网络空间安全态势感知

Cyber Security Situational Awareness

2021-2022学年（春） 专业普及课

态势感知中评估和预测方法

卢志刚

luzhigang@iie.ac.cn

中国科学院大学
University of Chinese Academy of Sciences

提纲

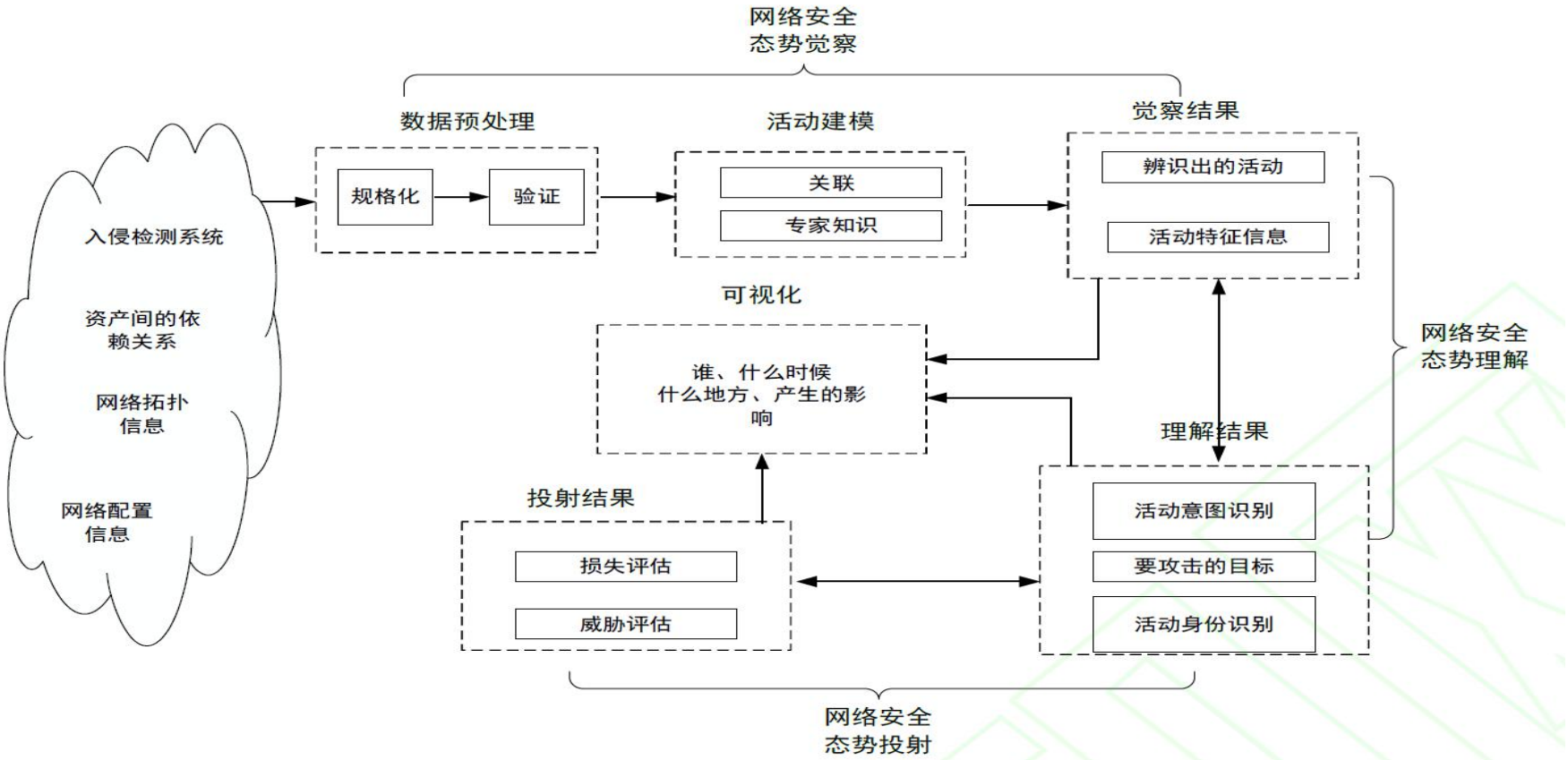
1、态势评估和预测的方法体系

2、态势评估方法

3、传统时间序列预测算法

4、机器学习预测算法

1、态势评估和预测的方法体系



网络安全态势感知NSSA 是对网络系统安全状态的认知过程,包括对从系统中测量到的原始数据逐步进行融合处理和实现对系统的背景状态及活动语义的提取,识别出存在的各类网络活动以及其中异常活动的意图,从而获得据此表征的网络安全态势和该态势对网络系统正常行为影响的了解.

1、态势评估和预测的方法体系

态势感知评估和预测的数学问题本质

- 态势理解
- 态势评估
- 态势预测
- 态势可视化

网络安全态势感知的核心环节
数据融合领域的研究重点

即在融合各安全信息并进行简单处理的基础上，通过一些数学方法或者数学模型，经过分析，得到一个对当前网络安全状态的整体描述。

简言之，该过程即态势因子集合到态势集合的映射。

模型中的评估/预测方法

分类方法1：2012-哈工大-博士-基于融合决策的网络安全态势感知技术研究_李志东

网络安全态势评估有多种分类方法
按评估的侧重点不同可分为风险评估和威胁评估
按评估的实时性可分为静态评估和动态评估
按评估的形式可分为定性评估和定量评估
按评估依据的理论技术基础可分为 3 类

类别	方法
基于数学模型的方法	层次分析法、集对分析法、模糊综合评价法、多属性效用函数法、距离偏差法等
基于知识推理的方法	模糊推理、贝叶斯网络、马尔可夫过程、D-S 证据理论等
基于模式分类的方法	聚类分析、粗糙集、灰关联分析、神经网络、支持向量机等

模型中的评估/预测方法

分类方法1：2012-哈工大-博士-基于融合决策的网络安全态势感知技术研究_李志东

基于数学模型的方法综合考虑网络安全态势的影响因素，构造从安全指标集合到安全态势的映射，藉此将态势评估归结为多属性聚集计算或多指标综合评价问题，能给出明确的数学表达式，也能得出确定性的结果。

基于知识推理的方法凭借专家知识及经验建立评估模型，通过逻辑推理评估安全态势，可以细分为基于产生式规则的推理(如模糊推理等)、基于图模型的推理(如贝叶斯网络、马尔可夫过程等)和基于证据理论的推理。其基本思想是借助模糊理论、概率论、证据理论等来表达和处理安全属性的不确定性，通过推理汇聚多属性信息。

基于模式分类的方法将安全要素或评价指标视为模式属性、离散型态势值视为模式类别、连续型态势值视为分类器输出的弥合值，先通过训练构建模型，再按照模式分类的思想评估安全态势。

模型中的评估/预测方法

分类方法2：2016-东南大学-软件学报-网络安全态势感知综述_龚俭

将网络安全态势分为态势觉察、态势理解和态势投射

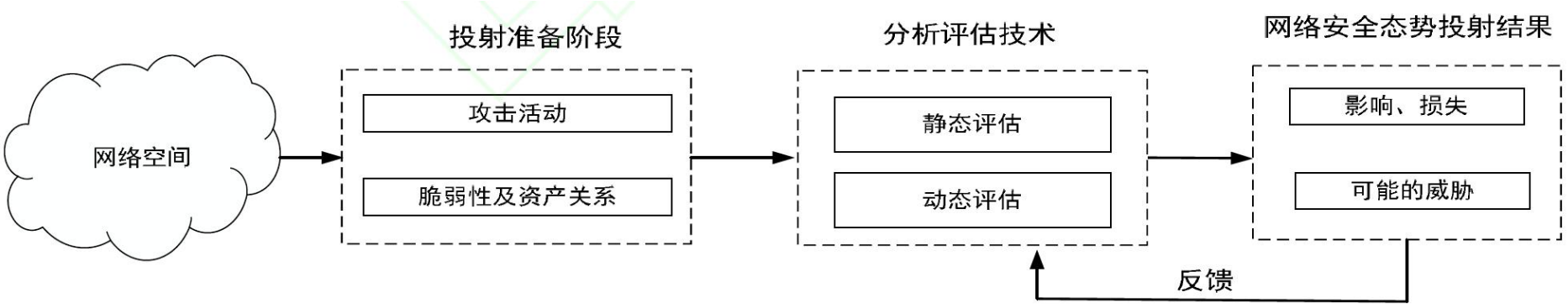
态势投射→态势评估

存在**静态评估**和**动态评估**两种风险评估技术

静态评估是指在攻击发生之前,主动地分析和评估被管系统中存在的风险和隐患,支持全面预防性的安全响应决策

动态评估是指在攻击发生之时,基于当前的安全警报进行实时评估和预判型评估,以支持有针对性的动态安全响应决策

研究方法大致分为三类:知识推理方法、统计方法和灰度理论方法



模型中的评估/预测方法

分类方法2：2016-东南大学-软件学报-网络安全态势感知综述_龚俭

研究方法大致分为三类:知识推理方法、统计方法和灰度理论方法

基于知识推理的方法：凭借专家知识及经验建立评估模型,通过逻辑推理分析整个网络的安全态势

- 相关方法有两类:一类是基于图模型的推理,如贝叶斯网络、模糊认知图(Fuzzy Cognitive Maps,简称 FCM)等;另一类是基于证据理论的推理,如D-S 证据推理.

统计分析的方法：综合考虑影响网络安全的态势要素,构建一个评估函数,实现态势要素和整个网络态势空间的映射

- 常用的统计方法有权重分析方法和层次分析法(Analytic Hierarchy Process AHP),该方法的关键是求得态势要素的重要性权值.

灰度理论方法：灰色系统理论以”部分信息已知,部分信息未知的小样本、贫信息”的不确定性系统作为研究对象,并在此基础上提取有用信息，安全态势风险值的变化作为“灰色系统”

- 灰度理论、马尔科夫理论、GM(1,1)模型

预测算法定义

什么是预测？

预测分为两种类型：

- 预测分类标签：与分类问题一样
- 预测连续数值：基于连续值函数预测（以下的预测只此类问题）

为了方便理解，把预测和分类看做两个问题。

预测与分类的区别：

Classification：预测分类类标签，离散数值或标签；分类过程基于训练集及训练集给出的分类属性。

Prediction：构建连续数值函数，预测未知或缺失的值。

态势感知中的预测

态势评估：定性、定量分析网络当前的安全状态和薄弱环节，并给出相应的应对措施，这一步是态势感知的核心；

态势预测：通过对态势评估输出的数据，预测网络安全状况的发展趋势，这一步是态势感知的目标。需要注意的是态势评估输出的数据都是以时间为自变量的数据，因此态势预测主要针对**时序数值型连续随机变量**的预测问题。

模型中的评估/预测方法

- 态势评估方法
  - ✓ 基于逻辑的方法：依据信息之间的内在逻辑，对信息进行融合。 其中， 融合的数据源为单一数据源，数据获取存在不确定性; 逻辑关系的获取存在很大的难度， 不适用于存在不确定性数据的网络系统。
  - ✓ 基于数学模型的方法：根据各项态势因素构造评定函数，建立态势因素集合到态势状态的映射。 但是， 权值的选择没有统一的标准；需要准确的数据源，即先验数据。但先验数据有时存在不完整性、不确定性。
  - ✓ 基于概率统计的方法：需要依赖一个较大的数据源；模型需要的存储量和匹配计算的运算量相对较大,影响态势评估的实时性; 特征提取模型构建和先验知识的获取都存在一定的困难；
  - ✓ 基于知识推理的方法：通过模糊量化多源多属性信息的不确定性，利用规则进行逻辑推理，实现网络安全态势的评估。 计算复杂度高；而且当证据出现冲突时,方法的准确性无法保证。
- 态势预测方法
  - ✓ 时间序列预测法： 对高精度时序模型的建立，依赖模型参数的最佳估计；模型的阶数要合适且建模过程复杂；对具有非线性关系、非正态分布特性的宏观网络态势值所形成的时间序列数据，处理效果不理想。
  - ✓ 机器学习深度学习分类： 对结果难以提供可信的解释；训练时间长，过度拟合或者训练不足。依靠经验风险最小化原则，易导致泛化能力下降且模型结构难以确定；样本数量有限时，精度难以保证；样本数量很多时，模型维度过多，泛化能力不高。

提纲

1、态势评估和预测的方法体系

2、态势评估方法

3、传统时间序列预测算法

4、机器学习预测算法

2、态势评估方法

- 2.1 实际应用中的数学模型方法
- 2.2 层次分析方法
- 2.3 贝叶斯方法
- 2.4 D-S证据理论

2.1 实际应用中的数学模型方法

从宏观角度，研究报警的整体时序特征，试图寻找特定网络的受攻击特征，攻击爆发程度，周期性；
多采用网络流量分析技术，以发现流量的变化趋势和突变。

分类过滤


峰会官网防护监测
进流量 0.02 兆(Mb)
出流量 0.06 兆(Mb)
访问量 16 千次
攻击量 1,228 次
网站服务质量
0~0.3秒
0.3~1秒
1~10秒
>10秒
无法访问
网站漏洞分布
未发现安全问题
访问源区域排行
浙江 20.38%
美国 13.93%
上海 10.02%
北京 8.74%
日本 7.23%
韩国 4.16%
广东 4.02%
香港 3.32%
澳大利亚 2.36%
加拿大 1.57%
时间 源IP 区域 攻击URL 攻击方式 风险等级
13:59:28 112.25.132.10 江苏 g20.org/year/201512/t201… 服务器信息泄露 中
攻击网站URL排行
1 312 /news/
2 268 /robots.txt
3 250 /images/css.xxs
4 250 /images/jquery.min.xx
5 250 /images/style.xxs
6 250 /images/daimabiji.xx
7 250 /images/SuperSlide.2.1.xx
8 116 /English
9 108 /images/page.xx
10 88 /export/rss2/index.xml
访问源IP排行
美国 69.191.211.210 2308
美国 198.7.61.14 1602
韩国 211.45.102.36 930
上海 116.231.81.102 818
北京 219.143.74.42 718
法国 62.210.182.126 612
韩国 61.72.230.109 606
广东 121.34.187.221 540
韩国 112.223.116.28 512
浙江 60.190.225.138 456
网站攻击趋势
访问分析
流量分析
访问次数
攻击次数
次
1,000
800
600
400
200
0
00:00 01:16 02:32 03:48 05:04 06:20 07:36 08:52 10:08 11:24 12:40 13:56
/分钟
技术支持单位：杭州安恒信息技术有限公司
攻击源IP排行
法国 62.210.182.126 610
广东 121.34.187.221 334
重庆 113.250.170.83 224
上海 116.231.81.102 216
重庆 183.66.114.52 200
俄罗斯 194.28.238.89 192
重庆 222.179.204.119 184
香港 49.246.230.40 178
美国 40.77.167.79 142

统计分析


峰会官网防护监测
进流量 0.02 兆(Mb)
出流量 0.06 兆(Mb)
访问量 16 千次
攻击量 1,228 次
网站服务质量
0~0.3秒
0.3~1秒
1~10秒
>10秒
无法访问
网站漏洞分布
未发现安全问题
访问源区域排行
浙江 20.38%
美国 13.93%
上海 10.02%
北京 8.74%
日本 7.23%
韩国 4.16%
广东 4.02%
香港 3.32%
澳大利亚 2.36%
加拿大 1.57%
时间 源IP 区域 攻击URL 攻击方式 风险等级
13:59:28 112.25.132.10 江苏 g20.org/year/201512/t201… 服务器信息泄露 中
攻击网站URL排行
1 312 /news/
2 268 /robots.txt
3 250 /images/css.xxs
4 250 /images/jquery.min.xx
5 250 /images/style.xxs
6 250 /images/daimabiji.xx
7 250 /images/SuperSlide.2.1.xx
8 116 /English
9 108 /images/page.xx
10 88 /export/rss2/index.xml
访问源IP排行
美国 69.191.211.210 2308
美国 198.7.61.14 1602
韩国 211.45.102.36 930
上海 116.231.81.102 818
北京 219.143.74.42 718
法国 62.210.182.126 612
韩国 61.72.230.109 606
广东 121.34.187.221 540
韩国 112.223.116.28 512
浙江 60.190.225.138 456
网站攻击趋势
访问分析
流量分析
访问次数
攻击次数
次
1,000
800
600
400
200
0
00:00 01:16 02:32 03:48 05:04 06:20 07:36 08:52 10:08 11:24 12:40 13:56
技术支持单位：杭州安恒信息技术有限公司
攻击源IP排行
法国 62.210.182.126 610
广东 121.34.187.221 334
重庆 113.250.170.83 224
上海 116.231.81.102 216
重庆 183.66.114.52 200
俄罗斯 194.28.238.89 192
重庆 222.179.204.119 184
香港 49.246.230.40 178
美国 40.77.167.79 142

TopN排序


峰会官网防护监测
进流量 0.02 兆(Mb)
出流量 0.06 兆(Mb)
访问量 16 千次
攻击量 1,228 次
网站服务质量
0~0.3秒
0.3~1秒
1~10秒
>10秒
无法访问
网站漏洞分布
未发现安全问题
访问源区域排行
浙江 20.38%
美国 13.93%
上海 10.02%
北京 8.74%
日本 7.23%
韩国 4.16%
广东 4.02%
香港 3.32%
澳大利亚 2.36%
加拿大 1.57%
时间 源IP 区域 攻击URL 攻击方式 风险等级
13:59:28 112.25.132.10 江苏 g20.org/year/201512/t201… 服务器信息泄露 中
攻击网站URL排行
1 312 /news/
2 268 /robots.txt
3 250 /images/css.xxs
4 250 /images/jquery.min.xx
5 250 /images/style.xxs
6 250 /images/daimabiji.xx
7 250 /images/SuperSlide.2.1.xx
8 116 /English
9 108 /images/page.xx
10 88 /export/rss2/index.xml
访问源IP排行
美国 69.191.211.210 2308
美国 198.7.61.14 1602
韩国 102.36 930
上海 116.231.81.102 818
北京 219.143.74.42 718
法国 62.210.182.126 612
韩国 61.72.230.109 606
广东 121.34.187.221 540
韩国 112.223.116.28 512
浙江 60.190.225.138 456
网站攻击趋势
访问分析
流量分析
访问次数
攻击次数
次
1,000
800
600
400
200
0
00:00 01:16 02:32 03:48 05:04 06:20 07:36 08:52 10:08 11:24 12:40 13:56
/分钟
技术支持单位：杭州安恒信息技术有限公司
攻击源IP排行
法国 62.210.182.126 610
广东 121.34.187.221 334
重庆 113.250.170.83 224
上海 116.231.81.102 216
重庆 183.66.114.52 200
俄罗斯 194.28.238.89 192
重庆 222.179.204.119 184
香港 49.246.230.40 178
美国 40.77.167.79 142

模式匹配


木马
50%
感染式病毒
13%
蠕虫
6%
黑客工具
1%
灰色软件
18%
风险软件
11%
136，236
116，148
捕获样本
境内捕获
放马URL
197，448
354，840
6，321，336
恶意代码
放马IP
53，616
42，264
广州市
绍兴市
北京市
湛江市
杭州市
宁波市
合肥市
上海市
合计
2013-03
2013-04
2013-05
2013-06
2013-07
2013-04
2013-05
2013-06
2013-07

2.1 实际应用中的数学模型方法

上述的方法原理都比较简单，实施起来也比较容易，而且效率也基本能满足实时监控的要求，因而被广泛使用。

但是，网络安全本身就具有突发性和快速变化的特点使这些方法的准确性较低，并不能给出深层次的结果，不能满足多种目的的需要。

2.2 层次分析方法

层次分析法（The Analytic Hierarchy Pricess，以下简称 AHP）是由美国运筹学家、匹兹堡大学萨第（T.L.Saaty）教授于本世纪 70 年代提出的。于 1977 年在国际数学建模会议上发表了“无结构决策问题的建模—层次分析法”一文，此后 AHP 在决策问题的许多领域得到应用。

主要思路：层析分析法首先将问题层次化，根据问题的性质和要求达到的总目标，将问题分解为不同的组成因素，并按照因素之间的相互关联影响以及隶属关系将因素按不同的层次聚集组合，形成一个多层次的分析结构模型，并最终把系统分析归结为最低层（供决策的方案、措施等），相对于最高层（总目标）的相对重要性权值的确定或相对优劣次序的排序问题。

方法：层次分析法引入 1—9比率标度方法，并写成矩阵的形式，即构成所谓的判断矩阵。形成矩阵后，即可以通过计算判断矩阵的最大特征根及对应的特征向量，计算出某一层元素相对于上一层次某一个元素的相对重要性权值。在计算出某一个层次相对于上一层次各个因素的单排序权值后，用上一层次因素本身的权值加权综合，即可以计算出某层因素相对去上一层整个层次的相对重要性权值，即层次总排序的权值。

2.2 层次分析方法

该方法把复杂的问题分解为若干层次，按自下而上、先局部后整体的策略逐层计算权数。该方法将定性分析与定量分析相结合,是一种无结构的多准则决策方法,通过思维过程的层次化和数量化,达到分析复杂问题的目的。**态势评估中的应用：网络中各种要素的重要程度，或者说各种要素被用户的关注程度，也就是各种要素在网络态势计算中，反映出来的其权值的大小，这是一个元素排序的问题。**

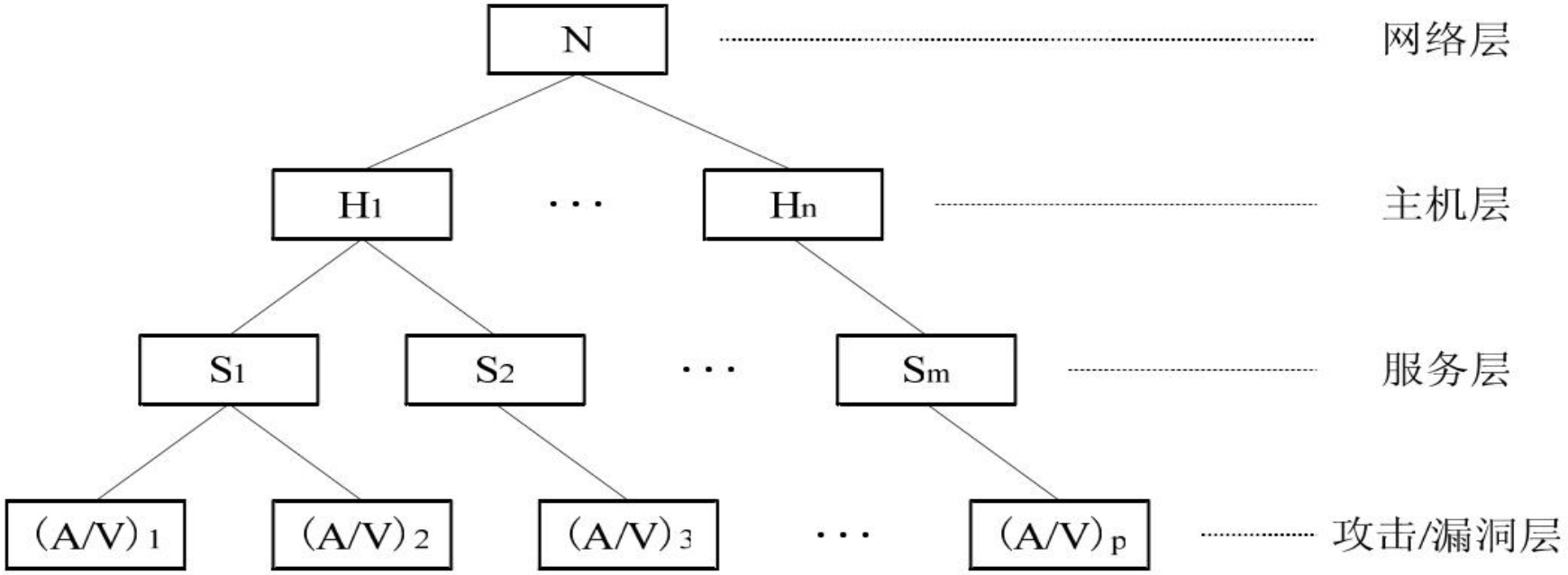


图 2.6 层次化网络安全态势评估模型

2.2 层次分析方法

态势评估分析指标的选取

网络安全态势指标的选取需综合考虑不同层次(宏观网络、局部网络、主机、服务、攻击/漏洞), 不同信息来源(流量、报警、日志、静态配置) 和不同需求(普通用户、管理者、维护者

三个方面的指标体系：

（1）不同层次的网络安全指标

（2）针对不同使用对象的网络安全指标

（3）使用不同检测/监控手段的网络安全指

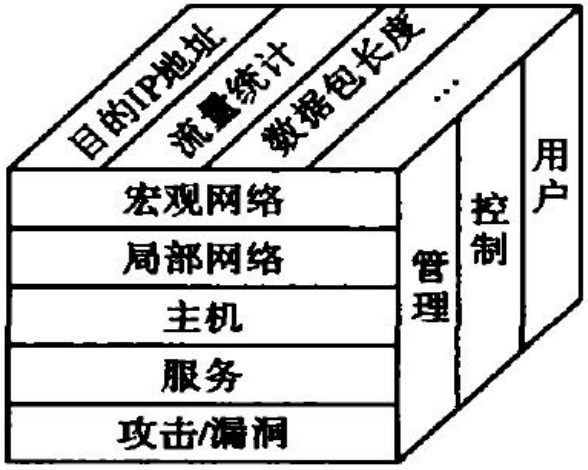


图 2　宏观网络安全指标选取

表 1　网络安全态势评估指标体系

二级指标	一级指标
脆弱性	网络漏洞数目及等级 关键设备漏洞数目及等级 子网内安全设备数目 子网内各关键设备提供的服务种类及其版本 子网内各关键设备的操作系统类型及其版本 子网内各关键设备开放端口的总量 网络拓扑
容灾性	网络带宽 子网内安全设备数目 子网内各关键设备的操作系统类型及其版本 子网内各关键设备访问主流安全网站的频率 子网内各关键设备提供的服务种类及其版本 网络拓扑 子网内主要服务器支持的并发线程数
威胁性	报警数目 子网带宽使用率 子网内安全事件历史发生频率 子网内各关键设备提供的服务种类及其版本 子网数据流入量 子网流入量增长率 子网内不同协议数据包的分布 子网内不同大小数据包的分布 流入子网内数据包源 IP分布
稳定性	子网内关键设备平均存活时间 子网流量变化率 子网内不同协议数据包分布比值的变化率 子网内不同大小数据包分布比值的变化率 子网数据流总量 流出子网数据包目的 IP的分布 子网内存活关键设备数目 子网平均无故障时间

2.3 贝叶斯方法

贝叶斯推理法：贝叶斯推理在给定一先验似然估计和附加证据条件下，能更新一个假设的似然函数。贝叶斯网络是概率分析和图论结合的产物，它是一种有向图模型，用于不确定性知识的表达和推理。简单来说，贝叶斯网络表现为一个赋值的因果关系网络图。

主要思路：就是进行概率推理，给定贝叶斯网络的所有事件（节点）联合概率分布，理论上能够回答所有的推理问题。

方法：在网络安全态势评估中引入贝叶斯网络时，必须合理构造贝叶斯网络，适当简化网络结构，降低计算难度。主要有三种推理防范：

（1）**因果推理**，由原因推出结果；

（2）**诊断推理**，由结果推出原因；

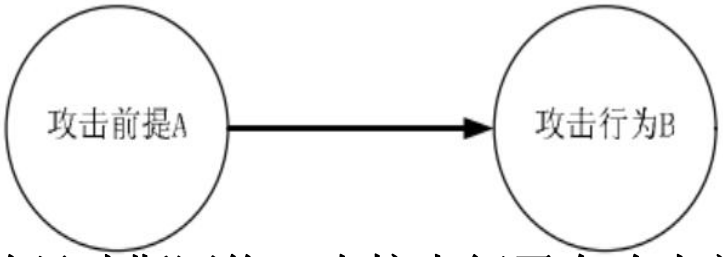
（3）支持推理，提供解释支持现象。

因果推理：主要可以帮助解决一些已知了一些安全态势数据和对于发生哪类威胁的概率关系，当收集到多种不同数据的时候，通过推理判断发生某个威胁的可能有多大。

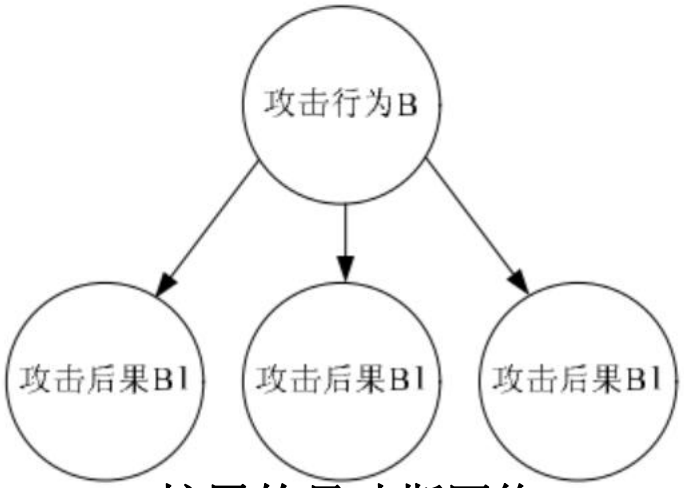
诊断推理：在已知一些安全隐患可能有什么安全问题引起的概率时，结合系统发生一定的安全报警，通过反推的方法了解到问题可能由什么隐患引起的。

2.3 贝叶斯方法-推理模型

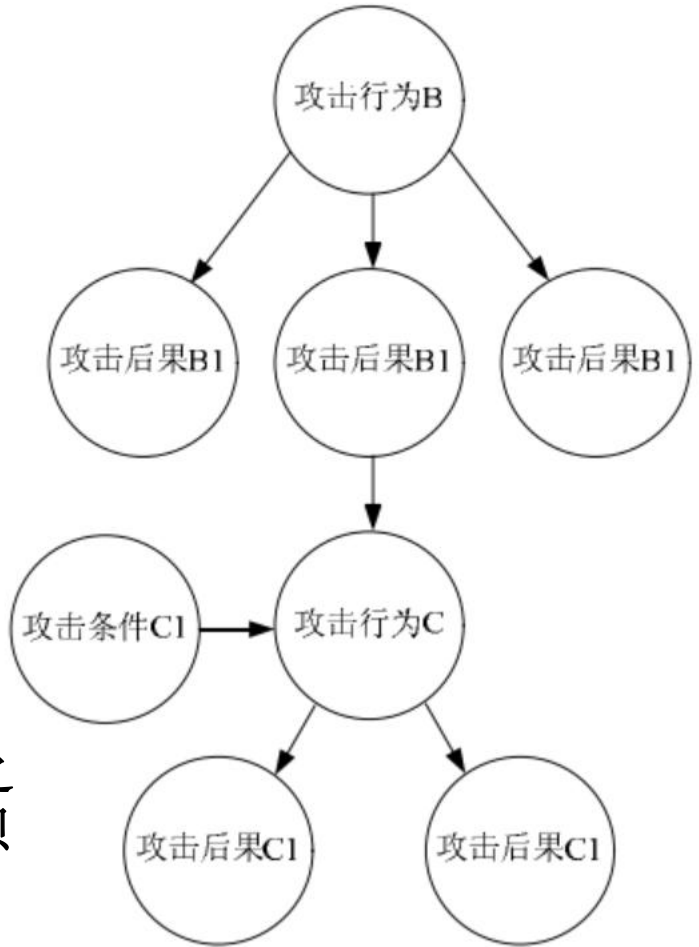
采用贝叶斯法对网络态势评估有很好的实用价值，可以对整体网络安全态势的等级进行划分，也可以对具体某类态势进行判断有很好的作用。



简单的贝叶斯网络：直接表征了在攻击前提 A 下攻击行为 B 发生的推理过程。产生攻击行为 B 的安全隐患与攻击前提 A 的概率情况有着直接关系。



扩展的贝叶斯网络

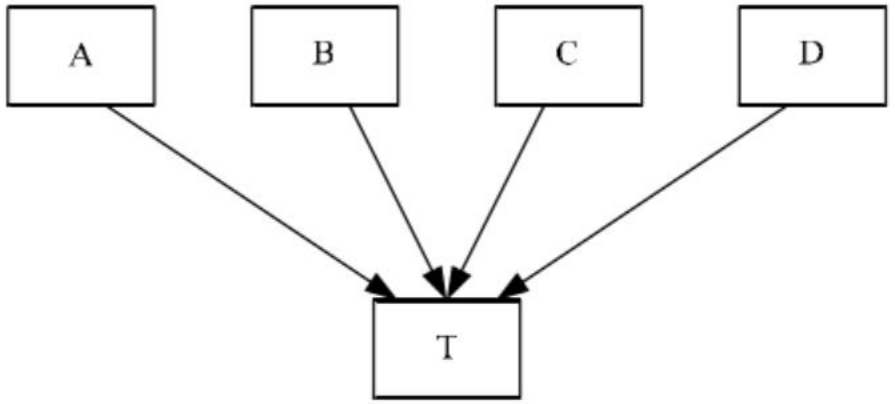


攻击行为之间关联的贝叶斯网络

2.3 贝叶斯方法-态势计算中的应用

根据不同的评估指标，构建贝叶斯网络：

A.综合安全态势
B.主机安全态势
C.攻击威胁安全态势
D.病毒疫情安全态势
T.网络安全态势等级



（1）通过测试数据训练，确定每个指标关于安全态势等级的条件概率。
（2）贝叶斯网络的拓扑结构和条件概率密度 CPD 定义好以后，接着训练每一个CPD 的参数（根节点不做训练），确定了参数就确定了各个节点的 CPD，也就可以对网络开始进行安全评估。
（3）可以采用最大似然估计(Maximum Likelihood Estimation，MLE)来计算（训练）节点的 CPD 参数。

2.4 D-S证据理论

Dempster

1. D-S证据理论

DS 证据理论（Dempster-Shafer envidence theory）也称为DS理论。**是进行不确定性推理的重要方法**。它不仅能够强调事物的客观性，还能强调**人类对事物估计的主观性**。

2. 优点：

D-S 理论处理不确定性和无知性之间的区别，它并不计算一个命题的概率，而是计算证据可能支持命题的概率，给出信息的信任测度和似然测度，具有一定程度的怀疑能力。

3. 缺点：

- 证据需要是独立的（有时候不容易满足）；
- 证据合成理论没有坚固的理论基础，合理性和有效性争议大；
- 计算上存在潜在的指数爆炸。

不过以上缺点都有相关的理论去改善原来的模型理论，提高普适性。

2.4 D-S证据理论

- 其核心是D-S 证据合成规则，将来自两个独立信息源的证据组合为一个新的证据，并且可以推广到多个证据的情况。
- 主要特点：满足比贝叶斯概率论更弱的条件。
- 将D-S 理论引入计算机领域进行网络态势评估，其分析过程如下：
- 首先建立证据和命题之间的逻辑关系，即实体、态势因子到态势状态的汇聚方式，确定基本概率分配；
- 然后根据到来的证据，即每一则事件发生的上报信息，使用证据合成规则进行证据合成，得到新的基本概率分配，并把合成后的结果送到决策逻辑进行判断，将具有最大置信度的命题作为备选命题。
- 当不断有事件发生时，这个过程便得以继续，直到备选命题的置信度超过一定的阈值，证据达到要求，即认为该命题成立，态势呈现某种状态。

2.4 D-S证据理论

- 使用D-S 理论进行态势评估，克服了用概率描述不确定性的不足，不需要精确了解概率分布，也不需要显式表示不确定性。
- 通过建立命题和集合之间的对应关系，把命题的不确定性问题转化为集合的不确定性问题，给出信息的信任测度和似然测度。
- 当先验概率很难获得时，D-S 理论较概率论更为有效。
- 使用D-S 理论的另一大优点就是形式灵活多变，可与模糊逻辑、神经网络、专家系统相结合，进一步提高推理的准确性。
- 该方法的缺点是计算复杂度高。

提纲

1、态势评估和预测的方法体系

2、态势评估方法

3、时间序列预测算法

4、机器学习预测算法

3、 时间序列预测算法

- 时间序列预测法是将预测目标的历史数据按时间的顺序排列成为时间序列，然后分析它随时间的变化趋势，外推预测目标的未来值。
- 也就是说，时间序列预测法将影响预测目标的一切因素都由“时间”综合来描述。
- 因此时间序列预测法主要用于分析影响事物的主要因素比较困难或相关变量资料难以得到的情况。
- 时间序列预测法可分为确定性时间序列预测法和随机时间序列预测法。

3、 时间序列预测算法

- 现实中的时间序列的变化受许多因素的影响。
- 有些起着长期的、决定性的作用，使时间序列的变化呈现出某种趋势和一定的规律性。
- 有些则起着短期的、非决定性的作用，使时间序列的变化呈现出某种不规则性。
- 时间序列分析具有以下三种变化形式：

  趋势变动：指现象随时间变化朝着一定方向呈现出持续稳定地上升、下降或平稳的趋势。

  周期变动：指现象受季节影响，按某固定周期呈现出的周期波动变化。

  随机变动：指现象受偶然因素的影响而呈现出的不规则波动。

3、 时间序列预测算法

- 网络态势预测可视为一个时间序列分析问题。
- 根据系统观测得到的时间序列数据，通过曲线拟合和参数估计来建立数学模型。
- 假定时间序列 $x=\{x_i \mid x_i \in U, i=1,2,\cdots,L\}$，网络态势预测就是通过序列的前$R$ 个时刻的态势值，预测出后M 个时刻的值，实现从输入U^R 到输出U^M 的映射，即

$$\left(x_{R+1}, x_{R+2}, \cdots, x_{R+M}\right)=f\left(x_1, x_2, \cdots, x_R\right)$$

从而达到时间序列预测的目的。

3、 时间序列预测算法

- 马尔科夫与隐马尔科夫预测算法
- ARIMA

3.1 马尔科夫与隐马尔科夫预测算法

1. 马尔科夫预测算法

基本原则：对于一个系统，由一个状态转至另一个状态的转换过程中，存在着转移概率，并且这种转移概率可以依据其紧接的前一种状态推算出来，与该系统的原始状态和此次转移前的过程无关。一系列的马尔可夫过程的整体称为马尔可夫链。

转移概率矩阵：从一个状态转换到另一个状态的可能性，我们称之为状态转移概率。所有状态转移概率的排列即是转移概率矩阵。

2. 马尔可夫分析的基本假定：

预测期系统状态数保持不变

系统状态转移概率矩阵不随时间变化

状态转移仅受前一状态影响，即马尔可夫过程的无后效性

3. 马尔可夫过程用于预测基本步骤：

首先确定系统状态，然后确定状态之间转移概率，再进行预测，并对预测结果进行分析-若结果合理，则可提交预测报告，否则需检查系统状态及状态转移概率是否正确。

3.1 马尔科夫与隐马尔科夫预测算法

案例一：马尔科夫预测攻击变化情况

第一步，要调查不同攻击行为占比情况，得到整体攻击行为占有率向量A。通过对系统状况调查，得出目前系统攻击者的行为比例向量 A=(0.2, 0.5, 0.3)，即目前，在系统的10个攻击中，第一类攻击占有20%，第二类攻击占有50%，第三类攻击占有30%。

3.1 马尔科夫与隐马尔科夫预测算法

案例一：马尔科夫预测攻击变化情况

第二步，调查攻击者行为变动情况，得出整个攻击行为的转移概率矩阵B，经过调查，攻击者下一步行为，在现在采用第一类攻击行为中进行调查，下一步仍然采用第一类攻击的占有50%，40%采用第二类攻击，10%将采用第三类攻击；在现在采用第二类攻击行为中进行调查，下一步采用第一类攻击的占有20%，采用第二类攻击的占有50%，30%将采用第三类攻击；在现在采用第三类攻击中进行调查，下一步采用第一类攻击的占有30%，而有40%将采用第二类攻击。据此得出下一步整个攻击发布情况的转移概率矩阵。

$$\begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

3.1 马尔科夫与隐马尔科夫预测算法

案例一：马尔科夫预测攻击变化情况

第三步，用向量S乘以矩阵B即可得出下一步三类攻击的占有率分别是29%、45%和26%。

第四步,若这种变化成为相对稳定状况，也即转移概率矩阵将对攻击占有率不起变动作用，我们就可以计算出攻击行为相对稳定之后的攻击占有率。设x = (x1, x2, X3)是稳定以后的市场占有率，则x不随时间的推移而变化，也即市场占有率处于动态平衡，即有xB=x，详细写出来即为

$$\text{即}(x_1 x_2 X_3)\begin{bmatrix}0.5 & 0.4 & 0.1\\ 0.2 & 0.5 & 0.3\\ 0.3 & 0.4 & 0.3\end{bmatrix} = (x_1 x_2 X_3)$$

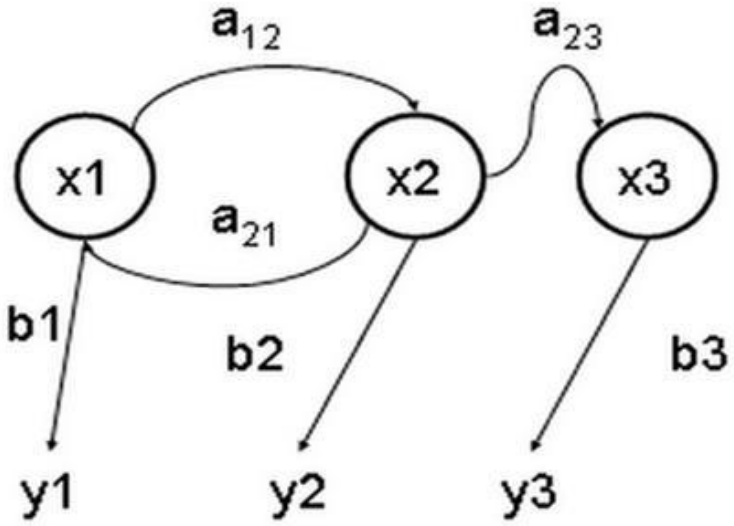
可以联立方程组：

$$\begin{cases}0.5X_1 + 0.2X_2 + 0.3X_3 = X_1\\ 0.4X_1 + 0.5X_2 + 0.4X_3 = X_2\\ 0.1X_1 + 0.3X_2 + 0.3X_3 = X_3\\ X_1 + X_2 + X_3 = 1\end{cases}$$

解方程组得x1 = 32.92,x2 = 36.36,x3 = 30.69，攻击行为在系统状况为稳定情况下的比例为第一类攻击32.92%，第二类攻击36.36%，第三类攻击30.69%。

3.1 马尔科夫与隐马尔科夫预测算法

隐马尔可夫模型 (Hidden Markov Model)：是一种统计模型，用来描述一个含有隐含未知参数的**马尔可夫过程**。在正常的**马尔可夫模型**中，状态对于观察者来说是直接可见的。这样状态的转换概率便是全部的参数。而在**隐马尔可夫模型**中，状态并不是直接可见的，但受状态影响的某些变量则是可见的。每一个状态在可能输出的符号上都有一概率分布。因此输出符号的序列能够透露出状态序列的一些信息。具体的，下图是一个三个状态的**隐马尔可夫模型**状态转移图，其中x 表示隐含状态，y 表示可观察的输出，a 表示状态转换概率，b 表示输出概率。隐藏的状态和可观察到的状态之间有一种概率上的关系，也就是说某种隐藏状态 X 被认为是某个可以观察的状态 y_1 是有概率的，假设为 $P(y_1 \mid X)$。如果可以观察的状态有3种，那么很显然 $\mathbf{P(y_1 \mid X)+P(y_2 \mid X)+ P(y_3 \mid X) = 1}$。



3.2 ARIMA时序分析算法

- 典型方法：基于时间序列分析的网络安全态势预测算法。
- ARIMA （Autoregressive Integrated Moving Average Model）模型的全称叫做自回归移动平均模型
- 是统计模型(statistic model)中最常见的一种用来进行时间序列预测的模型。
- 对利用网络安全态势量化评估算法计算得到的多个时段的网络安全态势值样本进行时间序列分析，从而实现对未来网络安全态势发展趋势的预测。
- 时间序列分析是根据系统观测得到的时间序列数据，通过曲线拟合和参数估计来建立数学模型的理论和方法。

3.2 ARIMA时序分析算法

- 所谓时间序列(Time Series)就是按照时间顺序记录的一列有序数据。
- 对时间序列进行观察、研究，找寻它变化发展的规律，预测它将来的走势就是时间序列分析。
- 时间序列预测方法是通过时间序列的历史数据揭示现象随时间变化的规律，将这种规律延伸到未来，从而对该现象的未来做出预测。

3.2 ARIMA时序分析算法

- 时间序列分析方法最早起源于1927年，数学家耶尔(Yule)提出建立自回归(AR)模型来预测市场变化的规律。
- 1931年，数学家瓦尔格(Walker)在AR模型的启发下，建立了移动平均(MA)模型和自回归移动平均(ARMA)模型，初步奠定了时间序列分析方法的基础，当时主要应用在经济分析和市场预测领域。
- 20世纪60年代，时间序列分析理论和方法迈入了一个新的阶段，伯格(Burg)在分析地震信号时最早提出了最大熵谱(MES)估计理论，后来有人证明AR模型的功率谱估计与最大熵谱估计是等效的，并称之为现代谱估计。
- 它克服了用传统的傅里叶功率谱分析所带来的分辨率不高和频率泄漏严重等固有的缺点，从而使时间序列分析方法不仅在时间域内得到应用，而且扩展到频率域内，特别是在各种工程领域内应用功率谱的概念更加方便和普及。

3.2 ARIMA时序分析算法

- 到20世纪70年代，美国统计学家博克斯和英国统计学家詹金斯提出来了Box．Jenkins方法，是迄今最通用的时间序列预测方法。
- 随着信号处理技术的发展，时间序列分析方法不仅在理论上更趋完善，尤其是在参数估计算法，定阶方法及建模过程等方面都得到了许多改进，进一步地迈向实用化，各种时间序列分析软件也不断涌现，逐渐成为分析随机数据序列不可缺少的有效工具之一。
- 随着时间序列分析方法的日趋成熟，其应用领域越来越广泛，主要集中在预报预测领域，例如气象预报、市场预测、地震预报、人口预测、汛情预报、产量预测、安全检测和质量控制等等。

3.2 ARIMA时序分析算法

- 基本思想

将预测对象随时间推移而形成的数据序列视为一个随机序列，即除去个别的因偶然因素引起的观测值外，时间序列是一组依赖于时间的随机变量，这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性，而这种自相关性一旦被相应的数学模型描述，就可以从时间序列的过去值和现在值预测未来值。

3.2 ARIMA时序分析算法

- 参考Box—Jenkins的建模方法，时间序列预测建模的一般过程主要包括如下几个步骤：

  序列检验(检验平稳性和纯随机性)

  模型识别

  参数估计

  模型检验

  预测

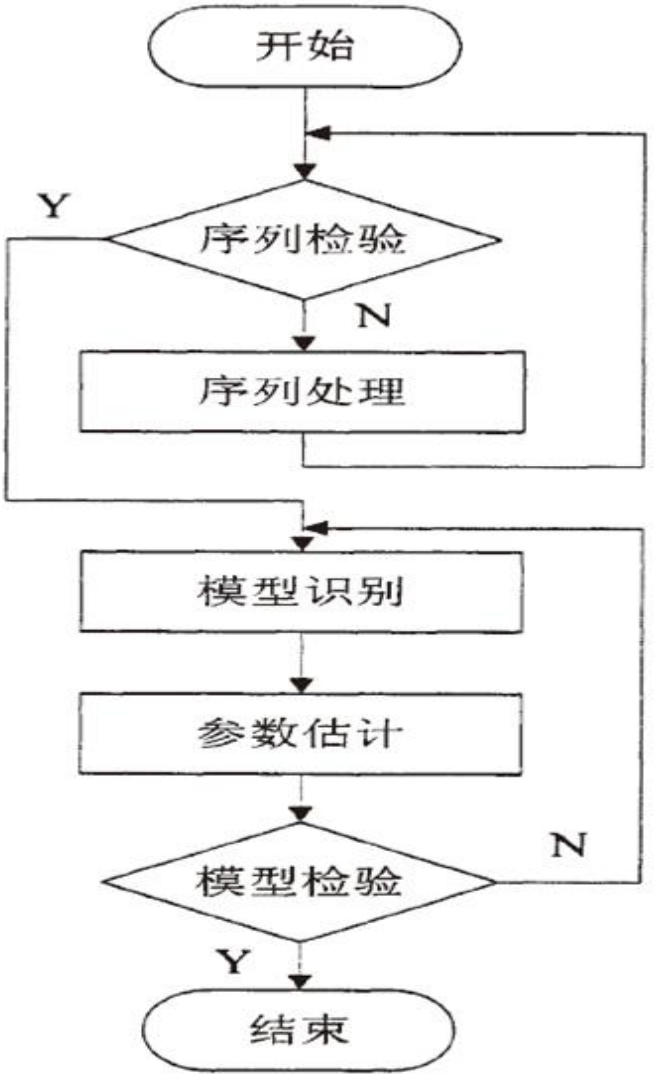


图 4.1 时间序列流程图

3.2 ARIMA时序分析算法

ARIMA模型运用的基本流程有几下几步：

1. 数据可视化，识别平稳性。
2. 对非平稳的时间序列数据，做差分，得到平稳序列。
3. 建立合适的模型。

   平稳化处理后，若偏自相关函数是截尾的，而自相关函数是拖尾的，则建立AR模型；

   若偏自相关函数是拖尾的，而自相关函数是截尾的，则建立MA模型；

   若偏自相关函数和自相关函数均是拖尾的，则序列适合ARMA模型。
4. 模型的阶数在确定之后，对ARMA模型进行参数估计，比较常用是最小二乘法进行参数估计。
5. 假设检验，判断（诊断）残差序列是否为白噪声序列。
6. 利用已通过检验的模型进行预测。

3.2 ARIMA时序分析算法

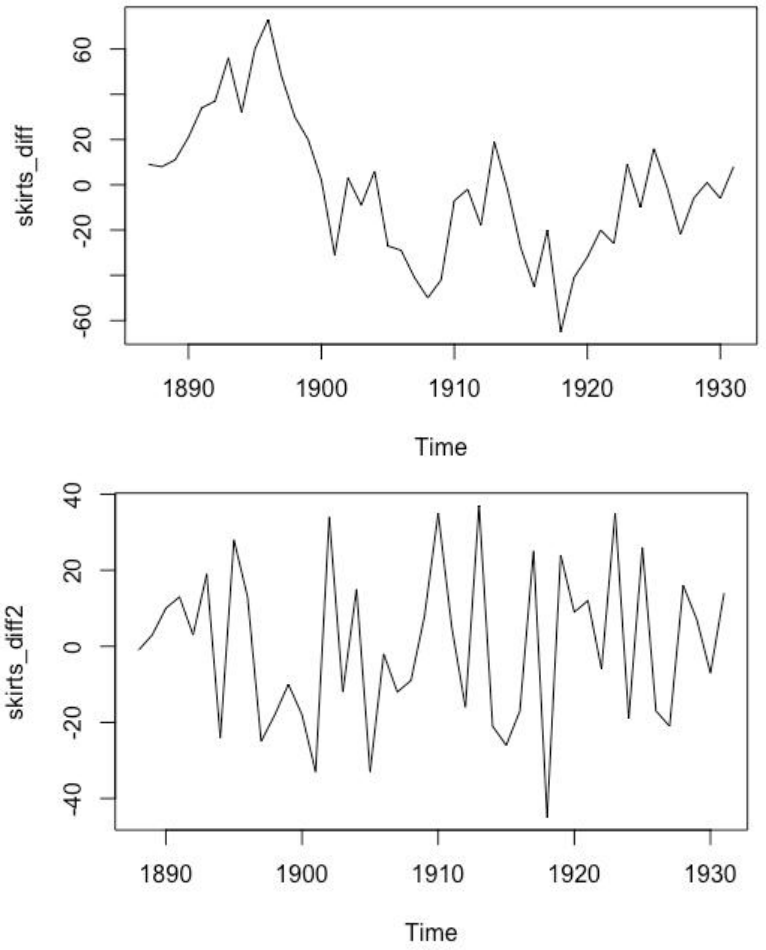
平稳性检验

- 采用ARIMA模型预测时序数据，必须是稳定的（平稳性），如果不稳定的数据，是无法捕捉到规律的。平稳性要求序列的均值、方差和协方差不发生明显变化，
- 可以通过看图检验平稳性，也可以通过单位根（unit root）检验。单位根检验是指检验序列中是否存在单位根，因为存在单位根就是非平稳时间序列了。
- 而Augmented Dickey-Fuller test可以测试一个自回归模型是否存在单位根。通过ADF检测的ADF值和p_value值看是否满足平稳性要求。
- 1）ADF值越小那么越拒绝原假设，越说明序列不存在单位根，那么时间序列越平稳。
- 2）p_value值要小于一个显著值，时间序列就是平稳的。一般以0.01为显著值。

3.2 ARIMA时序分析算法

参数d的确认

- 参数d的确认
- d就是差分的阶数，首先通过ADF检验，看原时间序列的平稳性，如果原时间序列是平稳的，那么d=0；如果原数据不平稳，那么做差分，通过ADF检验直到时间序列平稳。一般差分次数不超过2次。
- 从一阶差分的图中可以看出，数据仍是不平稳的，继续差分
- 二次差分后的时间序列在均值和方差上看起来是平稳了



3.2 ARIMA时序分析算法

参数估计

➢ARIMA（p，q）中，AR是“自回归”，p为自回归项数；MA为“滑动平均”，q为滑动平均项数

➢使用auto.arima()函数，自动获取最佳的ARIMA模型

- AR(p)模型

描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测

- MA(q)模型

移动平均模型关注的是自回归模型中的误差项的累加

ARMA(p,q)

自回归模型和移动平均模型的结合

1、AR（p）(p 阶自回归模型)

$$x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t$$

其中 u_t 白噪声序列，δ 是常数（表示序列数据没有 0 均值化）

2、MA（q）（q 阶移动平均模型）

$$x_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

$$x_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) u_t = \Theta(L) u_t$$

其中 $\{u_t\}$ 是白噪声过程。

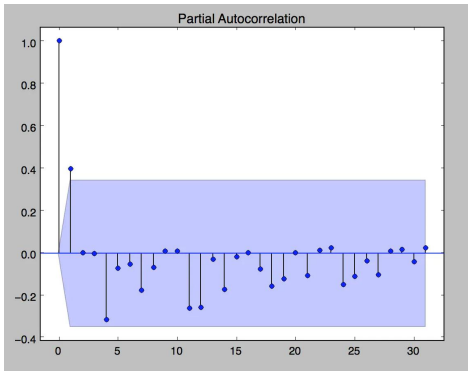
3.2 ARIMA时序分析算法

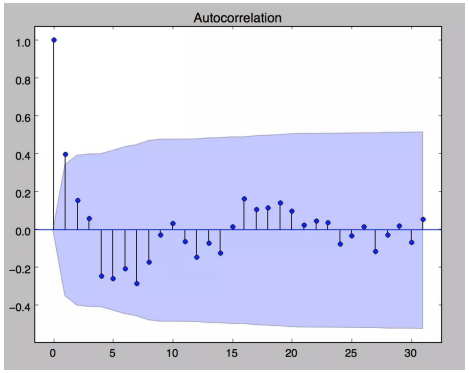
参数估计

➢ 检查平稳时间序列的自相关图和偏自相关图（ACF及PACF图）确定参数

- 观察法 ARMA（1，2）

模型	ACF	PACF
AR(p)	衰减趋于零（几何型或振荡型）	p阶后截尾
MA(q)	q阶后截尾	衰减趋于零（几何型或振荡型）
ARMA(p，q)	q阶后衰减趋于零（几何型或振荡型）	p阶后衰减趋于零（几何型或振荡型）



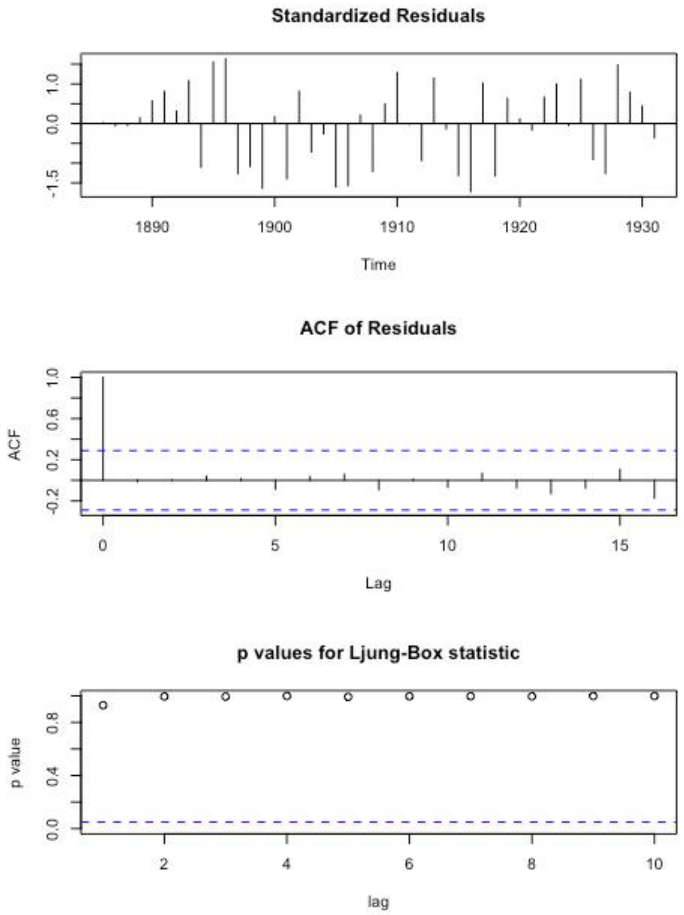


- AIC 和 BIC 值达到最小的那一组为理想阶数。

3.2 ARIMA时序分析算法

模型检验

- 观察ARIMA模型的残差是否是平均值为0且方差为常数。第一个图表代表估计模型**误差**的绘图。图中竖线的长度比较相似，都处在稳定范围之内，即估计的模型没产生不符合要求的误差分布。
- 第二张绘图，对残差进行**自相关性检测**。显示估计的模型没造成误差之间的任何关系，符合数据生成时每个数据都是独立的这个前提的。ACF图大部分都在置信空间内，部分超出也只超出一点点。
- 第三张图，Ljung-Box 指标。这个指标可对每一个时间序列的**延迟**进行显著性的评估。判定技巧是，P-value点的高度越高，模型越可信。Ljung-Box测试显示：所有的P-value>0.05，说明残差为高斯白噪声序列。



3.2 ARIMA时序分析算法

实例

- UDP flood， 攻击将尽可能多地向目标发送UDP数据报。 由于UDP中没有流量控制， 攻击者不仅会消耗目标的带宽， 而且会消耗CPU和内存的利用率，会导致对目标的拒绝服务。 为了检测这种攻击， ARIMA可以用来预测“正常”的交通模式。
- TCP SYN泛滥， 我们将看到不完整的TCP握手过程的数量增加。 因此， 通过测量SYN数据包的数量以及完成的TCP握手过程的数量， 它将提供可能攻击的指示。 在正常情况下， SYN数据包的数量肯定会与完成握手过程的数量相匹配。我们可以使用ARIMA来预测SYN数据包与完成TCP握手数量的正确比率。利用ARIMA进行的预测被认为是预期的预测正常数据。

3.2 ARIMA时序分析算法

采用ARIMA方法进行预测。
i. 实时捕获流量
ii. 根据捕获的流量创建统计数据
iii. 预测当前交通的正常模式
iv. 测量预测交通和实际交通之间的差异
v. 错误超过阈值时， 程序将发出警报。

从图中可以看出， 流量突然增加， 这可能意味着DoS攻击。

这种技术的一个已知限制是， 在每秒流量小于1MB的小容量网络环境中， 可能会产生误报。 最适合高交通量的数据中心环境。

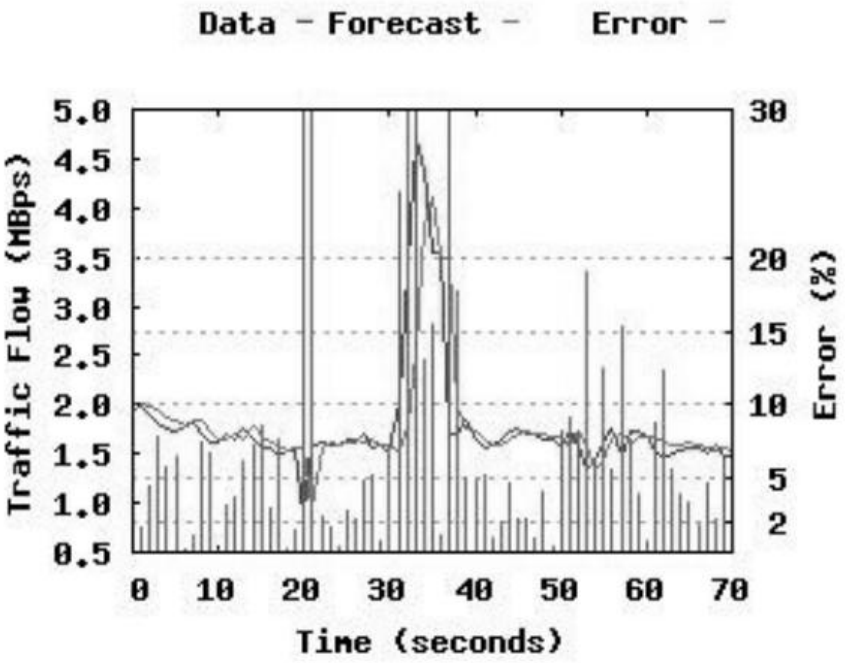


Figure 5. Output during a suddent increase in traffic (simulated DOS attack).

提纲

1、态势评估和预测的方法体系

2、态势评估方法

3、传统时间序列预测算法

4、机器学习预测算法

态势感知中的机器学习

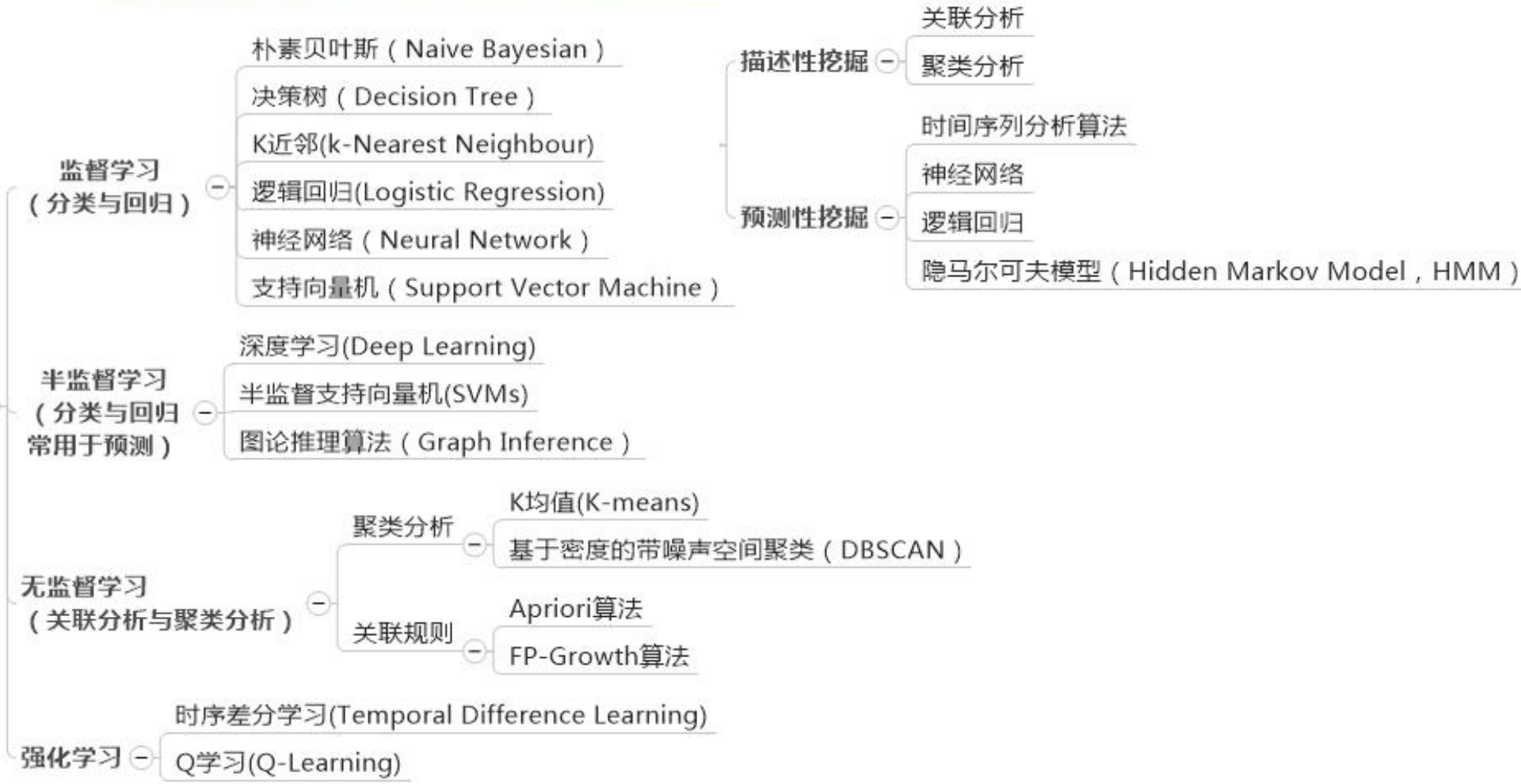
传统的数据分析耗费时间与精力多，非常依赖专家经验。

态势感知需要对海量数据快速完成聚类、分类、关联等处理。

机器学习能从部分数据中自动分析获取规律，并应用于未知数据。

利用(广义)机器学习进行态势感知中的**数据挖掘**

态势感知中的机器学习-典型算法



态势感知中的机器学习-典型研究

算法	思想	效果
人工神经网络	利用典型的网络攻击行为训练神经网络分类引擎，网络流量的特征向量被提交到分类引擎中进行研判。	可准确检出已知类型的攻击。泛化能力与精确性受样本质量影响大（样本多时，模型维度多但泛化能力低；样本少时，精确性低）。
支持向量机	由正常与异常序列训练得到支持向量与相关参数，并以之判别执行迹中的短序列是否异常。	从少量的正常与异常执行迹数据能得到较高检测率，预测效果较好。对复杂攻击(多分类问题）检测与训练效果一般。
遗传算法	分析网络数据集并通过遗传算法进行适应度评价生成规则集。	考虑的信息相对全面，可检测复杂的攻击。但易早熟，对新空间探索能力有限，效果不稳定。
时间序列分析法	通过序列的前N个时刻的态势值预测后M个态势值。	有较高的可操作性，但高精度的时序模型取决于参数的最佳估计，且建模相当复杂。对非线性关系、非正态分布的网络态势效果不理想。

降维算法：
主成份分析PCA；
深度学习DL。

交叉验证：
精确度precision；
召回率recall。

1.特征选择/降维处理

2.模型选择

3.模型验证/优化

4.模型部署

i.聚类：
K均值；
DBSCAN。

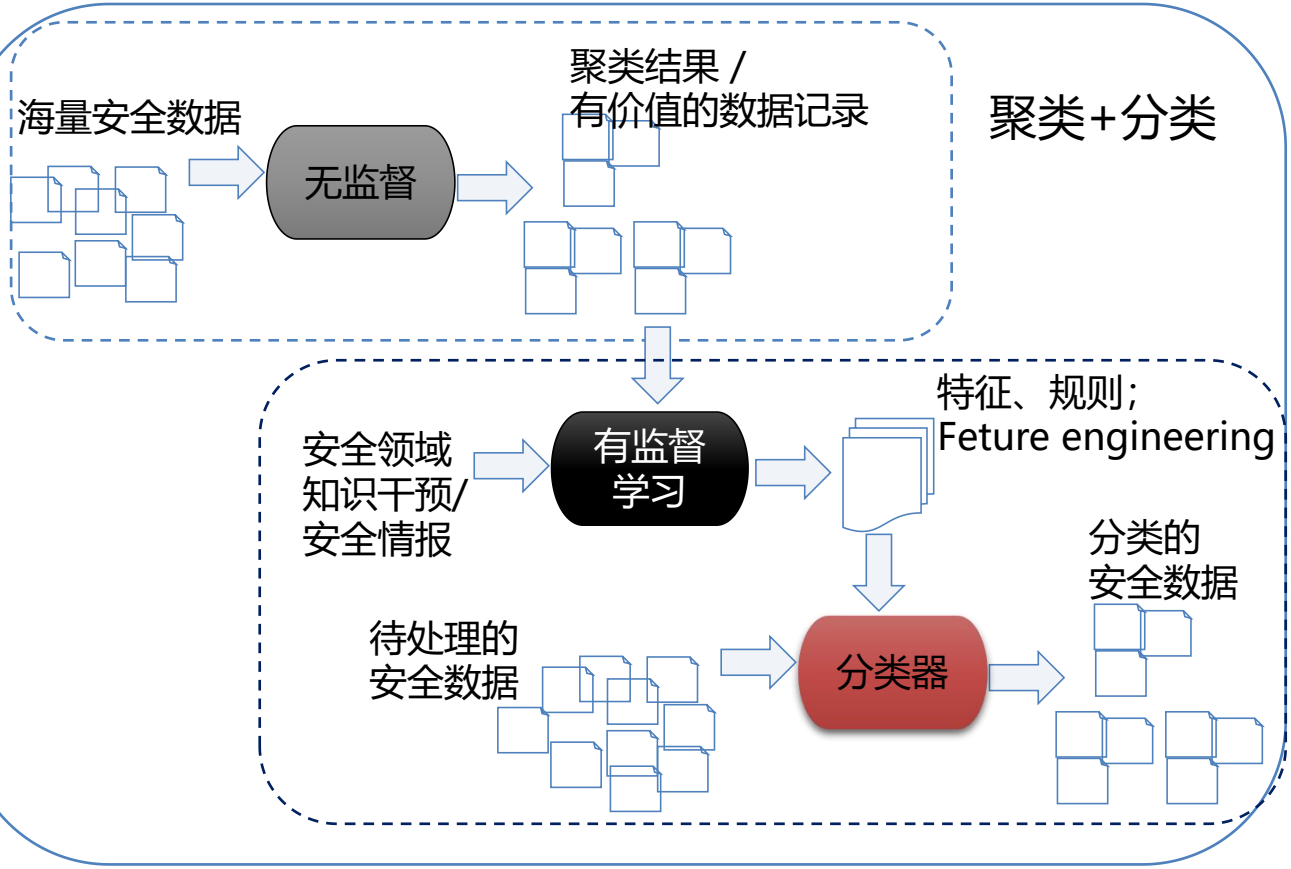
ii.分类：
决策树C4.5->随机森林；
朴素贝叶斯；
K-近邻；
神经网络；
支持向量机。

iii.关联：
Apriori算法；
FP-growth算法。

iv.预测：
时间序列；
逻辑回归->softmax算法；
HMM。

v.集成：
AdaBoost

聚类+分类



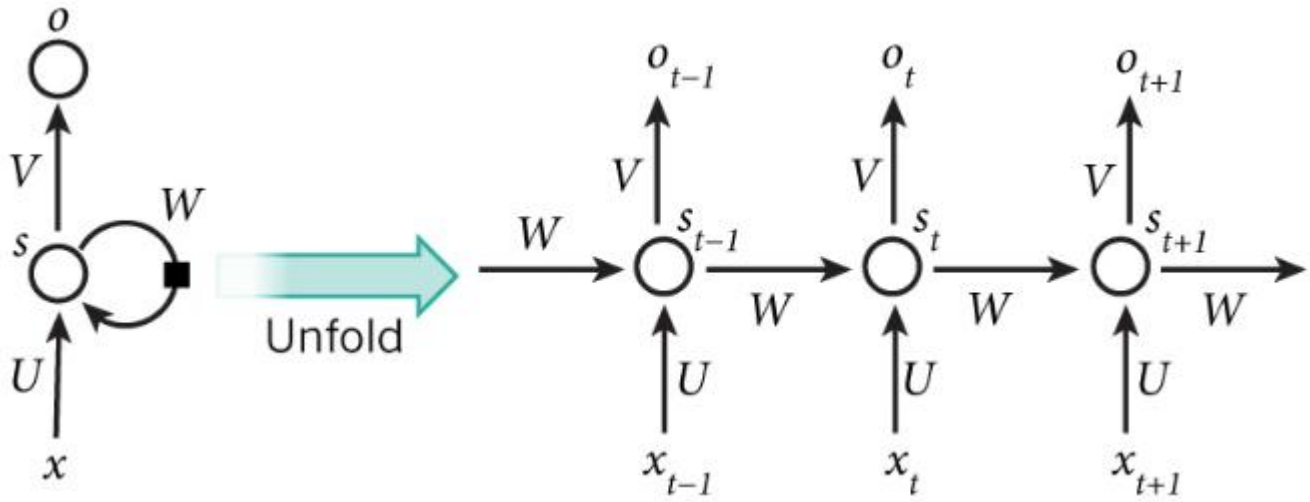
4、机器学习预测算法

- RNN与LSTM预测算法

4.1、RNN与LSTM预测算法

RNN 模型

目前时间序列模型最强大的的工具就是递归神经网络（recurrent neural network, RNN）。相比于普通神经网络的各计算结果之间相互独立的特点，RNN的每一次隐含层的计算结果都与当前输入以及上一次的隐含层结果相关。通过这种方法，RNN的计算结果便具备了记忆之前几次结果的特点。



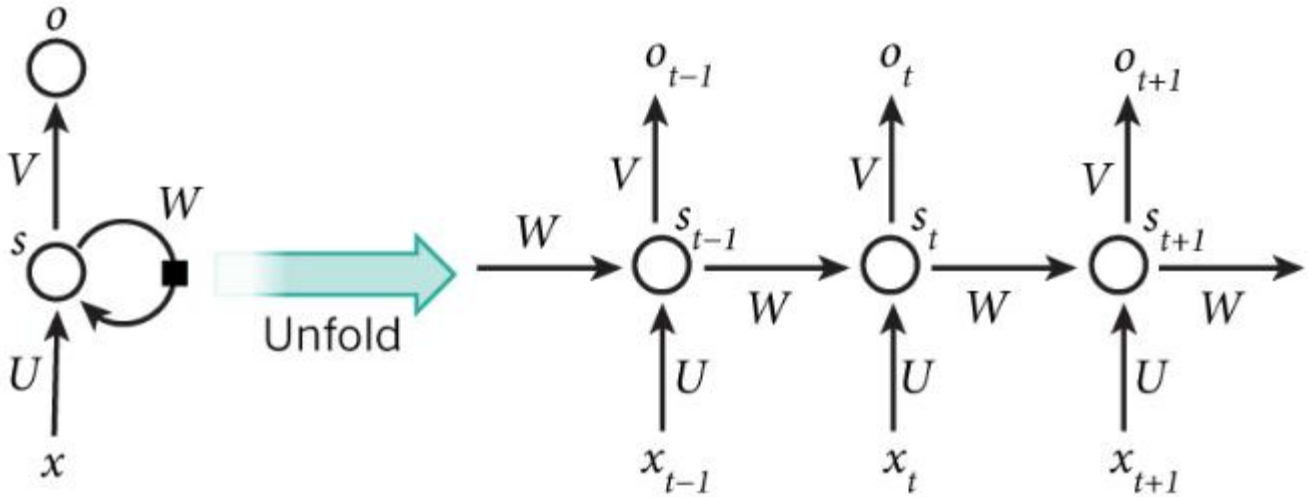
4.1、RNN与LSTM预测算法

典型的RNN网路结构如下：

x为输入层，o为输出层，s为隐含层，而t指第几次的计算；V,W,U为权重，其中计算第t次的隐含层状态时为St = f(U*Xt + W*St-1)，实现当前输入结果与之前的计算关联的目的。

RNN的局限：

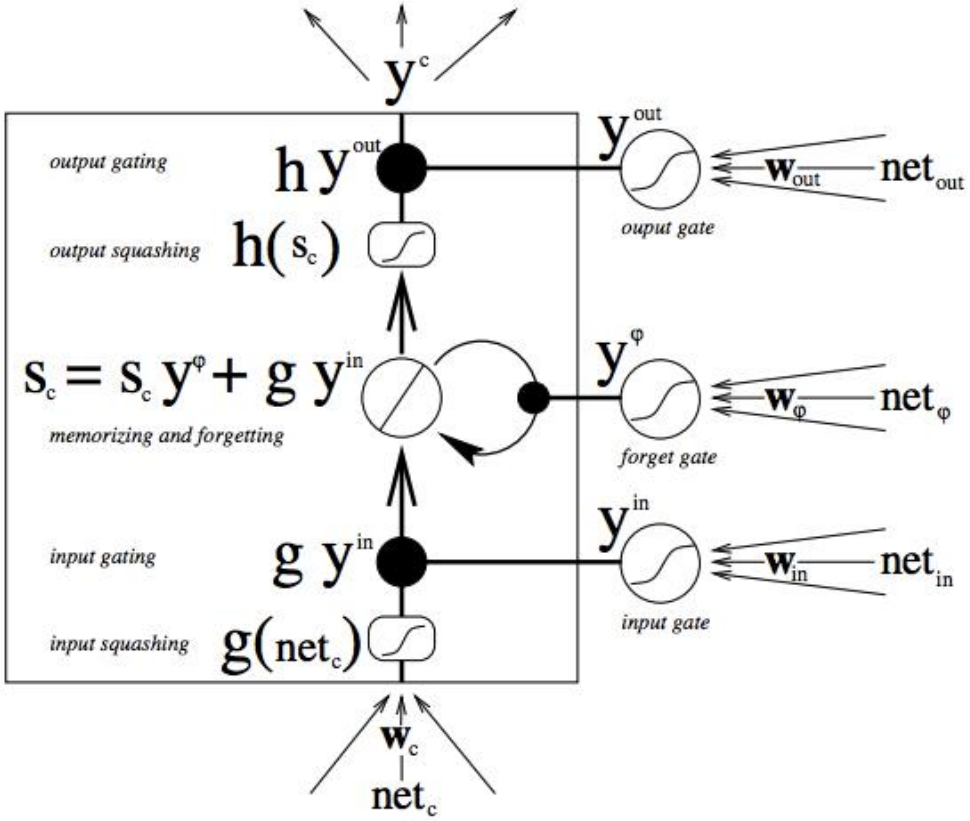
RNN模型如果实现长期记忆的话需要将当前的隐含态的计算与前n次的计算关联，即St = f(U*Xt + W1*St-1 + W2*St-2 + ... + Wn*St-n)，这样计算量会呈指数式增长，导致模型训练的时间大幅增加，因此RNN模型一般直接用来进行长期记忆计算。



4.1、RNN与LSTM预测算法

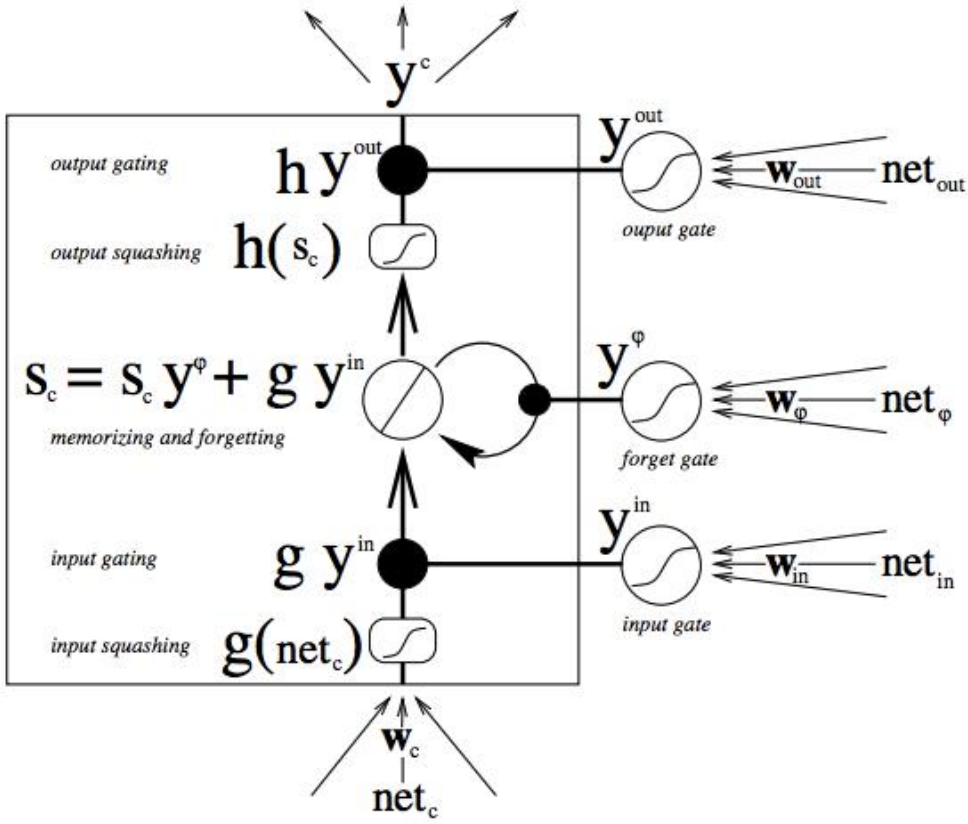
LSTM （Long Short-Term Memory）模型

LSTM模型是一种RNN的变型， LSTM的特点就是在RNN结构以外添加了各层的阀门节点。



4.1、RNN与LSTM预测算法

具体的，阀门有3类：遗忘阀门（forget gate），输入阀门（input gate）和输出阀门（output gate）。这些阀门可以打开或关闭，用于将判断模型网络的记忆态（之前网络的状态）在该层输出的结果是否达到阈值从而加入到当前该层的计算中。如图中所示，阀门节点利用sigmoid函数将网络的记忆态作为输入计算；如果输出结果达到阈值则将该阀门输出与当前层的的计算结果相乘作为下一层的输入（**PS：这里的相乘是在指矩阵中的逐元素相乘**）；如果没有达到阈值则将该输出结果遗忘掉。每一层包括阀门节点的权重都会在每一次模型反向传播训练过程中更新。



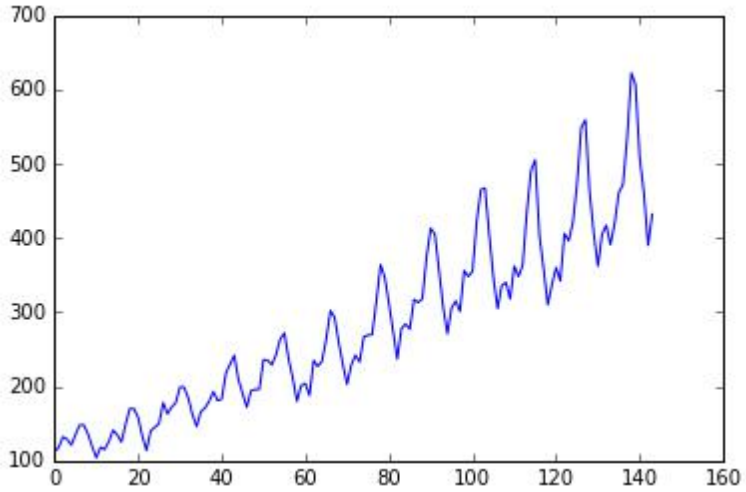
4.1、RNN与LSTM预测算法

案例四：预测未来 1 个月的攻击数量

【从这 12 年的数据可以看到上升的趋势，每一年内的 12 个月里又有周期性季节性的规律】

步骤一：当激活函数为 sigmoid 或者 tanh 时，要把数据正则话，此时 LSTM 比较敏感

设定 67% 是训练数据，余下的是测试数据



4.1、RNN与LSTM预测算法

案例四：预测未来 1 个月的攻击数量

步骤二：建立 LSTM 模型

输入层有 1 个input，隐藏层有 4 个神经元(LSTM cell值)，输出层就是预测一个值，激活函数用 sigmoid，迭代 100 次，batch size(每次训练样本数)为 1

借助python keras库实现LSTM：

```
# create and fit the LSTM network
model = Sequential()
model.add(LSTM(4, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=2)
```

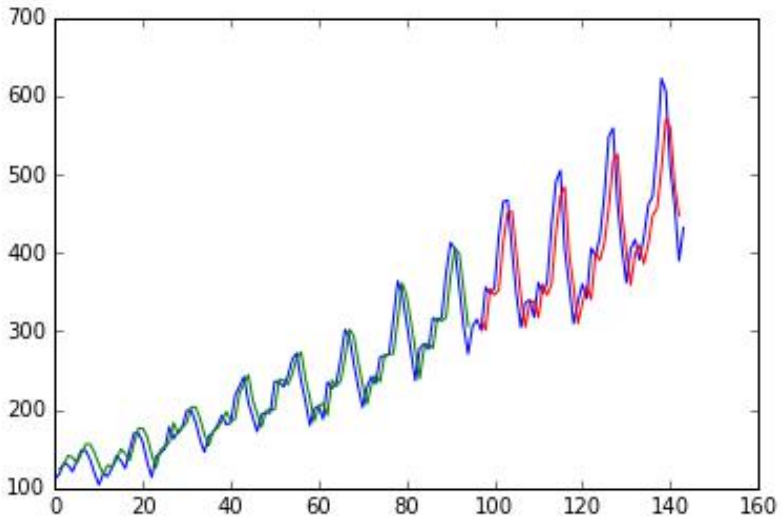
verbose（日志显示参数）

4.1、RNN与LSTM预测算法

步骤三：训练与预测

画出结果：蓝色为原数据，绿色为训练集的预测值，红色为测试集的预测值

```
# make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
```



4.1、RNN与LSTM预测算法

步骤四：计算误差
计算 **mean squared error**

```
1  trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))
2  print('Train Score: %.2f RMSE' % (trainScore))
3  testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
4  print('Test Score: %.2f RMSE' % (testScore))
```

Train Score: 22.92 RMSE
Test Score: 47.53 RMSE

典型方法比较

类型	方法	结果形式	建模时间	空间开销	评估时间	特征数量	可扩展性	可理解性	知识来源
基于数学模型	AHP	打分	短	小	短	少	难	易	经验知识
	集对分析	分级	短	小	短	少	易	中	经验知识+差异度
基于知识推理	模糊集	分类	中	中	中	中	中	易	专家知识
	马尔科夫	分类	中	大	中	中	中	易	专家知识+可能性
	贝叶斯	分类	中	大	中	中	中	易	专家知识+可能性
基于模式识别	灰色关联	分类	中	中	长	多	中	中	经验知识+关联度
	神经网络	分类	长	中	中	多	易	难	机器学习
	粗集理论	分类	长	中	中	多	易	难	机器学习

预测算法选取原则

- 没有单一的方法适用于所有的数据集；
- 在考虑使用哪种方法的时候，需要考虑精确度，训练时间，鲁棒性，可解释性，可扩展性等问题。

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

谢 谢!