

# Bayesian Networks Model

## 1. 建模结构

模型通过两个系统评价指标来衡量系统的风险：Apdex 和 Health。其中Health指标的计算还需进一步确定；Apdex指标主要是由CPU使用率、内存使用率、http请求数、thread数影响，是一个综合指标（比如用http平均响应时间来刻画）。

$T_1$	$T_2$	$\dots$	$T_n$
$cpu$	$cpu$	$\rightarrow$	$cpu$
$memory$	$memory$	$\rightarrow$	$memory$
$http$ 请求数	$http$ 请求数	$\rightarrow$	$http$ 请求数
$Threads$	$Threads$	$\rightarrow$	$Threads$
$\Downarrow$	$\Downarrow$	$\dots$	$\Downarrow$
$p(A)$	$p(A)$	$\dots$	$p(A)$

1. 通过历史数据构建贝叶斯网络，刻画上述系统指标对 $p(A)$ 的因果关系；
2. 构建每一个系统指标的时间序列，推测每个系统指标在未来某一时刻的值；
3. 通过贝叶斯网络得到 $p(A)$ 在未来时刻的值。

## 2. 设想

### 2.1. 关于预测没有出现过的样本

在离散情况下，贝叶斯方法是对历史数据的统计。对于没有出现过的样本，不能做出预测。如果将离散的历史数据拟合成连续的分布函数，就可以刻画没有出现的样本的概率，然后进行预测。

疑问：将拟合函数延伸到未出现情况的可靠性。

## 3. 明年任务：贝叶斯

1. 风险定位：通过因果推理定位已知风险的根源问题；
2. 因果推理预测：通过因果推理预测未知潜在风险。

## 4. 数据分析

### 4.1. 产品功能

1. 异常检测
2. 贝叶斯预测

## 5. 套路有限

1. AD
2. BN $\Rightarrow$ (关联影响)
3. TSP(应用最广)
  - (a) 是否能够预测，需要说明原因；
  - (b) 预测。

## 6. 行业经验

对于客户行业的了解，有利于我们抓到客户的痛点，为分析类似业务寻求方向。

行业数据：

1. 微信号：狗熊会
2. 新浪财经