人工智能在网络安全领域的发展

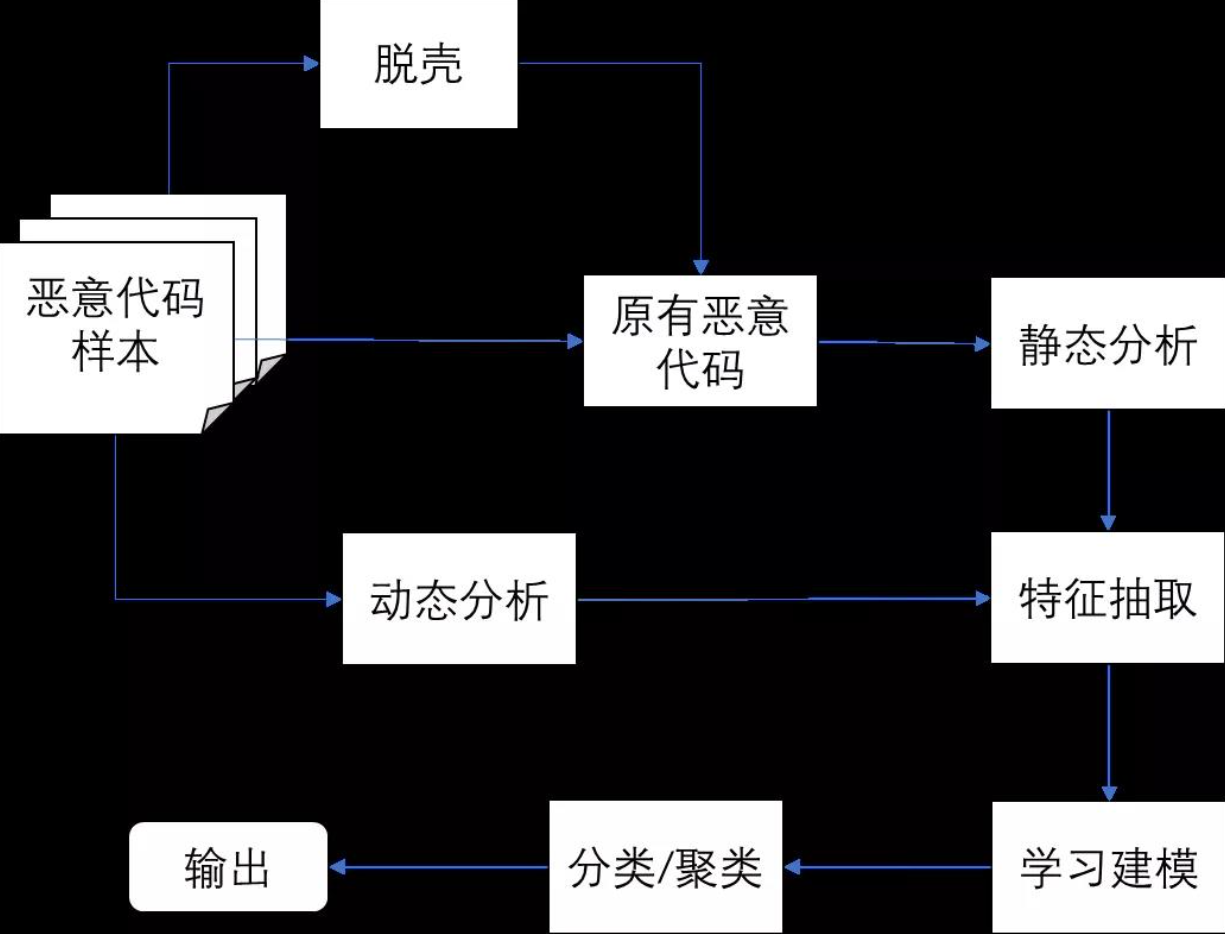
由于无处不在的联网计算机、云计算和移动技术，网络攻击变得更加普遍和多样化。大量连接的设备为网络犯罪分子提供了大量可供攻击的接入点，而这些接入点缺乏足够的安全性。物联网的兴起引发了一波不断扩大的网络攻击浪潮，网络欺诈经常出现在新闻中。

传统的网络安全方法需要大量的人力来识别威胁、提取其属性并将其编码到可以检测威胁的软件中。但是人类无法再扩展以充分保护动态的企业攻击面，并且传统方法不像当今的网络攻击那样复杂。近年来，人工智能已成为增强人类信息安全团队工作的必要技术。人工智能使网络安全团队能够形成强大的人机合作关系，从而突破我们的知识界限并推动网络安全大于其各个部分的总和。基于 AI 的网络安全系统制定重要的优先级决策，不仅基于可用于被攻击系统的内容，还基于最有可能用于攻击系统的内容，这其实就是最近非常火热的RASP设备，在阿里云相当多的网站上已经上线了该设备，这种设备可以对IP进行自动封禁，让一些黑客难以攻击。举个简单的例子，XSS攻击（Cross-Site Striking）的攻击手段是 <script>alert(1)</script>，可以插入进一些搜索框中，如果RASP检测到了相关于<script>这一关键字，则会对你进行IP封禁。人工智能和机器学习可以帮助防御者了解网络罪犯的最新情况，自动化威胁检测，并比传统的软件驱动或手工技术更有效地做出响应。

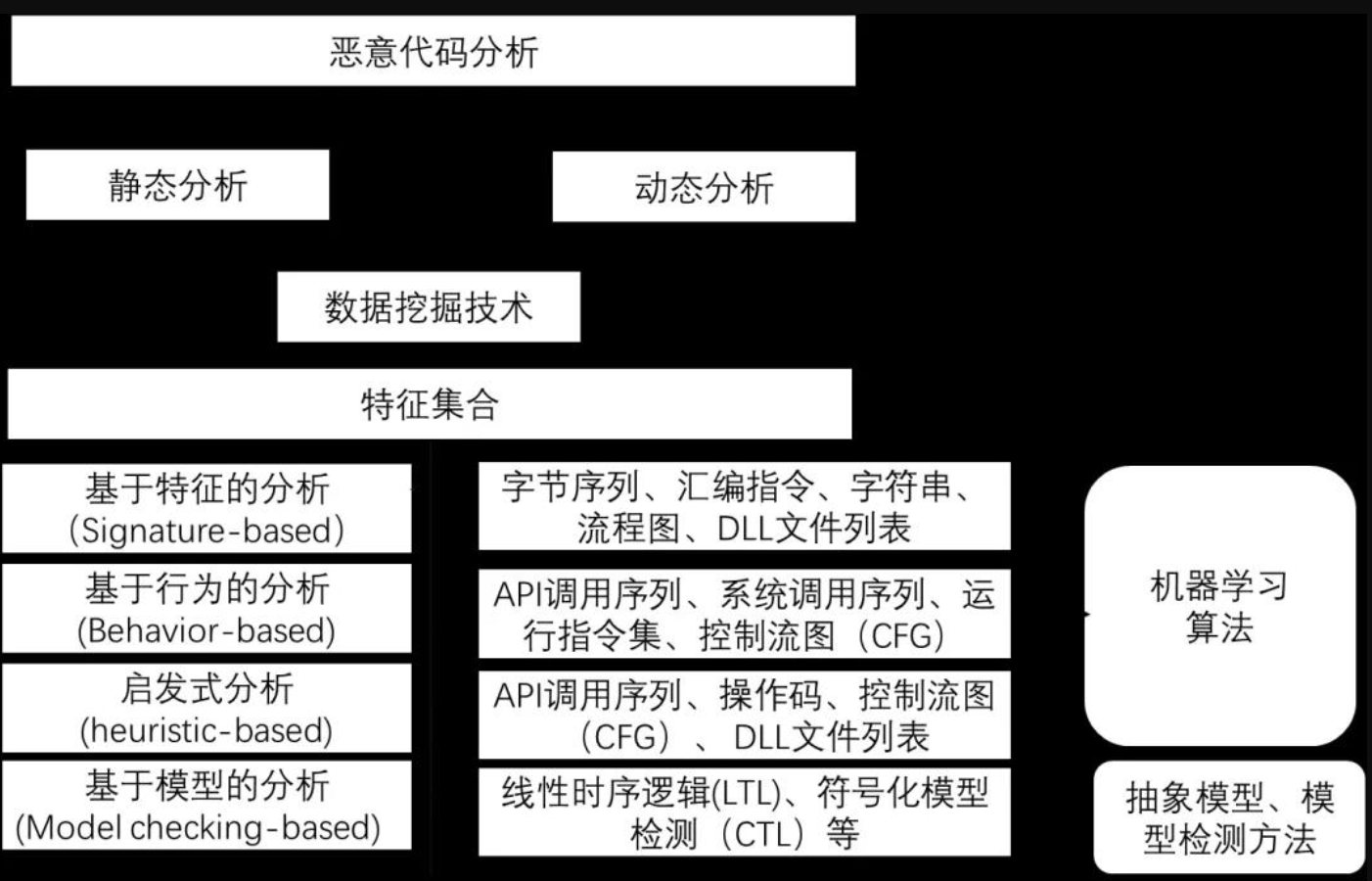
基于人工智能的网络安全系统可以提供关于全球和行业特定危险的最新信息，从而更好地制定重要的优先级决策，不仅基于可能被用来攻击系统的内容，而且基于最有可能被用来攻击系统的内容。这也是深度学习技术带来的重大意义。人工智能可以识别风险并确定风险的优先级。使用人工智能软件检测威胁并不新鲜，因为大多数网络安全公司都部署了人工智能算法来自动调查和识别攻击指标。但是现在，通过使用复杂的算法，人工智能系统正在接受训练以检测恶意软件、运行模式识别，甚至在恶意软件或勒索软件攻击进入系统之前检测它们的最细微行为。实时检测偏差和行为变化有助于组织更快、更智能地做出响应。

早期恶意软件是为了某种简单的恶意目的而编写，更容易被检测到，这种恶意软件可以定义为传统恶意软件。随着与地下经济体系深度融合，恶意代码可能带来巨大的经济利益，其整体隐蔽性越来越强，这种可以在内核模式下运行、比传统恶意软件更具破坏性且更难检测的恶意软件被定义为新一代恶意软件。

新一代恶意软件的特征提取和分类变得越来越难，一个恶意软件实例可能同时呈现多个类别的特征，往往很容易绕过运行在内核模式下的防火墙、防病毒程序等防护软件。一般来说，传统的恶意软件由一个进程组成，不使用复杂的技术来隐藏自己，而新一代恶意软件可能会注入多个不同的现有进程或新进程，采用特定混淆技术来进行隐藏并在系统中持久潜伏，一旦发起攻击可能更具破坏性和持久性。

为了应对越来越隐蔽的恶意代码威胁，研究人员将数据挖掘技术广泛应用于恶意软件检测过程。这是一种基于人工智能、机器学习、模式识别、统计学、数据库、可视化等技术的决策支持过程，高度自动化地分析数据，进行归纳性推理，从中挖掘出潜在的模式，为决策提供支持。特别是基于人工智能的数据挖掘算法检测恶意应用程序，已经成为一个新的研究方向，受到越来越多国内外学者的关注，其分析检测流程如图3所示。

这种方法的核心在于特征提取和学习建模(分类或聚类)[2]。其中，特征提取过程通过静态分析或动态分析方法，捕获文件样本的特征或特征序列，这些特征通常提取自二进制字符串、API调用和程序执行行为等特征集，然后对这些特征或特征序列进行筛选，选择分类效果好的特征作为学习建模的分类输入参数。学习建模过程在特征分析的基础上，运用分类或聚类等智能算法自动化地将文件样本分类至不同的类别，根据类别判定文件是否属于恶意软件。

基于人工智能的恶意软件分析检测方法根据文件特征的获取方式、类型和分析手段的不同，可以分为4类，如图4所示

基于特征的分析方法一般是基于n-gram等算法提取的字节序列、操作码序列、函数调用序列、可执行文件的特征和字符串等特征，然后使用信息增益、文档频率、Fisher Score及层次特征选择等方法选择分类能力较强的特征，最后采用人工神经网络、贝叶斯网络、朴素贝叶斯、决策树模型、K近邻、支持向量机、随机森林、改进的决策树、改进的神经网络等实现分类。

基于行为的分析方法通过监控工具观察样本恶意软件行为，抓取其典型行为生成行为特征，生成API调用、系统调用、运行指令、控制流图等特征模式库，后期检测将采集软件的运行行为特征与库中的运行模式进行匹配，以判断是否属于恶意软件。其中从行为中提取特征将使用n-gram、n-tuple、bagging、知识图谱等数据挖掘技术，Hellinger距离、余弦距离、卡方等距离算法常用于计算特征之间的相似性，特征分类一般使用机器学习方法。

启发式检测方法是一种发现分析、学习和解决问题的方法，是基于规则估计或有根据的猜测来找到特定问题的解决方案。启发式这个词源于古希腊词“为了发现”。启发式检测技术一般特指静态启发式检测技术，通过对文件的API调用、操作码（Opcode）、控制流图、动态链接库等特征进行分析，在此基础上反汇编病毒程序和可疑程序，进行基于程序的指令级代码分析，模拟跟踪代码执行的流程，与已知恶意软件活动的代码进行比较进行判别。

软件模型检测是用来在程序执行过程中证明性质正确性的算法。它源于逻辑和定理证明，这两者都给出了基础问题形式化的基本概念，以及提供了分析逻辑问题的算法流程。模型检测最初是为了验证系统的正确性而开发的，也可被用来检测恶意软件。在这种检测方法中，恶意软件行为一般通过线性时序逻辑(LTL)、符号化模型检测（CTL）、计算树逻辑语言（CTPL）、控制流图等形式显示特定的特征。程序行为是通过查看一个或多个系统调用的流关系来创建的，并通过使用诸如隐藏、传播和注入等属性来定义行为，通过比较这些行为，可以确定该程序是恶意软件还是良性的。

此外，还有通过关联规则挖掘及序列模式挖掘等方法实现分类的方法，都是现阶段研究的热点。

当然现在人工智能不只是能应用于防御手段，在攻击面也存在，比如在很多年前就存在的OCR技术，这一技术作为人工智能的初步应用是可以作为攻击手段存在的。在普通的用户登录注册界面，我们都会需要输入验证码，而验证码的出现导致攻击者无法进行用户名与密码的爆破，从而无法进行有效的攻击。当人工智能应用于这一领域之后，便能够自动识别验证码并进行填写。对应的GitHub项目有：<https://github.com/f0ng/captcha-killer-modified>

参考文献：

[1]毛蔚轩,蔡忠闽,童力.一种基于主动学习的恶意代码检测方法.软件学报,2017,28(2):384-397 [doi: 10.13328/j.cnki.jos.005061].

[2] Ömer Aslan,Refik Samet, A Comprehensive Review on Malware Detection Approaches, 2020.01, IEEEAccess ( Volume: 8 ), P 6249 – 6271[DOI: 10.1109/ACCESS.2019.2963724].