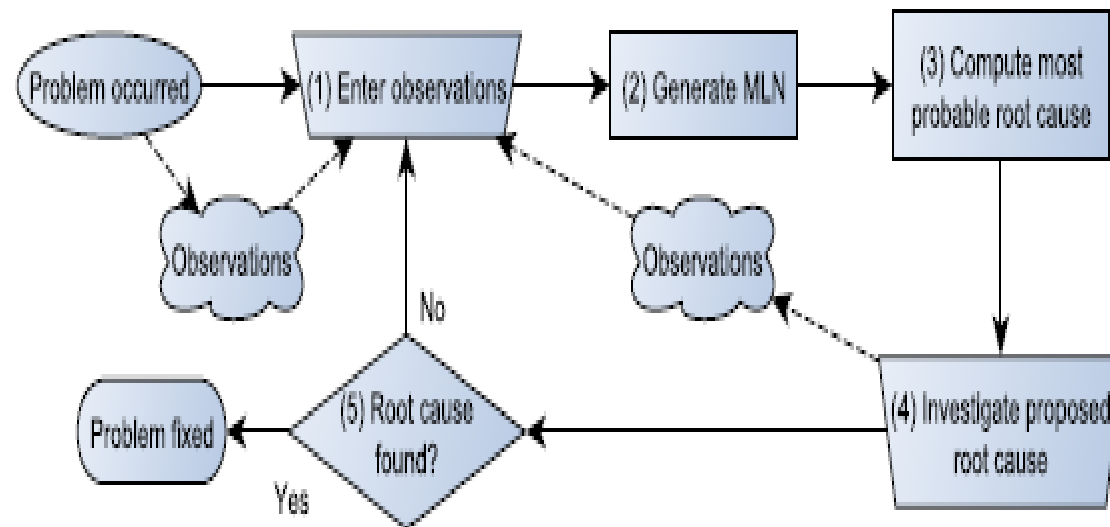
1. 针对发布之前的每个Build，统计测试用例通过数量和已修复Bug 数量 (Pass and Success)；
2. 统计By Design, Not Repro和Test Defect 的Bug数量(Fail but Success)；
3. 统计产品环境(和UAT环境) 被测试人员漏掉但是被发现客户发现的 Bug 数 (Pass but Failure)；
4. 统计当前活跃的Bug 数和发现但是没有被修复的Bug 数(Fail and Failure)；
5. 计算Likelihood, Prior, Evidence，以便最后推算出后验概率 Posterior；
6. 统计每个build 的后验概率，观察质量变化趋势。

1. Integrating Artificial Intelligence in Software Testing by Roni Stern,

- 1) 执行一个测试集
- 2) 某个用例失败发现缺陷
- 3) AI系统生成可能的失败原因
- 4) 假如只有一个失败原因，把原因直接反馈给相应工程师
- 5) 否则，AI规划一个新的测试集，排除某些原因



2. Root Cause Analysis through Abduction in Markov Logic Networks by Joerg Schoenfish, etc., 采用马尔科夫逻辑网络快速的找到问题的根源。通过马尔科夫网络对观察到的现象O和原因R进行概率建模，网络的概率由能量函数定义，然后通过最大似然概率估计网络，



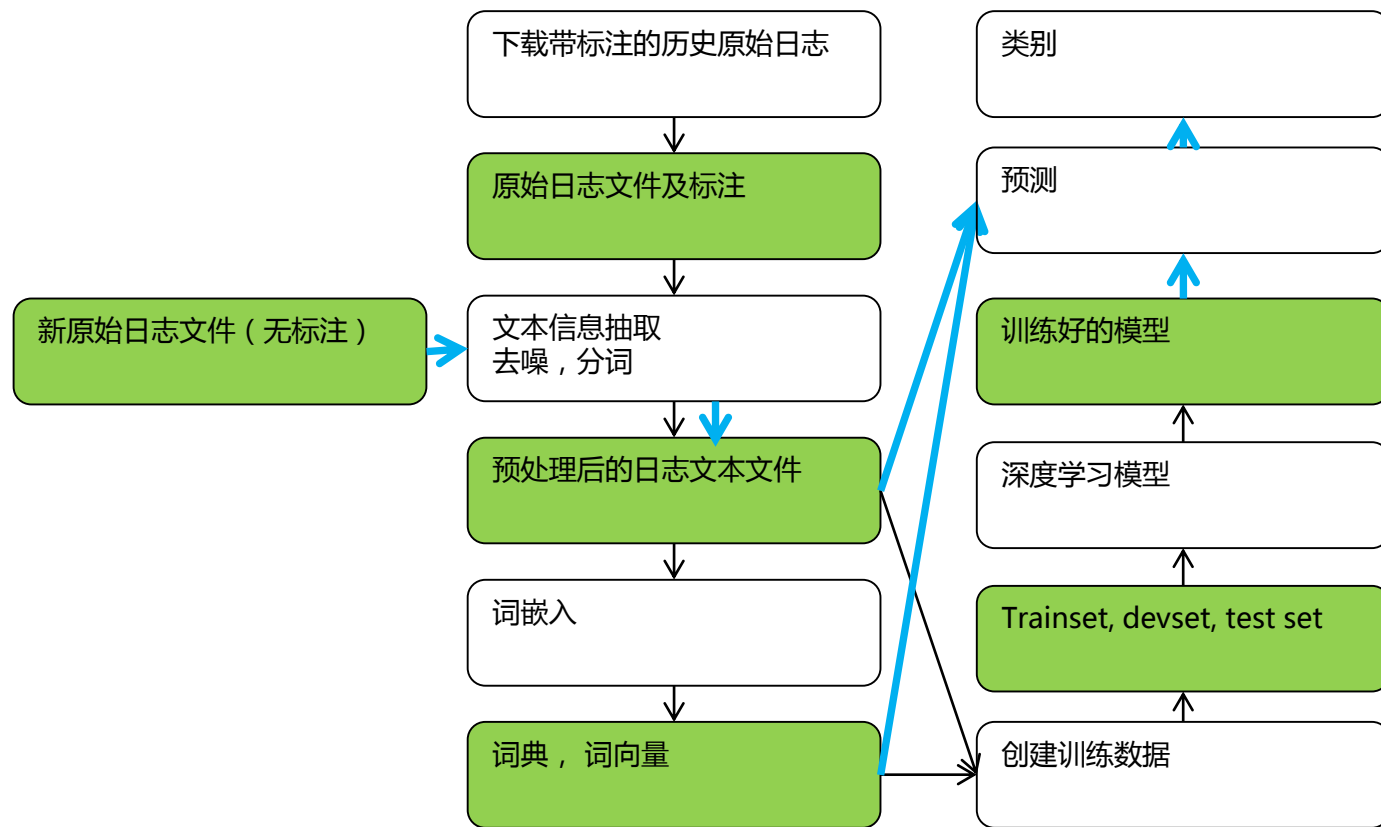
内容

- AI基本介绍
- AI在测中的应用
- 深度学习做日志分析案例分享

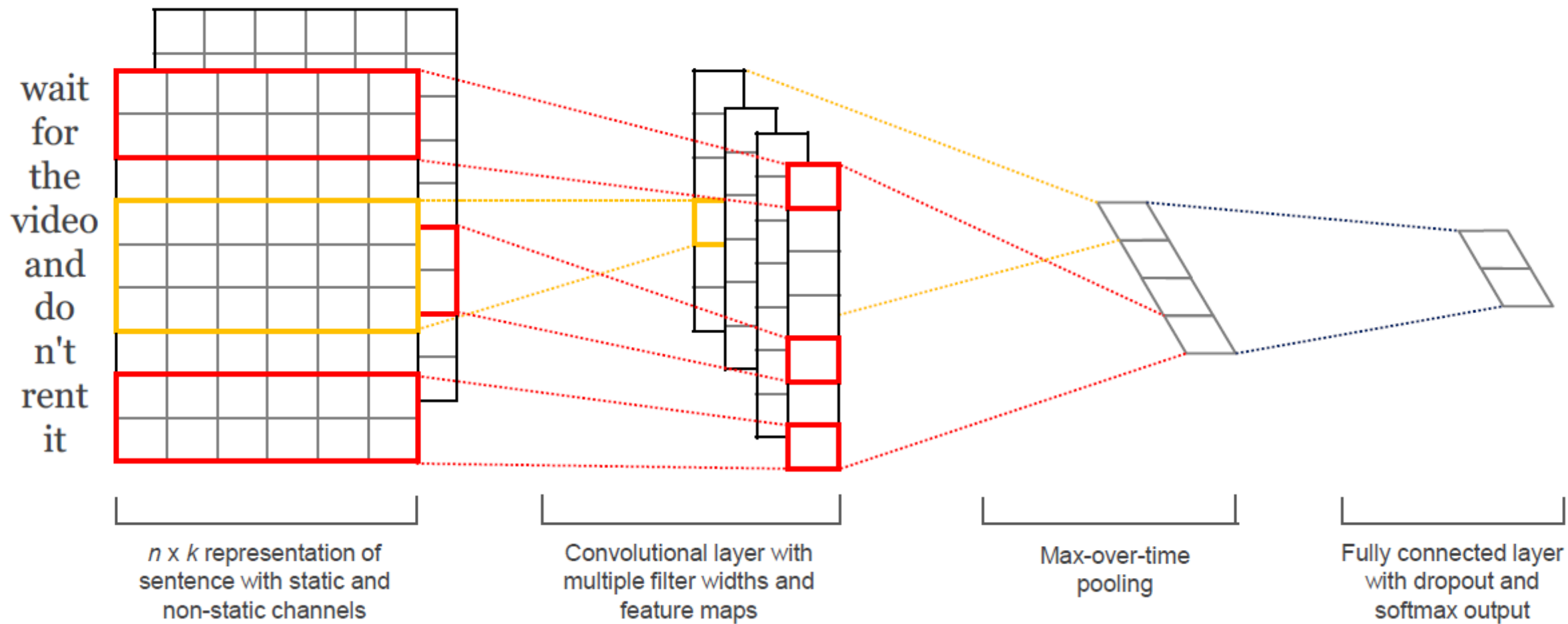
- 问题及数据来源：

- 问题：输入一个原始日志文件，输出一个出错原因类别
- 方法：由于人工抽取特征的困难性，采用深度学习对日志进行处理；采用监督学习训练分类器
- 数据来源：历史日志及人工分析的结论类别

- 方案：



论文：Convolutional Neural Networks for Sentence Classification by Yoon Kim from New York University

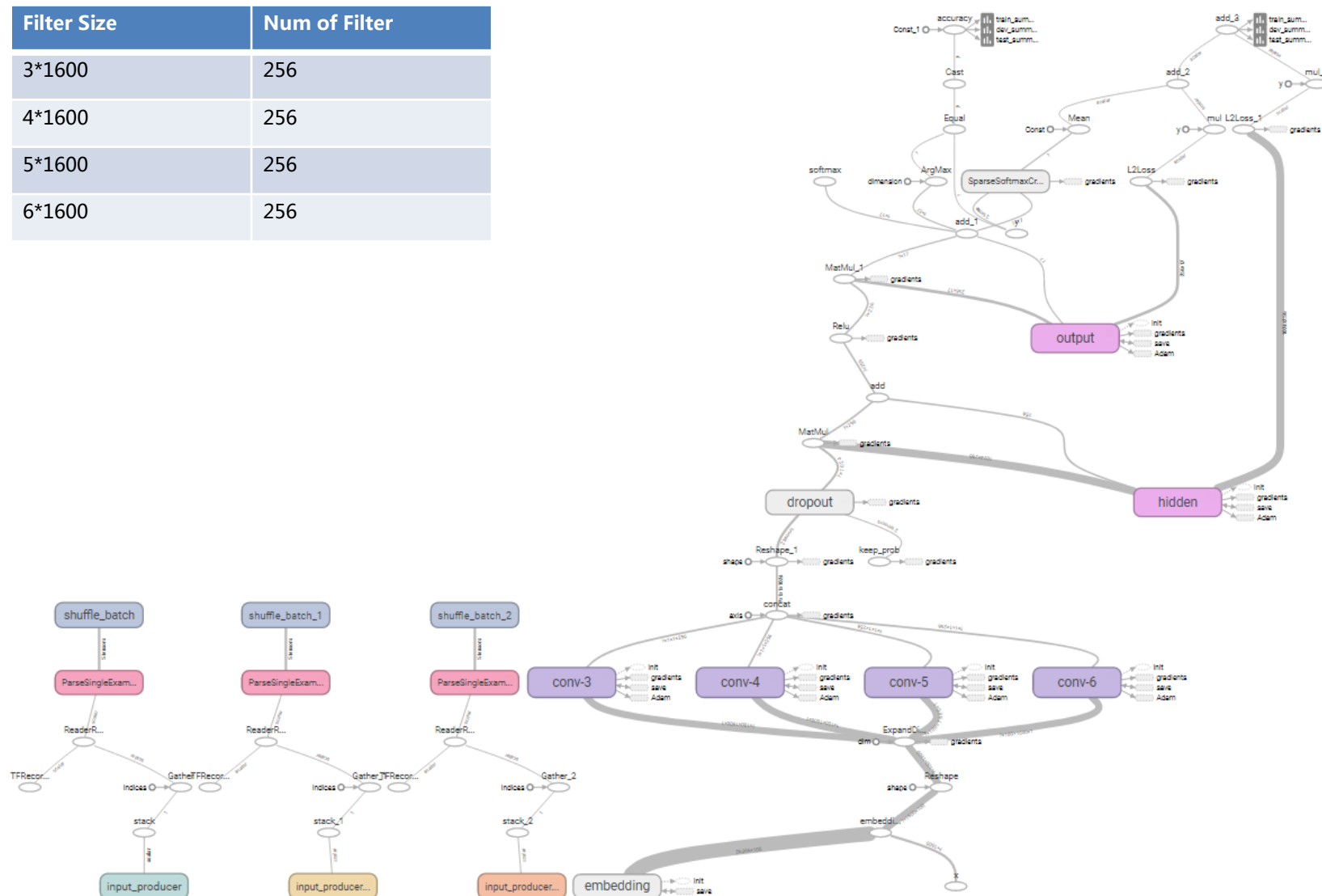
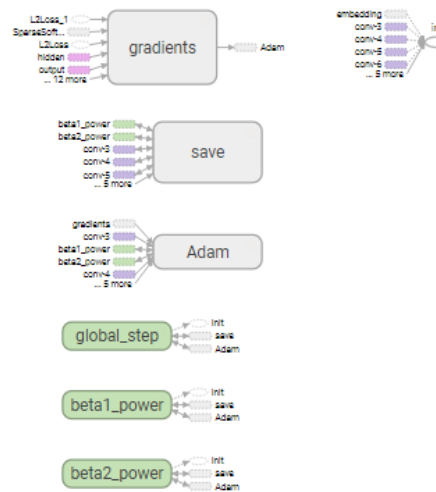


实现：模型计算图

Main Graph

Filter Size	Num of Filter
3*1600	256
4*1600	256
5*1600	256
6*1600	256

Auxiliary Nodes



- 训练数据采用过采样
- 测试数据不采用过采样
- 训练数据和测试数据取自不同的时间段
- 验证集从训练数据中分出一部分
- 参数：文档每行的最大词数，文档最大函数。如果文档太长，需要考虑如何截断。目前采用去尾部的数据，感觉上取出错位置上下文的截断效果要好一些，不过没有测试，同时要也考虑智能调测打印的数据
- 如果某些类别的样本太少，可以归为others类

- 双向LSTM
- 层级注意力机制GRU
- CNN
- 目前实验结果看 CNN要好于 层级注意力机制的GRU好于不带注意力机制的双向LSTM
- CNN好在收敛快，泛华准确率高。
- CNN好的原因之一：日志并不是完全的自然语言（虽然有点像），日志中的信息是关键词，句。可以对词，句子做卷积然后采用max pooling，让网络自动抽取关键词，句。
- 这里面有太多的参数需要去调试，实验（reg，dropout，等，网络结构，超参数等）
- 采用GPU训练，比cpu有8~10倍加速

- 训练过程保存check point，日志等。

THANK YOU

Copyright©2016 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.