O'REILLY®

机器学习与安全

Machine Learning and Security

Clarence Chio David Freeman 著
侯荣涛 王玉祥 徐旦华 侯丽倩 译

中国电力出版社

机器学习与安全

Clarence Chio David Freeman 著
侯荣涛 王玉祥 徐旦华 侯丽倩 译

ebastopol • Tokyo O'REILLY®
lly Media, Inc. 授权中国电力出版社出版

中国电力出版社

图书在版编目（CIP）数据

机器学习与安全 /（美）克拉伦斯·基奥（Clarence Chio），（美）戴维·弗里曼（David Freeman）著；侯荣涛等译. — 北京：中国电力出版社，2019.8

书名原文：Machine Learning and Security

ISBN 978-7-5198-3004-5

I. ①机… II. ①克… ②戴… ③侯… III. ①机器学习 IV. ①TP181

中国版本图书馆CIP数据核字(2019)第052598号

北京市版权局著作权合同登记 图字：01-2019-0697号

出版发行：中国电力出版社

地　　址：北京市东城区北京站西街19号（邮政编码100005）

网　　址：http://www.cepp.sgcc.com.cn

责任编辑：刘 炽（liuchi1030@163.com）

责任校对：黄蓓 闫秀英

装帧设计：Randy Comber，张 健

责任印制：杨晓东

印　　刷：三河市航远印刷有限公司

版　　次：2019年8月第一版

印　　次：2019年8月北京第一次印刷

开　　本：750毫米×980毫米 16开本

印　　张：23.5

字　　数：448千字

印　　数：0001—3000册

定　　价：88.00元

对本书的赞誉

未来在线网络的保密与安全取决于安全人员根据互联网规模和速度部署机器学习以发现和阻止恶意攻击行为的能力。有关这个主题，Chio 和 Freeman 在他们这本权威性的著作中阐述了捍卫该领域安全的最新的学术思想和融汇机器学习深刻知识的教程。

——Alex Stamos
Facebook 安全主管

该书是一本期盼学习如何运用机器学习技术检测计算机系统异常，保护终端用户安全的优秀实用指南。

——Dan Boneh
斯坦福大学计算机科学教授

对于期盼了解机器学习在安全方面作用的读者，本书给出了非常清晰的结果。

——Nwokedi C. Idika 博士
Google 安全与隐私组织软件工程师

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

作者介绍

Clarence Chio 是一位工程师和企业家，他在 DEF CON 及其他十多个国家举办的安全与软件工程会议上做过机器学习与安全方面的讲演、研讨和培训课程。他之前曾经是 Shape Security 公司安全研究团队的成员，是 Intel 的社区发言人，也是 Oracle 的安全顾问。Clarence 从事的工作是为初创公司在安全数据科学方面提供咨询建议，他创立和领导的网络安全活动小组聚集了旧金山湾区最多的安全数据方面的科学家从事数据挖掘研究。他拥有斯坦福大学学士和硕士学位，主攻数据挖掘和人工智能。

David Freeman 作为研究员和工程师，在 Facebook 公司致力于处理完整性和滥用问题。他曾在 Linkedln 领导了反滥用工程和数据科学团队。在那里他建立了检测欺诈和滥用情况的统计学模型，并与 LinkedIn 的大型机器学习社团合作构建了可扩展的建模和评分基础设施。他是国际机器学习与安全会议，如 NDSS、WWW 和 AISec 的作者、主持人和组织者，发表了 20 多篇关于计算机安全数理统计方面的学术论文。他在加州大学伯克利分校获得数学博士学位，并在 CWI 和斯坦福大学进行了密码学和安全方面的博士后研究。

封面介绍

本书封面上的动物是西伯利亚蝮蛇（Gloydius halys），也称哈利斯蝰蛇，它生活在乌拉尔山脉以东的亚洲广大地区，包括俄罗斯和中国的部分地区。这种蛇有毒，属于蝮蛇科。蝮蛇因其口鼻部的深坑处长有特殊的热感应器官而得名。这个凹坑器官有助于它找到调节温度所需的冷暖适宜的地方，还能够感知和攻击猎物。大多数种类的蝮蛇也会生下幼崽而不是产卵。

西伯利亚蝮蛇的长度约为 21~23 英寸，雌性比雄性略长。它的鼻子微微翘起，皮肤长有大的颜色深浅相间的横向条纹图案（根据亚种不同，颜色从灰色到浅棕色或黄色不等）。

这种蛇以伏击的方式进行狩猎，它静静地爬在地上等待猎物（如鸟或小型哺乳动物）足够靠近时开始进攻，毒液会麻痹或杀死猎物，然后将其整个吞食，和大多数蛇一样，西伯利亚蝮蛇通常不会伤人，只有在它们感觉受到威胁时才会咬人。

封面图片来自 Lydekker's Royal Natural History。

目录

前言

机器学习正在改变着世界。从通信、金融到交通运输、制造业，甚至农业[注1]，几乎每一个技术领域都被机器学习和人工智能改变着，或者很快就会被改变。

计算机安全也在吞噬世界。随着我们在工作、娱乐和社交生活上对计算机越来越多的依赖，攻破这些系统的价值在相应地增加，吸引了越来越多希望从中赚钱或只是进行破坏的攻击者。此外，随着系统变得越来越复杂和相互关联，要确保攻击者没有漏洞或后门可钻变得越来越困难。其实，在本书出版之前，我们就了解到目前使用的很多微处理器都是不安全的[注2]。

通过机器学习可以（潜在）解决所有显现的问题。将机器学习应用到计算机安全领域是很自然的，其实质是为机器学习蓬勃发展提供坚实稳健的数据集。事实上，对于新闻中给出的所有安全威胁，我们听到的只是许多关于人工智能如何“彻底改变”我们解决安全问题的方式的说法。由于它承诺能够限制攻击者在某些最复杂领域方面的进展，机器学习被吹捧为将最终结束攻击者和防御者之间的“猫和老鼠”游戏的一种技术。走在大型安全会议的浏览区内，显而易见，越来越多的公司开始利用机器学习来解决安全问题。

注1：Monsanto, “How Machine Learning is Changing Modern Agriculture,” *Modern Agriculture*, September 13,2017, *https://modernag.org/innovation/machine-learning-changing-modern-agriculture/*。

注2：“Meltdown and Spectre,” Graz University of Technology, accessed January 23, 2018, *https://spectreattack.com/*。

反映这两个领域对相互结合的日益增长的兴趣，伴随而来的还有认为其是炒作的冷嘲热讽。那么，我们如何达到平衡呢？人工智能应用于安全的真正潜力是什么？如何区分市场无价值的东西和有前途的技术？我应该用什么来解决我的安全问题呢？要想回答这些问题，最好的方法就是深入研究科学，理解核心概念，做大量的测试和实验，让结果自己说话。然而，这样做需要对数据科学和计算机安全有一定的了解。在工作中，通过建立安全系统，领导了反滥用团队，会议演讲，我们遇到了一些具有这两方面知识的人，也遇到了很多具有一方面知识而想要学习另一方面知识的人。

这就是本书的目标。

本书的内容

我们撰写本书的目的是为讨论两个无处不在而又不可避免进行结合的概念：机器学习和安全提供一个架构。尽管在某些文献上（及多个会议的研讨会上：计算机与通信会议（CCS）的人工智能与安全（AISec），美国人工智能协会（AAAI）的计算机网络安全的人工智能（AICS）和神经信息处理系统（NIPS）大会中的关于机器诈骗）也阐述了这两个学科的相互结合，但现有的大部分工作都是学术或理论的。特别是，我们没有找到一个提供具体的、能够对安全从业者了解数据科学，帮助机器学习从业者有效思考现代安全问题的可运行的代码实例向导。

在安全检查范围广泛的主题空间，我们提供如何应用机器学习增强或取代基于规则或启发式解决入侵检测，恶意软件分类或网络分析等问题。除了探索核心机器学习算法和技术，我们关注的挑战是在安全空间内建立可持续的、可靠的和可伸缩的数据挖掘系统。通过有效和指导性的讨论，我们向你展示如何在一个不友好环境中考虑数据，如何识别可能被噪声淹没的重要信号。

本书面向的读者对象

如果你工作在安全领域并且希望使用机器学习来改进你的系统，这本书是为

你准备的。如果你已经学习过机器学习，现在想用它来解决安全问题，这本书也是为你准备的。

我们假设你有一些基本的统计学知识，在第一次阅读本书时可以跳过大多数复杂的数学内容，而不会失去概念。我们还假设你熟悉一种编程语言。我们的示例是用 Python 语言写的，我们还提供了对所需 Python 包的引用来实现我们讨论的概念，不过你可以使用 Java、Scala、C++、Ruby 和许多其他语言中的开放源码库来实现相同的概念。

排版约定

本书使用了下述排版约定。

斜体（Italic）

表示新术语、URL、示例电子邮件地址、文件名、扩展名、路径名和目录。

等宽字体（`Constant Width`）

表示代码，在段内用以表示与代码相关的元素，例如变量或函数名、数据库、数据类型、环境变量、声明和关键字。

等宽粗体字（**`Constant width bold`**）

表示命令或其他用户输入的文本。

斜体等宽字体（*`Constant Width Italic`*）

表示该文本应当由用户提供的值或由用户根据上下文决定的值替换。

表示提示、建议或一般说明。

表示警告或提醒。

使用代码示例

可以从 *https://github.com/oreilly-mlsec/book-resources* 网址下载有用的补充材料（代码示例、练习等）。

本书是为了帮助完成你的工作。一般来说，如果本书提供示例代码，你可以在程序和文档中使用这些代码，而不需要得到我们的许可，除非对这些代码中的大量内容进行了复制。例如，如果你在编写程序时，使用了本书提供的几段代码，这不必得到许可。而出售或分发 O'Reilly 图书中的 CD-ROM 光盘则需要得到许可。通过引用本书和本书中的示例代码来解答问题也不需要得到许可，将本书中大量的示例代码引用到你的产品文档中确实需要获得许可。

我们赞赏使用本书时写上作品的所有者，但不做要求。所有者信息通常包括标题、作者、出版商和 ISBN。例如：“Machine Learning and Security, by Clarence Chio and David Freeman (O'Reilly). Copyright 2018 Clarence Chio and David Freeman, 978-1-491-97990-7”。

如果你对代码示例的使用超出了正常范围，或者超出了上面的许可范围，请通过 *permissions@oreilly.com* 与我们联系。

O'Reilly Safari

Safari（以前的 Safari Books Online）是面向企业、政府、教育和个人的会员制培训与参考平台。

Safari 的会员可以访问成千上万的书籍、培训视频、学习路径、交互式教程和推荐的书单。这些内容由 250 多家出版社提供，其中包括：O'Reilly Media、Harvard Business Review、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Adobe、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、

McGraw-Hill、Jones & Bartlett 和 Course Technology 等。

更多关于 Safari 的信息，请访问我们的网站：*http://oreilly.com/safari*。

联系我们

请把你对本书的意见和疑问发给出版社：

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街2号成铭大厦C座807室（100035）
奥莱利技术咨询（北京）有限公司

这本书有专属网页，你可以在那里找到本书的勘误、示例和其他信息。这个网页的地址是 *http://bit.ly/machineLearningAndSecurity*。作者还为本书在 *https://mlsec.net* 上创建了一个网站。

如果你对本书有一些评论或技术上的建议，请发送电子邮件到*bookquestions@oreilly.com*。

要了解O'Reilly图书、培训课程、会议和新闻的更多信息，请访问我们的网站，地址是：*http://www.oreilly.com*。

我们的 Facebook：*http://facebook.com/oreilly*。

我们的Twitter：*http://twitter.com/oreillymedia*。

我们的 Youtube：*http://www.youtube.com/oreillymedia*。

致谢

我们对 Hyrum Anderson、Jason Craig、Nwokedi Idika、Jess Males、 Andy Oram、Alex Pinto 和 Joshua Saxe 表示感谢，他们对本书早期的初稿进行了全面的技术审查和反馈。还要感谢 Virginia Wilson、Kristen Brown 和 O'Reilly 的所有工作人员，他们帮助我们将该项目从概念变为了现实。

Clarence 感谢 Christina Zhou 为本书度过的所有不眠之夜和周末休息的时间，感谢 Yik Lun Lee 对本书初稿的校对和发现代码的错误，感谢 Jarrod Overson 对我的鼓励，感谢吉娃娃 Daisy 在最艰难的时候陪在我身边。感谢 Anto Joseph 给我讲解安全方面的知识。感谢所有其他黑客、研究人员和以不同方式对本书产生影响的培训参与者。感谢我的同事在形状安全方面让我成为一个更好的工程师。感谢“数据挖掘”为网络安全演讲者和与会者成为推动该领域研究协会一员的贡献。最重要的是，感谢在新加坡的家人来自世界各地的支持，让我追逐我的梦想，追求我的激情。

David 感谢 Deepak Agarwal 鼓励我承担了这项任务。感谢 Dan Boneh 教我如何思考安全问题。感谢我在 LinkedIn 和 Facebook 的同事们为我说明安全问题在现实世界中的表现。还要感谢 Grace Tang 关于机器学习和专项训练的反馈。最大的感谢是对 Torrey、Elodie 和 Phoebe，他们为了完成这本书，和我一起忍受了许多不眠之夜和一些奇怪的短途旅行，而且从未动摇过对我的支持。

第 1 章

为什么要学习机器学习与安全？

起初，垃圾邮件泛滥。

当专家学者和科学家将足够多的计算机通过互联网连接起来构建一个有价值的通信网络之时，有些人就认识到这种免费传播和广泛分布的媒体是宣传粗制滥造产品、窃取账户凭证和传播计算机病毒的良好途径。

在这几十年里，计算机和网络安全领域面临到了巨大范围的威胁：入侵检测、Web 应用安全、恶意软件分析、社交网络安全、高级持续性威胁和应用密码技术等庞大领域，这里只例举很少部分。但即使在今天，垃圾邮件依旧是电子邮件或消息传递中重点关注的对象。对于普通公众而言，垃圾邮件可能是最能够直接触及他们生活的计算机安全问题。

机器学习不是那些同垃圾邮件斗争的工程师的发明，但它很快被专注于统计学的技术专家所采用，因为他们发现了它在处理不断变化的滥用源方面的潜力。电子邮件提供商和互联网服务提供商（ISP）可以查看丰富的电子邮件内容、元数据和用户的行为举止。通过使用电子邮件数据，可以构建基于内容的模型，以创建识别垃圾邮件的通用方法。可以从电子邮件中提取元数据和实体的影响力，在无需查看电子邮件内容的情况下来预测电子邮件是垃圾邮件的可能性。通过实例化用户行为的反馈回路，系统可以建立集体智慧并随着时间的推移在用户的帮助下提升系统的性能。

电子邮件过滤器因此逐渐演变为处理垃圾邮件发送者投放垃圾邮件所用的越来越多的各种欺骗方法。尽管今天发送的电子邮件中有 85% 是垃圾邮件（根

据一个研究组的调查结果），最好的现代垃圾邮件过滤器阻止了超过 99.9% 的垃圾邮件，而且对于那些主要利用电子邮件的用户来说，很少在他们的收件箱中查看到未经过滤和未检测的垃圾邮件。这些结果表明，在互联网早期开发的极其简化的垃圾邮件过滤技术已经取得了巨大的进步，该技术利用简单的词语过滤和电子邮件元数据影响力实现了较好的结果。

研究人员和从业者从这场与垃圾邮件的斗争中得到的主要经验是，懂得了利用数据击败恶意对手和提高我们进行技术交互质量的重要性。事实上，与垃圾邮件战斗的故事是一个在计算机安全的任意领域中使用数据和机器学习的代表性实例。今天，几乎所有的组织都严重依赖于科技，而几乎每一项技术都有安全方面的漏洞。受到与 20 世纪 80 年代的那些垃圾邮件发送者相同的核心动机（无管制，免费获得普通群体的可支配收入和私人信息）的驱使，恶意行为者可以对现代生活几乎所有方面构成安全威胁。其实，在计算机安全的所有领域，攻击者和防御者之间的战斗的基本性质都与在垃圾邮件战斗中的性质相同：一位积极的对手经常试图破坏性地使用计算机系统，每一方都力图在另一方发现其设计或技术中的缺陷之前，修复或利用这些缺陷。问题的陈述并未改变一丝一毫。

计算机系统和网络服务已经变得越来越集中化，许多应用程序已经发展到能够为数百万甚至数十亿用户提供服务。那些成为信息决策者的实体是更大的利用目标，但这也有利于实体利用数据和其用户基础获得更好的安全性。再加上现如今功能强大的数据处理硬件的出现，以及更完善的数据分析和机器学习算法的开发，这将是开发机器学习在安全领域潜能的最好时机。

在本书中，我们演示了机器学习和数据分析技术在各种安全和滥用问题领域中的应用。我们探讨对于如何评估不同机器学习技术在不同情况下的适用性的方法，并将重点放在那些有助于你更安全使用数据的原则上。我们的目标不是为你可能面临的每个安全问题提供答案，而是为你提供一个思考数据和安全性的架构，以及一个可以从中为你的问题选择正确解决方法的工具箱。