信息系统工程重点实验室

National Key Laboratory of Science and Technology on Information System Engineer

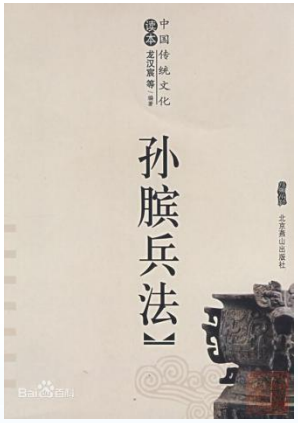
网络空间意图推断与行为分析研究组2019年汇报

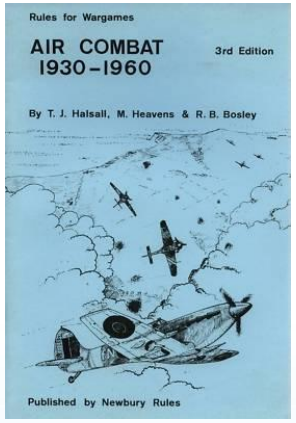
意图推断

- 什么是“意图”？
  - 行为意图推断是对网络空间中的内部、外部的异常行为进行分析判断：
    - 判断当前系统是否遭受外部攻击?
    - 攻击类型可能是哪些?
    - 是否存在内部违规操作?
    - 违规对象是什么?
    - 核心资产是否被侵害?
    - 最危险的核心资产是什么?

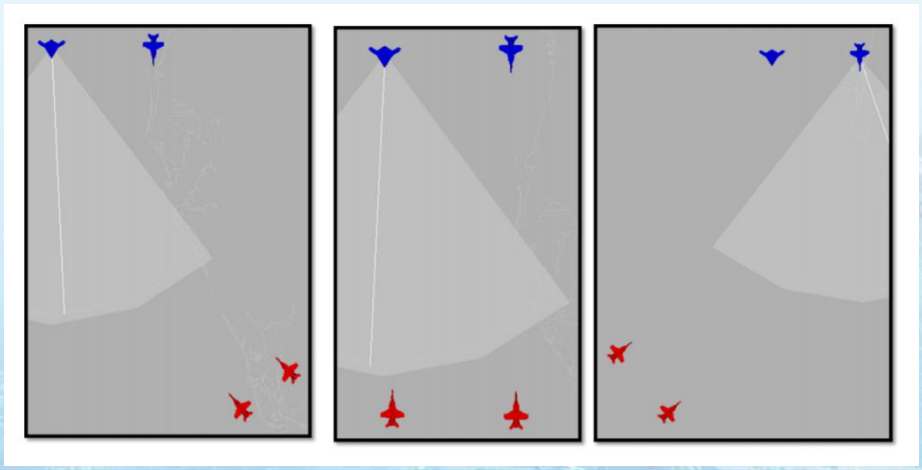
意图推断

- 通过阵形判断意图（威胁）
  - 中国古代阵法（十阵）
  - 现代空战对抗（编队进攻、防御）





圆阵 Vs. 疏阵



意图推断

- 意图推断的要素：

目的

类型

Types of Cybersecurity Attacks

Common Types of Cybersecurity Attacks

Phishing Attacks

SQL Injection Attacks (SQLi)

Cross-Site Scripting (XSS)

Man-in-the-Middle (MITM) Attacks

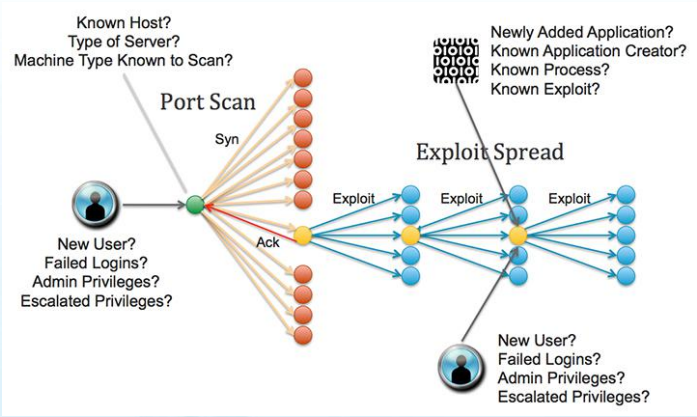
Malware Attacks

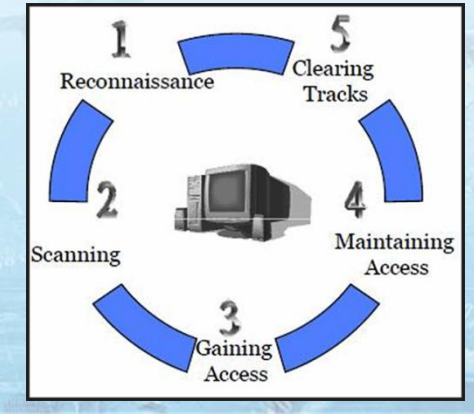
Denial-of-Service Attacks

Spear Phishing Attacks

Whaling Phishing Attacks

过程





行为目的推断

- 问题描述：通过分析对象的历史行动序列间的关系，推断该对象的意图或目的。
  - 常见关系：因果、时序、协同、惯例等。
  - 主要特点：区别于从有噪声数据中挖掘行为模式的“浅层”识别，其更注重通过关系分析，推断“深层”的行为意图。
  - 主要应用：人机交互、智能家居、影像监测等。
  - 目前的研究主要集中于在**物理空间**对**物理实体**对象的推断，较少有研究涉及网络空间虚拟实体。

行为目的推断

- 发展第一阶段：基于规则库，构建推断规则
  - ✓ Roger C. Schank and Robert P. Abelson. Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum Associates, Mahwah, NJ, 1977.
  - ✓ Robert Wilensky. Why John married Mary: Understanding stories involving recurring goals. Cognitive Science, 2(3):235-266,1978.
  - ✓ 对于复杂现实问题，规则的一致性、泛化性方面存在困难。

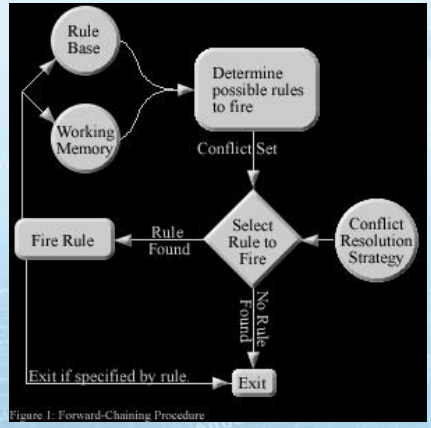
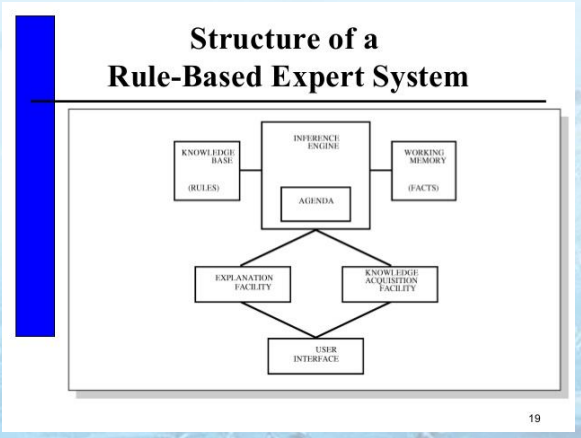
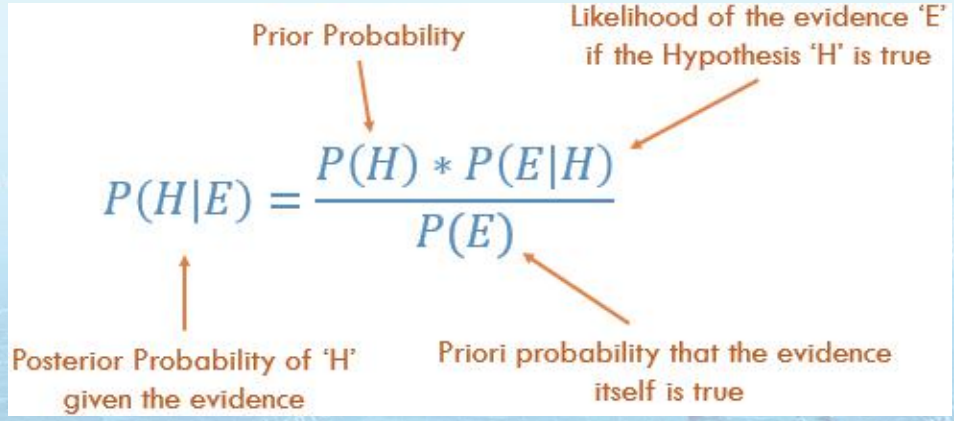


Figure 1: Forward-Chaining Procedure



行为目的推断

- 发展第二阶段：基于概率推断方法，Bayesian 推断
  - ✓ E. Charniak and R. P. Goldman. A Bayesian model of plan recognition. Artificial Intelligence, 64(1):53{79, November 1993.
  - ✓ Charniak and Goldman 将行为目的识别问题归纳为一个概率推断问题；
  - ✓ Bayesian方法是目前行为目的识别的主流方法。



行为目的推断

- 发展第三阶段：基于语法剖析的方法，Parsing problem
  - ✓ M. Vilain. Getting serious about parsing plans: A grammatical analysis of plan recognition. In Proceedings of National Conference on Artificial Intelligence, 1990.
  - ✓ Vilain将意图识别问题转换为一个语法剖析问题，将行动-意图的关系映射为句子-语义树的关系；
  - ✓ 基于Parsing的意图识别方法要求行动严格满足序列化要求，无法处理部分序列化的问题。

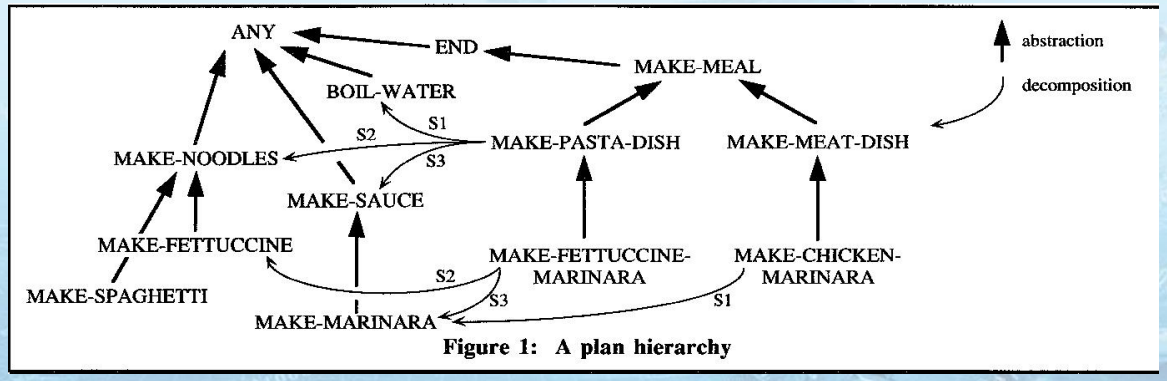
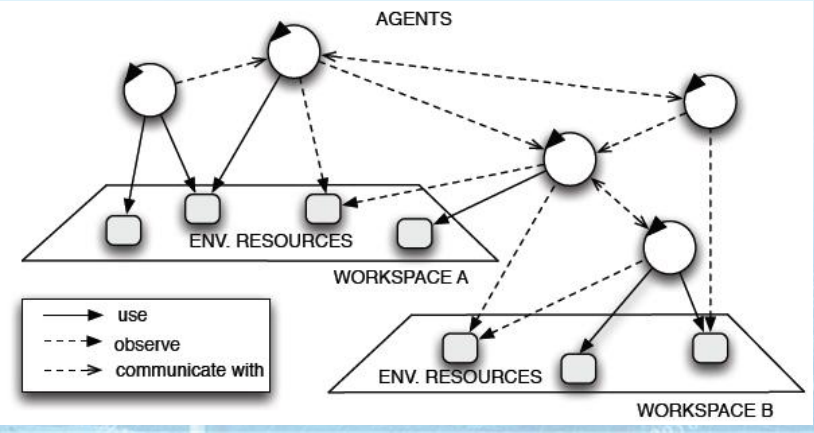
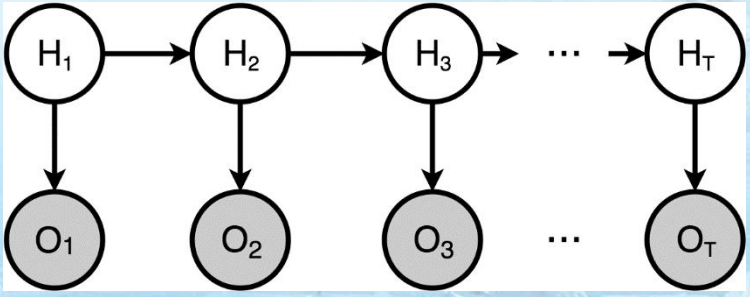


Figure 1: A plan hierarchy

行为目的推断

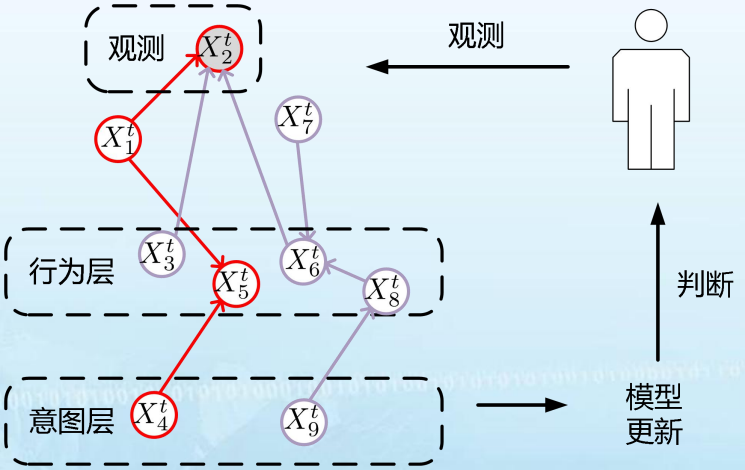
- 发展第四阶段：基于agent行动模型，HMMs
  - ✓ Hung H. Bui, Svetha Venkatesh, and Geoff West. Policy recognition in the Abstract Hidden Markov Model. Journal of Artificial Intelligence Research, 17:451-499, 2002.
  - ✓ Hung等通过建立agent模型反映对象的行为模式，用于实现意图识别；
  - ✓ 层次化的HMMs成为处理复杂意图识别的热点研究方法。



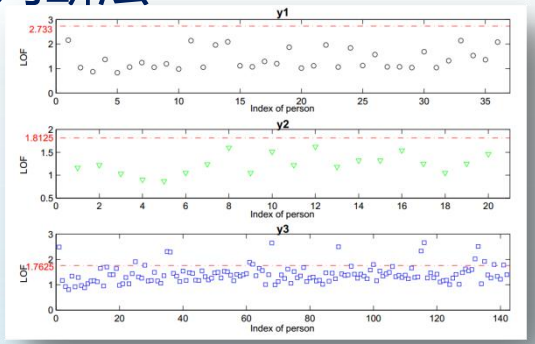


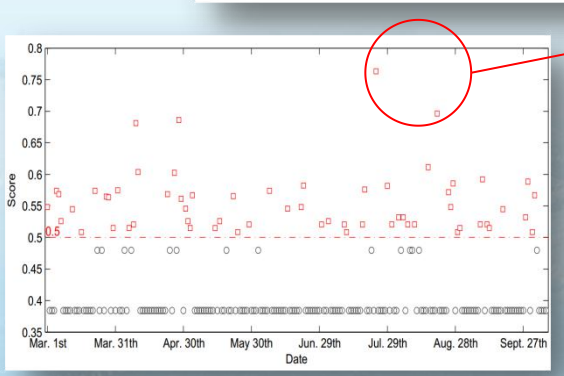
行为目的推断（案例）

- 问题描述：OA办公系统的用户行为目的分析
- 主要方法：从异常人员到异常时间的两阶段判断法
- 案例：异常消息群发目的分析



异常人员





异常时间

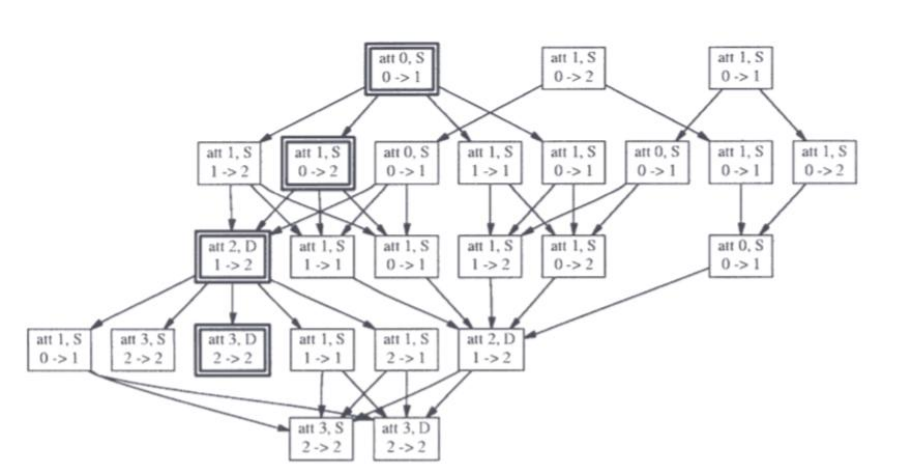
数据→异常行为检测→意图的智能推断

行为类型推断

- 发展第一阶段：基于逻辑和结构组合的推断模型建模方法
  - TVA model – [Jajodia et al., 2005] Center for Secure Information Systems, George Mason University
    - **将low-level vulnerabilities组合得到high-level attack goals**
    - **基于该工作，张永铮、方滨兴等人提出风险传播模型及网络节点相关性研究**
  - MulVAL – [Ou et al., 2005] Princeton University
    - **Multihost, multistage Vulnerability Analysis**
    - **基于OVAL描述逻辑语言[A Subset of Prolog]**
    - **vulExists(webServer, 'CAN-2002-0392', httpd)**
    - **vulProperty('CAN-2002-0392', remoteExploit, privilegeEscalation)**

行为类型推断

- 发展第二阶段：基于图模型的推断模型建模方法
  - Attack graphs – [Lippmann et al., 2002, 2005, 2009] MIT Lincoln Laboratory
    - 有限状态模型(S, τ, S_0, S_s)
    - S代表节点集，描述攻击的状态
    - τ代表边集，即状态转移关系
    - S_0, S_s 分别代表初始状态和攻击成功状态



- ✓ 攻击过程按树状结构展开；
- ✓ 从叶节点到根节点的任一路径描述了攻击达到目标状态经历的攻击算子步骤 atomic attacks；
- ✓ 加粗部分描述了其中的一条攻击路径。

行为类型推断

- 发展第三阶段：基于概率图模型的推断模型建模方法
  - Bayesian network-based attack graphs – [Frigault and Wang, 2008] Concordia University
  - NIST Interagency Report 7788 [Singhal and Ou, 2017]
    - 有向无环图
    - 节点代表状态
    - 边代表状态转移的概率关系

Case 3: To Reach the Goal State, both A and B must first be exploited and then C

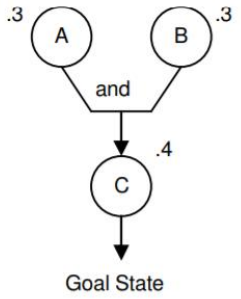


Figure 6: 3 Node Attack Graph

A	
T	F
.3	.7

B	
T	F
.3	.7

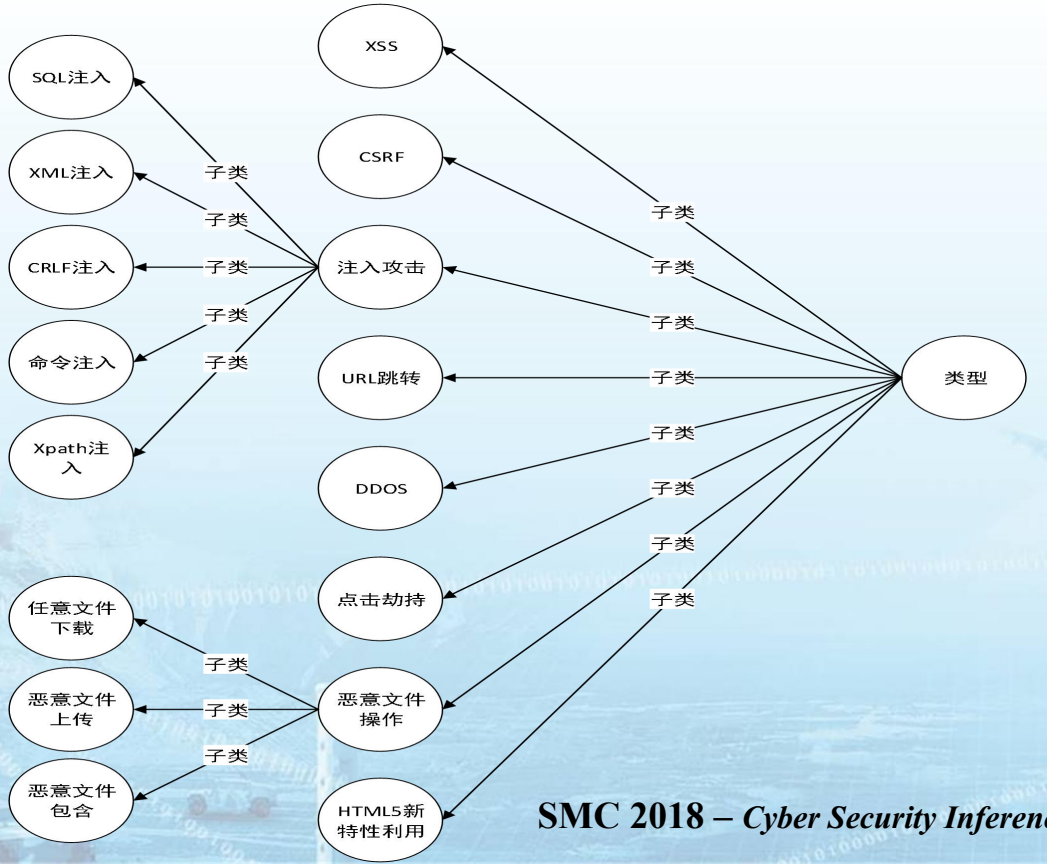
C			
A	B	T	F
F	F	0	1
F	T	0	1
T	F	0	1
T	T	.4	.6

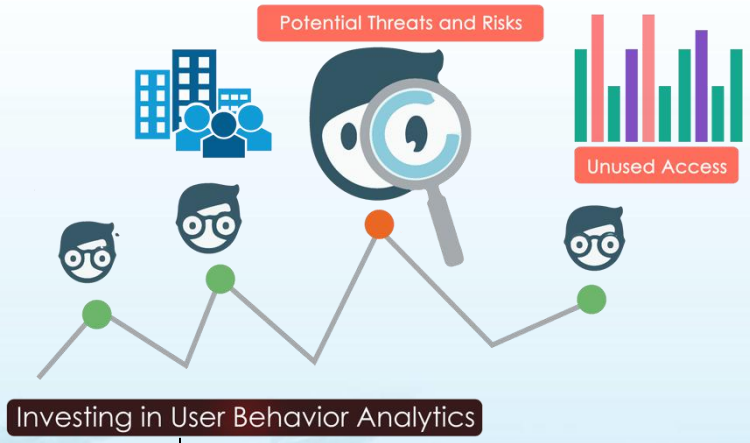
Figure 7: Conditional Probability Tables

- ✓ A、B条件同时满足时C满足；
- ✓ A、B的发生可以用边缘概率分布表示；
- ✓ “逻辑与”的攻击图可以通过设计C的条件概率表，来实现。

行为类型推断（案例）

- 目前关注14种类型的外部网络攻击行为，3种类型的内部用户威胁行为





① 内部浏览、阅读行为
② 内部交互留言行为
③ 内部数据、系统操作行为

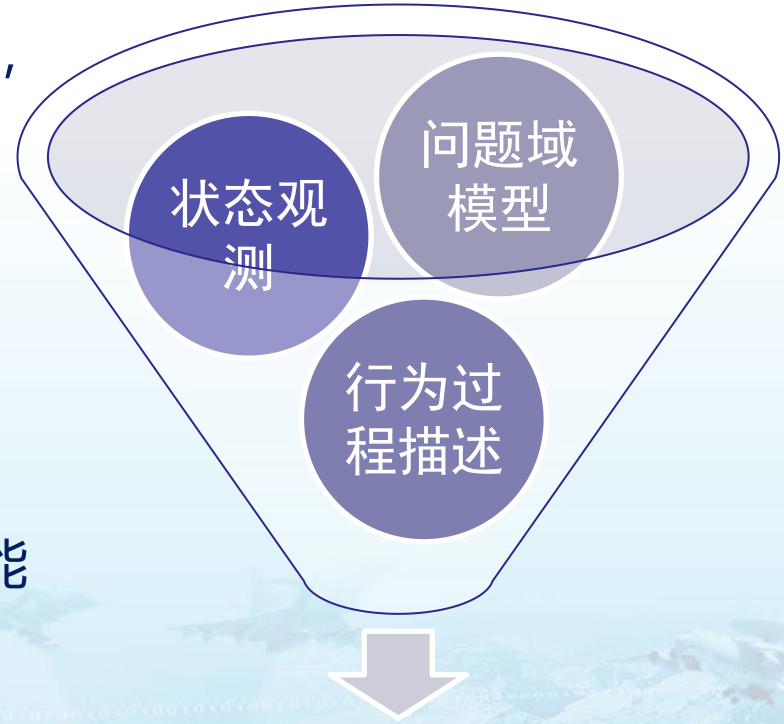
行为类型推断（案例）

- 问题描述：常见Web攻击行为类型推断
- 主要方法：基于机器学习（概率图）模型
- 案例：XSS攻击检测

```
<A HREF="h
http://6 6.000146.0X7.147/"">XSS</A>
<BODY
BACKGROUND="javascript:alert('XSS')">
<SCRIPT ="">"
SRC="http://3w.org/xss.js"></SCRIPT>
```

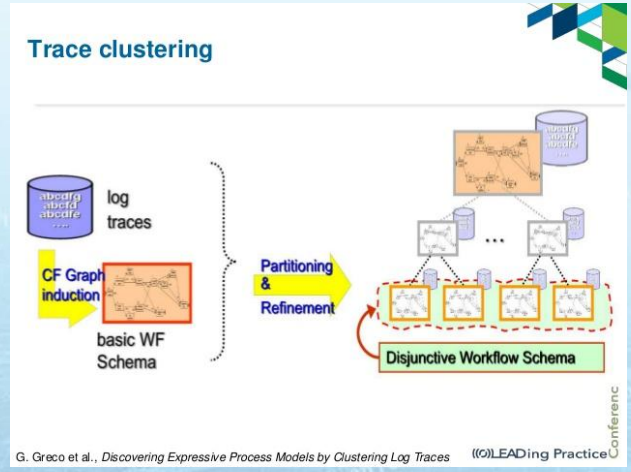
行为过程推断

- 问题描述：通过对问题的观测和建模，实现对象攻击行为过程的推断
  - 输入：状态观测、问题域描述、行为过程描述
  - 输出：攻击行动计划描述、可能的攻击动作、以及相关攻击概率
  - 如关键通道分析



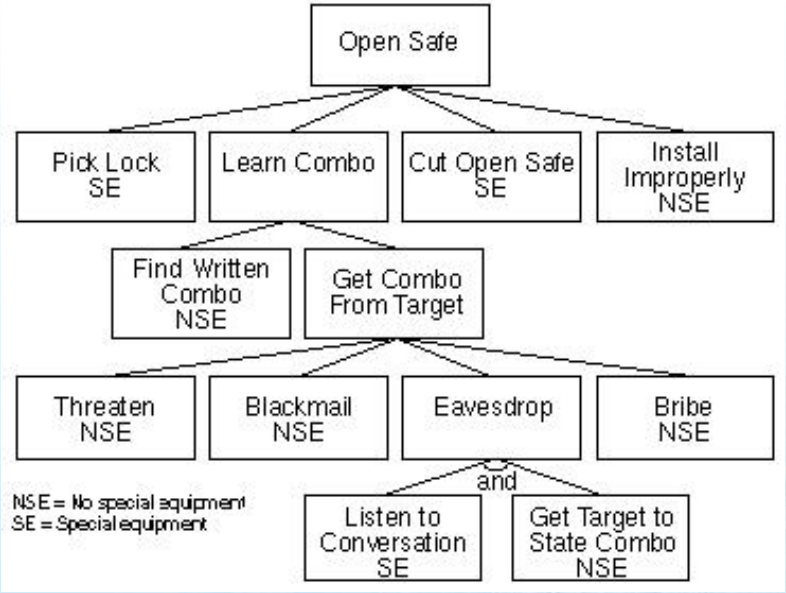
行为过程推断

- 发展第一阶段：基于聚类和攻击场景的数据挖掘方法
  - 入侵过程聚类Trace Clustering [Staniford et al. 2002，Valdes et al. 2001]
  - 基于攻击场景的数据挖掘方法
    - **STATL基于状态的攻击描述模型 [Eckmann et al. 2002]**
    - **LAMBDA基于逻辑攻击描述模型 [Cuppens & ortalo 2000]**
    - **利用挖掘方法识别攻击场景 [Dain & Gunningham 2001]**



行为过程推断

- 发展第二阶段：基于因果网络的方法 [Qin and Lee 2004]
  - Step 0：攻击聚类，警报分级（预处理）
  - Step 1：根据专家知识，建立攻击树
  - Step 2：用因果网络或者贝叶斯网络标记攻击序列变化概率
  - Step 3： 概率估计与攻击预测

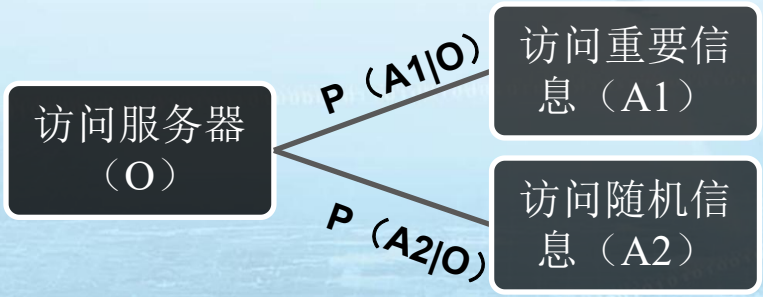


行为过程推断

- 发展第三阶段：基于相关性的多阶段的方法
  - 基于攻击相关性的多阶段攻击检测 [Cuppens and Miege 2002， Ning et al. 2004]
  - Step1： 用逻辑谓词描述攻击的条件和结果
    - 由谓词库和函数描述攻击行动和状态变化
    - 行动包括执行条件、可能效果、相关对象
    - 例如：具有相关性攻击建模语言（Correlated Attacking Modeling Language， CAML）[Cheung et al. 2003]

行为过程推断

- 发展第三阶段：基于相关性的多阶段的方法
  - 基于攻击相关性的多阶段攻击检测 [Cuppens and Miege 2002， Ning et al. 2004]
  - Step 2: 建立攻击行动相关图 [Ning & Xu 2003]
    - **逻辑相关性链接、时序相关链接**
  - Step 3: 学习多阶段攻击策略
  - Step 4: 分析结果



行为过程推断（案例）

- 问题描述：常见Web攻击过程推断
- 主要方法：贝叶斯网络诊断
- 案例：DDOS攻击还原

DDOS攻击观测

↓

计算其马尔科夫边界：

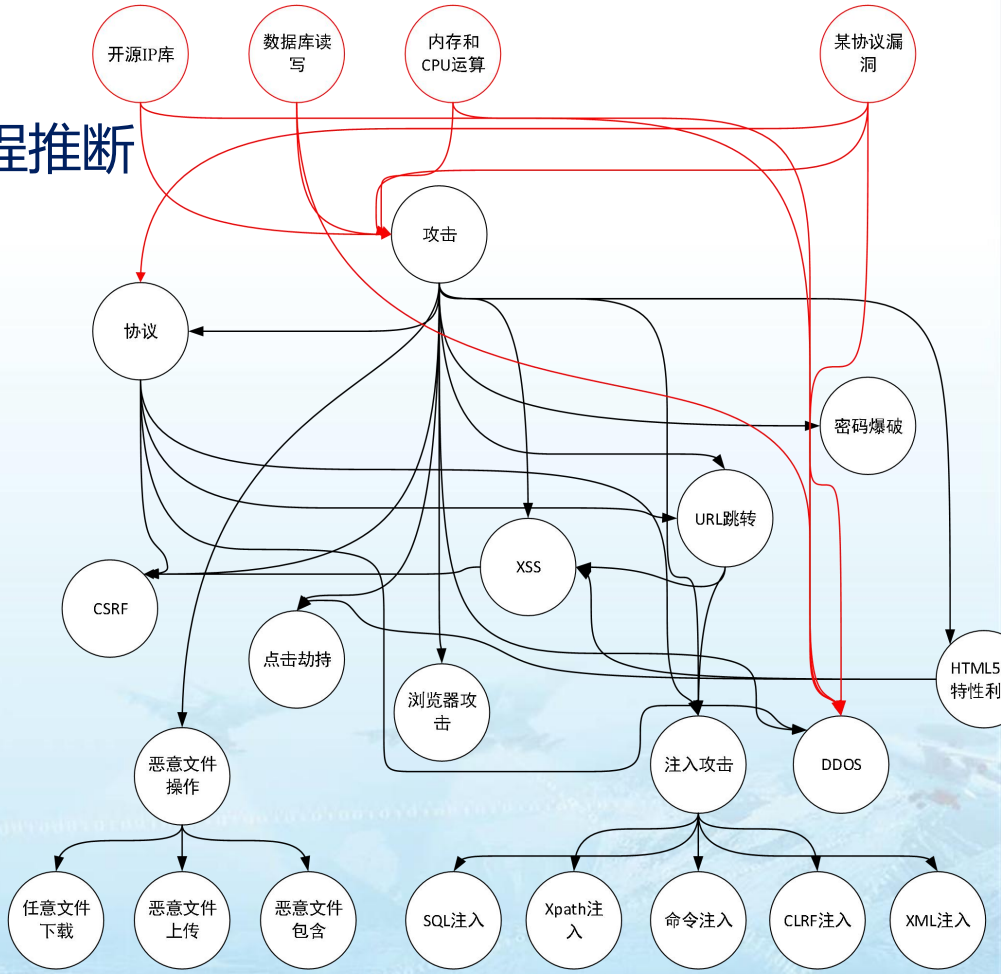
$MB(X) = pa(X) \cup ch(X) \cup (\cup_{Y \in ch(X)} pa(Y))$

↓

不同原因组合下的后验JS散度差

↓

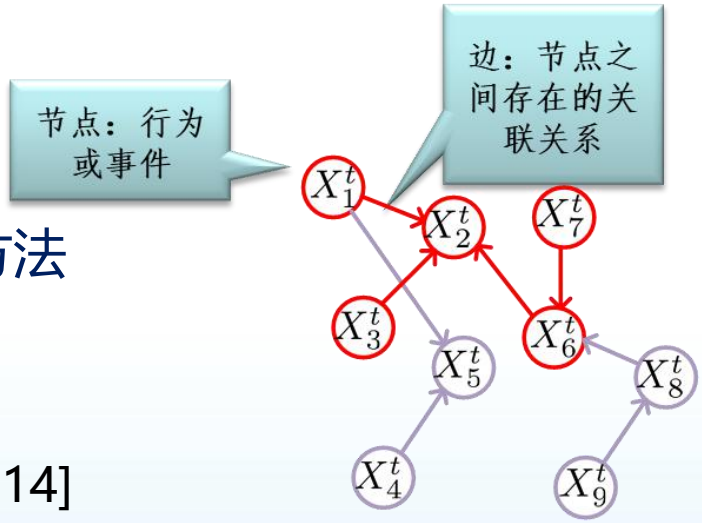
$JSD_{max}(X_i)$：IP库 > 数据库读写 > 协议 > …



预期创新

- 基于概率图的行为类型推断模型自学习方法
  - **目前较多的是攻击检测模型，如IDS系统**
  - **缺乏对异常原因、意图的自动分析和推断**
  - **模型可解释性差**
  - **用于推断的知识、背景、信息等难以综合利用与关联**
  - **推断计算空间庞大：**
    - **一般企业网络包含上千个节点 [Sharma et al. 2011; Raftopoulos and Dimitropoulos 2013];**
    - **主机节点一般有2-11个漏洞[WhiteHat Security 2015]，互联网网站节点平均有6.5个漏洞[Symantec 2015];**
  - **精确推断是NP-难问题。**

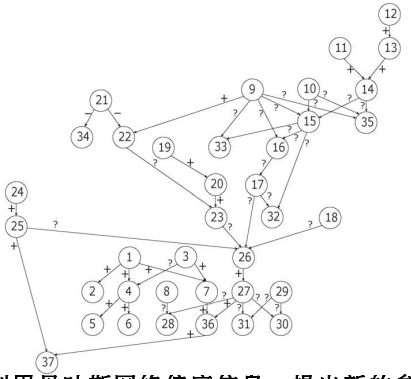
预期创新



- 基于概率图的行为类型推断模型自学习方法
  - **模型自学习**
    - ✓基于约束的学习
      - ✓Opt01SS [Villanueva et al., 2014]
    - ✓基于评分搜索的学习
      - ✓ILP method [Bartlett et al., 2017]
      - ✓Distributed PSO [Sahin et al., 2007]
    - ✓混合式的学习
      - ✓Parent reducing algorithm [Contaldi et al., 2016]
    - ✓参数学习
      - ✓Parameter learning with transferred priors and constraints [Zhou et al., 2015]
  - **存在问题：结构组合空间过于庞大，导致学习需要大量的标注数据支持。**

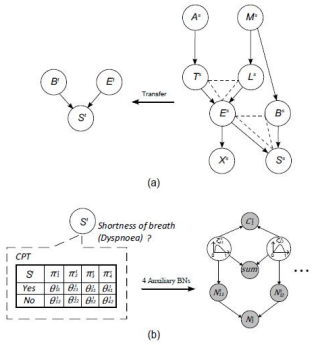
基于概率图的行为类型推断模型自学习方法

Decision Support Systems 2016
An empirical study of Bayesian network parameter learning with monotonic influence constraints



利用贝叶斯网络偏序信息，提出新的参数学习算法，提高了模型的学习精度

贝叶斯网络学习

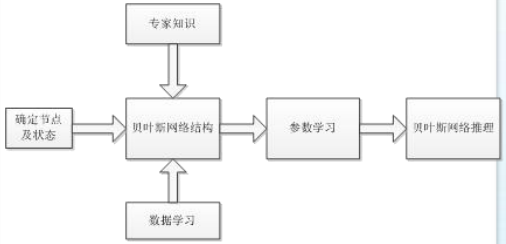


提出基于迁移和约束的贝叶斯网络学习算法，在12个标准测试集上领先

UAI 2015
Probabilistic Graphical Models parameter learning with transferred prior and constraints

针对现有贝叶斯网络结构学习算法中的启发式算法不稳定、准确率不高的问题，本文在现有启发式算法结合结构先验对现有方法进行改进；

启发式方法



将专家知识和数据学习都引入贝叶斯网络的结构学习中

结合专家知识

针对现有贝叶斯网络结构学习算法中的启发式算法无法找到全局最优解的问题，本文深入研究数值优化算法，将结构学习问题转化成非凸优化问题解决；

数值优化方法

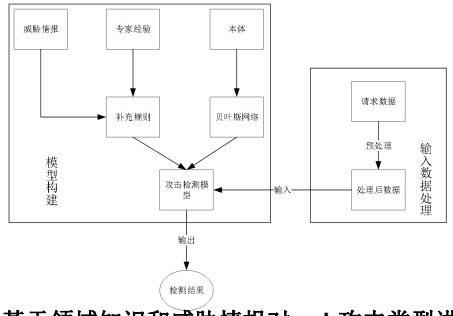
在研项目 1）国家自然科学基金青年项目, 61703416， “多任务贝叶斯网络学习及其应用”

论文
NO TEARS算法在XSS攻击检测中的应用研究.小型微型计算机系统. 2019
Bayesian Networks Structure Learning with Structure Prior, Information Sciences, 2020

信息系统工程重点实验室
National Key Laboratory of Science and Technology on Information System Engineer

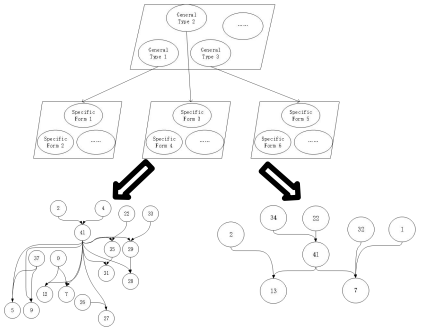
基于概率图的行为类型推断模型自学习方法

“纵横”网络空间安全创新论坛
基于领域知识和威胁情报的Web攻击检测研究



基于领域知识和威胁情报对web攻击类型进行推断

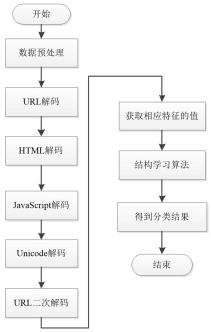
行为类型推断



将网络攻击进行两层建模，并利用推断算法对节点重要度进行获取

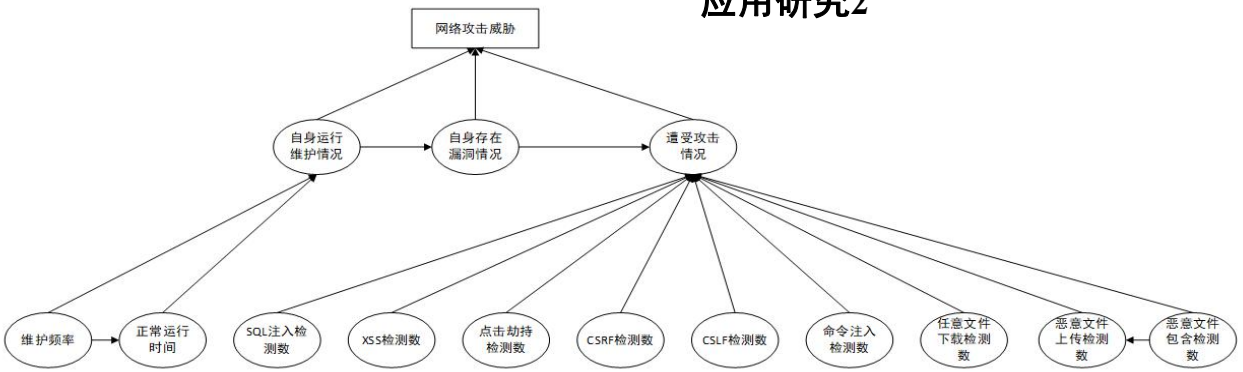
IEEE SMC 2018 Cyber Security Inference Based on a Two-level Bayesian Network Framework

应用研究1



NO TEARS算法在XSS攻击检测中的应用研究

应用研究2



基于贝叶斯网络，对网络威胁要素进行建模，利用评分搜索算法对模型进行结构学习，之后利用贝叶斯网络推理来对整体威胁进行风险分析和计算。

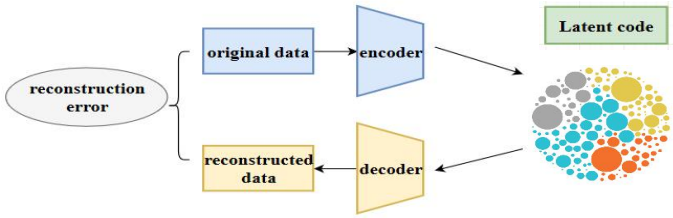
在研项目 1）国家自然科学基金青年项目, 61703416，“多任务贝叶斯网络学习及其应用”

在审专利 《基于两层贝叶斯网络模型的网络安全推断方法》

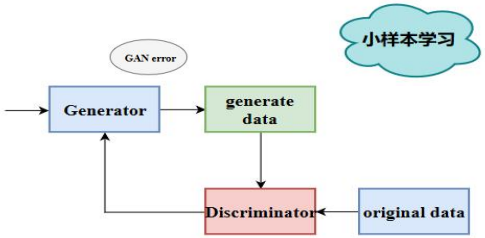
《一种检测XSS攻击的方法、装置及计算机可读存储介质》

《一种用户阅读兴趣主题漂移的检测方法》

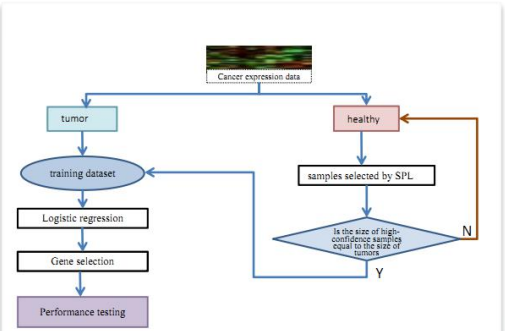
基于概率图的行为类型推断模型自学习方法



将原始数据映射到潜在空间，训练完成cvae，训练完成后的潜在空间作为生成器输入，旨在提供最优原始数据先验信息。



模型学习的方法创新



提出自适应自步学习对不平衡样本下采样

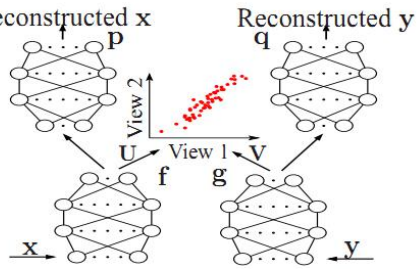
Under Review
Adaptive Sampling Using Self-paced Learning for Imbalanced Cancer Data Pre-diagnosis

在投一种新的多任务朴素贝叶斯学习方法

1. 基于 NB-STL 初始化，基于 $\mathbf{X}_r, \mathbf{y}_r$ 得到 $m=0$ 时的先验 $P_t^0(C_k)$ 和似然 $P_t(u_d \mid C_k)$，即 $\left[P_t^0(C_k), P_t(u_d \mid C_k)\right] = g(\mathbf{X}_r, \mathbf{y}_r)$
2. $P_t^*(C_k) \leftarrow P_t^0(C_k)$
3. $\hat{\mathbf{y}}_t^* = f\left(\mathbf{X}_r, P_t^0(C_k), P_t(u_d \mid C_k)\right)$
4. repeat
5. $\hat{\mathbf{y}}_t^m = f\left(\mathbf{X}_r, P_t^m(C_k), P_t(u_d \mid C_k)\right),\ t=1,2,\ldots,\mathcal{T}$
6. $\left(\mathbf{X}_{\mathcal{T}}, \hat{\mathbf{y}}_{\mathcal{T}}^m\right) = \left([\mathbf{X}_1; \mathbf{X}_2; \ldots; \mathbf{X}_{\mathcal{T}}], [\hat{\mathbf{y}}_1^m; \hat{\mathbf{y}}_2^m; \ldots; \hat{\mathbf{y}}_{\mathcal{T}}^m]\right)$
7. $P_{\mathcal{T}}^m(C_k) = g\left(\mathbf{X}_{\mathcal{T}}, \hat{\mathbf{y}}_{\mathcal{T}}^m\right)$
8. $P_t^{m+1}(C_k) = \alpha \cdot P_{\mathcal{T}}^m(C_k) + (1-\alpha) \cdot P_t^m(C_k)$
9. $\hat{\mathbf{y}}_t^{m+1} = f\left(\mathbf{X}_r, P_t^{m+1}(C_k), P_t(u_d \mid C_k)\right),\ t=1,2,\ldots,\mathcal{T}$
10. if $\xi\left(\left[\hat{\mathbf{y}}_1^{m+1}, \hat{\mathbf{y}}_2^{m+1}, \ldots, \hat{\mathbf{y}}_{\mathcal{T}}^{m+1}\right]\right) > \xi\left(\left[\hat{\mathbf{y}}_1^m, \hat{\mathbf{y}}_2^m, \ldots, \hat{\mathbf{y}}_{\mathcal{T}}^m\right]\right)$
11. $P_t^*(C_k) = P_t^{m+1}(C_k)$
12. endif
13. $m = m+1$
14. $\Delta = \sum_{t=1}^{\mathcal{T}} \sum_{k=1}^{K} \left| P_t^{m+1}(C_k) - P_t^m(C_k) \right|$
15. until $\Delta < \varepsilon$

针对现有多任务学习方法运行效率低和数据信息利用不足的问题，提出了一种新的多任务朴素贝叶斯学习方法。

- 不平衡样本学习
- 小样本学习
- 多任务学习



在不平衡样本中，将单视角问题转化为多视角问题，克服特征之间的高相关性带来的影响，提高的算法精度。

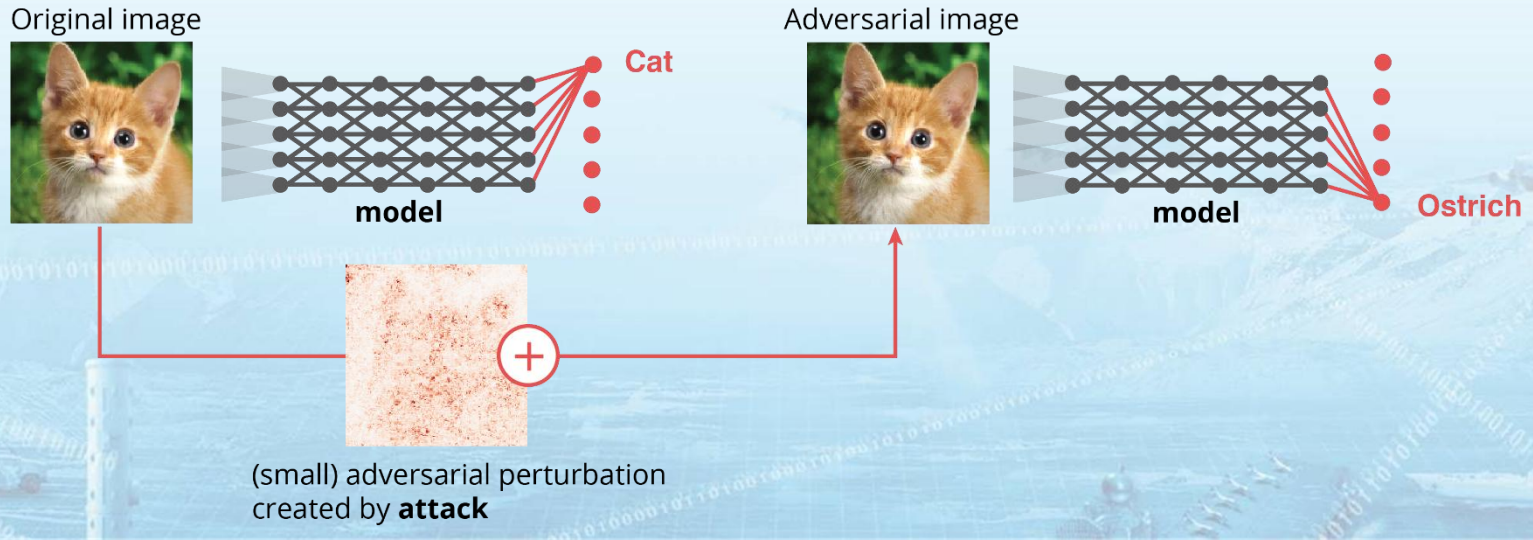
Generative Multi-view features adaptive high-confidence Sampling for Class-Imbalance Learning

在研项目

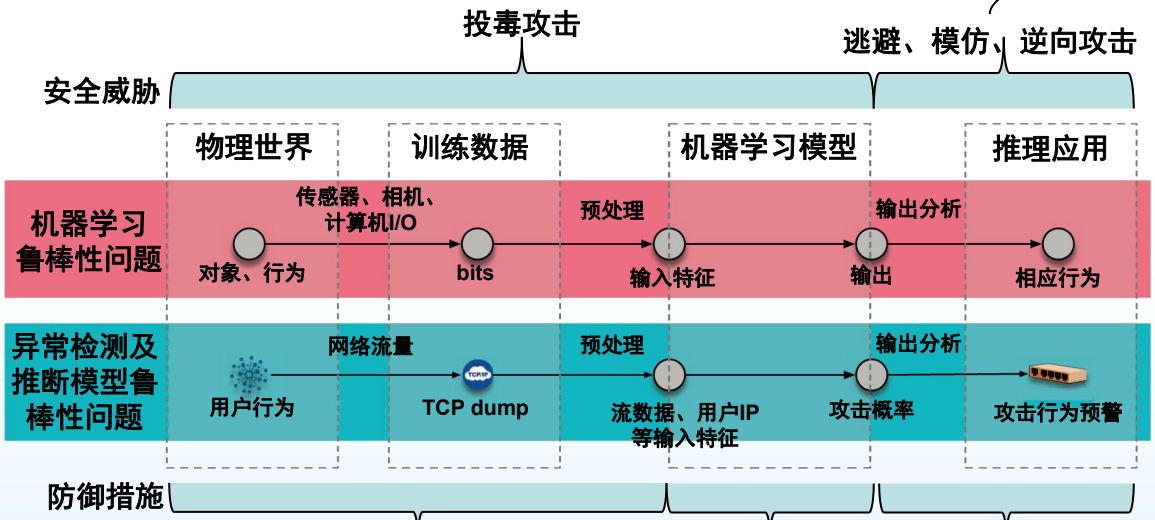
1）湖南省研究生创新项目, CX20190040，“贝叶斯网络机器学习新方法研究”；
2）湖南省研究生科研创新项目，CX20190039，“基于集成学习的高维小样本数据文本分类研究”

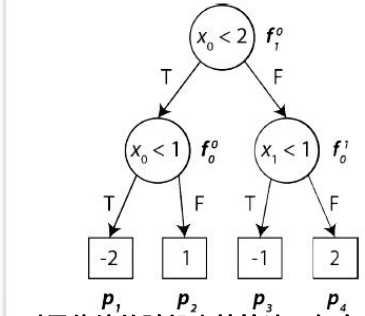
预期创新

- 异常检测及推断模型鲁棒性研究
  - **智能算法的鲁棒性、安全性问题**
  - **由于未来复杂的网络攻击也会针对智能防御算法进行设计**
  - **因此需要研究对抗条件下智能算法的鲁棒性、安全性，使得在欺骗攻击下也能达到可接受的性能**



异常检测及推断模型鲁棒性研究



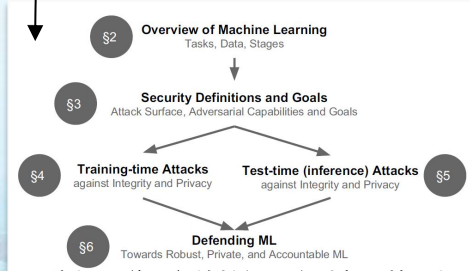


对于传统的随机森林算法，每个决策树在特征选择时具有较大随意性，攻击者可以计算模型中最有可能被选中的特征并对其进行攻击，针对随机森林算法易受攻击问题，开展相应的攻击和防御研究。

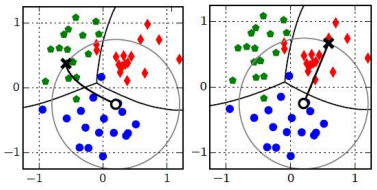
数据清洗

训练阶段：提高算法鲁棒性
测试阶段：安全评估机制

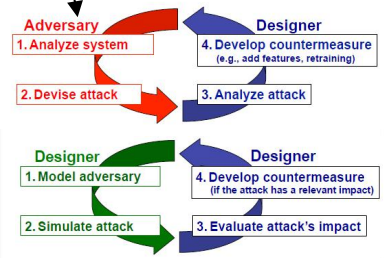
隐私保护技术



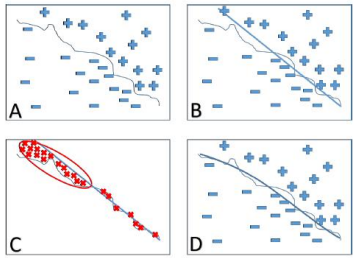
围绕机器学习鲁棒性问题全过程，梳理机器学习敌手模型、训练和推理过程中面临的安全威胁，算法安全评估技术以及防御机制，搭建机器学习鲁棒性研究框架，形成综述报告。



投毒攻击行为，主要采用在系统的训练过程中注入恶意数据的手段，研究如何对数据进行有效清洗对于提高模型鲁棒性有较好效果。



针对攻击与防御总是处于不断的博弈过程，研究机器学习中的博弈理论模型，迫使攻击者选择低成效的攻击策略。



对抗学习过程中，预测精度和对抗样本的鲁棒性不会同时提高，研究两者之间的交互作用具有较大意义。

项目论文

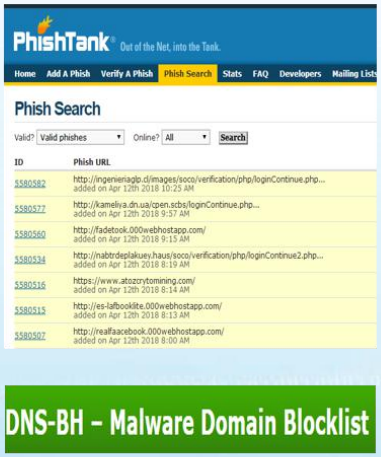
1）How can we deal with adversarial samples?
2）参加创新大赛-AI算法的抗攻击

预期创新

- 利用互联网开源威胁情报及领域本体知识对模型进行补充

IP库

恶意域名库

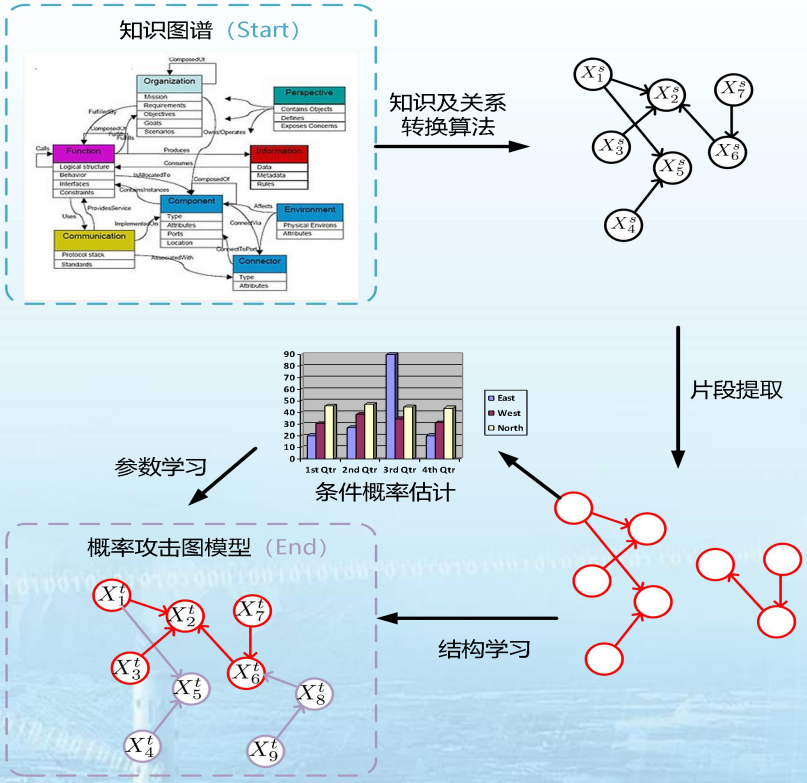


漏洞库

情报信息库

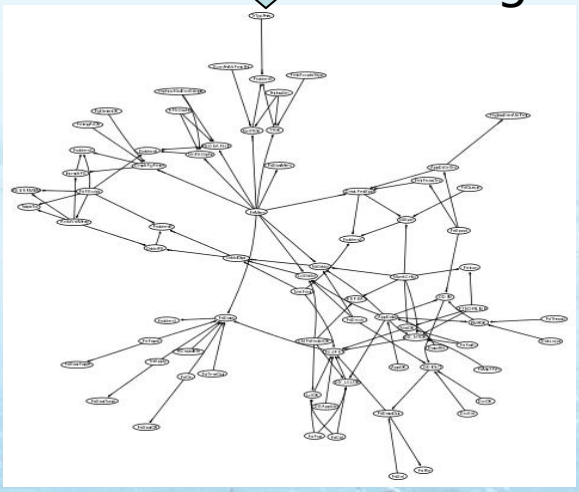
预期创新

- 利用互联网开源威胁情报及领域本体知识对模型进行补充



通过外部知识来补充节点、并减少搜索空间

Learning




对军队信息化建设和国防科技发展负有高度的责任心
献身国防科技事业为荣、以重点实验室一员为荣、以对科研事业的奉献为荣
责任
Duty
荣誉
Honor
Creativity
创造力
Impetus
推动力
创造性的开展基础性和前沿性研究
推动我国C⁴ISR领域全面的技术进步

敬请批评指正，谢谢！