基于人工免疫原理的检测系统模型及其应用

赵林惠/著

清华大学出版社

基于人工免疫原理的检测系统模型及其应用

赵林惠/著

清华大学出版社
北京

图书在版编目(CIP)数据

基于人工免疫原理的检测系统模型及其应用 / 赵林惠 著. —北京：清华大学出版社，2016

ISBN 978-7-302-42392-8

Ⅰ.①基… Ⅱ.①赵… Ⅲ.①免疫学—应用—人工智能—研究 Ⅳ.①TP18

中国版本图书馆 CIP 数据核字(2015)第 296367 号

责任编辑：王燊娉　胡花蕾
封面设计：赵晋锋
版式设计：周玉娇
责任校对：曹　阳
责任印制：刘海龙

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　邮　　编：100084
社 总 机：010-62770175　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：三河市君旺印务有限公司
装 订 者：三河市新茂装订有限公司
经　　销：全国新华书店
开　　本：170mm×240mm　印　张：16.25　字　数：211 千字
版　　次：2016 年 2 月第 1 版　印　次：2016 年 2 月第 1 次印刷
定　　价：58.00 元

产品编号：067516-01

前　言

免疫是在复杂环境条件下保障人体健康、安全的一套行之有效的机制和系统。模拟人类免疫能力的系统称为人工免疫系统。对于生物免疫原理和人工免疫系统的研究不仅具有重要的理论价值，而且有广阔的应用前景，近年来得到国内外学术界越来越广泛的关注。目前国内关于人工免疫系统的研究已取得许多成果，在国际上也具有相当的影响力，出版了一批专著，也发表了大量文章。在这种良好的研究氛围中，笔者结合自己在攻读博士学位期间和近几年工作中的研究，希望将生物免疫系统原理和机制在入侵检测、故障诊断、病毒检测等异常检测领域中的应用作进一步总结，并且抛砖引玉，帮助初涉人工免疫系统的学者了解这个应用领域，进而找到适当的切入点展开自己的研究。

事实上，人工免疫系统在优化、调度、控制、数据挖掘、路径规划等领域都得到了很好的运用和发挥。本书仅仅将重点放在异常检测这个应用领域，主要探讨基于人工免疫原理的各种异常检测模型或检测系统。作为人工免疫系统生物学基础的生物免疫系统，旨在区分外部有害抗原和自身组织并清除病原以保持有机体的稳定，而基于人工免疫的检测模型可以实时发现对象出现的异常进而进行快速处理，两者都是使受保护对象在不断变化的环境中维持系统的稳定性，因而具有极大的相似性。从这个角度来说，将人工免疫原理用于异常检测系统的设计较之其他方法具有不可比拟的优势。目前国内还没有专门讨论基于人工免疫原理的异常检测模型方面的专著，可以说这是本书的价值所在。

本书作者的研究成果得到国家部委预研基金资助(YJ0467011)、

北京理工大学基础研究基金、北京市属高等学校高层次人才引进与培养计划项目(CIT&TCD201404081)的资助。

本书有两大部分,共10章,主要内容包括生物免疫系统原理、人工免疫系统原理、人工免疫系统模型、各种检测系统模型及应用等,重点在于各种应用中具体问题的解决方法论述和分析,分别涉及入侵检测、故障诊断、垃圾邮件检测、文件病毒检测等应用。第一部分是本书的创新点所在,主要是利用危险理论、免疫网络理论和数据融合理论,提出了一种基于免疫的三级模块式自适应检测系统模型,详细论述了多变异模式独特型网络自适应检测器的设计和模板可调式自适应决策融合算法的设计。第二部分分别给出了基于人工免疫系统的正常模型的免疫算法设计及其在文件病毒检测中的应用、面向流数据特征提取的人工免疫网络模型及其在垃圾邮件检测中的应用、基于云模型与危险理论的人工免疫系统模型、基于人工免疫的故障检测与诊断等。

书中有些内容直接引用、参考了国内外许多文献,首先向所有被引用文献的作者表示由衷的感谢,并在参考文献中列出,恕不一一列举。感谢本书作者的导师北京理工大学戴亚平教授多年来的支持和鼓励,感谢师妹付冬梅、周丽华的支持和帮助。

感谢学校领导的大力支持以及家人的关心和鼓励。此外,出版社的工作人员为本书稿的整理打印做了许多工作,感谢你们为本书顺利问世所作的努力。

由于水平有限、经验不足,本书在内容的组织上难免存在疏漏和不妥之处,竭诚希望各同行专家以及广大读者提出宝贵的意见和建议,以使本书不断完善。

作　者

2015年9月

目　录

第1章　绪论

1.1　生物免疫系统原理

生物免疫系统作为一个高度进化的生物系统，旨在区分外部有害抗原和自身组织，清除病原并保持有机体的稳定。从计算的角度来看，生物免疫系统是一个高度并行、分布、自适应、自组织的系统，具有很强的学习、识别、记忆和特征提取能力，以及强大的信息处理能力。自20世纪40年代以来，随着医学在生物免疫系统研究领域的发展，人们对它的认识与理解不断得到深化与完善，逐渐形成了一门较为完善的学科——生物免疫学(immunnology)。

1.1.1　生物免疫系统的组成

生物免疫系统是一个极其复杂的自适应系统，是人工免疫系统的生物学基础。生物免疫系统是由免疫分子、免疫组织和免疫细胞组成的复杂系统。这些组成免疫系统的组织和器官分布在人体内的各处，涉及中枢免疫器官(骨髓、胸腺)和外周免疫器官(脾脏、淋巴结和黏膜免疫系统)，负责执行免疫功能，如图1-1所示。

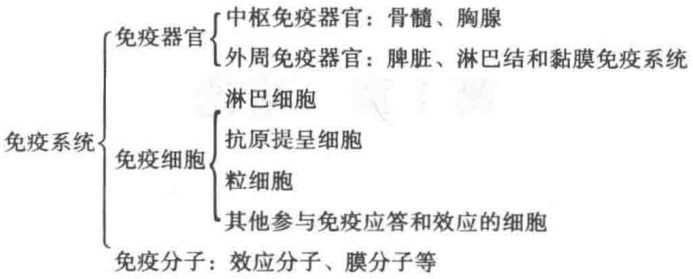

图 1-1　生物免疫系统的组成

免疫器官中执行免疫功能的主要是各类免疫细胞，如淋巴细胞(包括 T 淋巴细胞、B 淋巴细胞、自然杀伤细胞等)、抗原提呈细胞、粒细胞及其他参与免疫应答和效应的细胞。其中 T(B)淋巴细胞是参与适应性免疫应答的关键细胞，分别发挥细胞免疫和体液免疫效应；抗原提呈细胞则具有摄取、加工、处理抗原的能力，并可将经过处理的抗原肽提呈给特异性 T 细胞；各类粒细胞主要发挥非特异性免疫效应。

除免疫器官和免疫细胞外，多种免疫分子也被视为免疫系统的组成部分，包括活化的免疫细胞所产生的多种效应分子(如免疫球蛋白、细胞因子)、表达于免疫细胞表面的各类膜分子(如特异性抗原受体、CD 分子、黏附分子、主要组织相容性分子、各类受体)等。

人体免疫系统分为固有免疫系统(the innate immune system)和适应性免疫系统(the adaptive immune system)[1]。

①固有免疫系统是天生就有的，不随特异病原体变化，由补体、内吞作用系统和噬菌细胞系统组成。固有免疫系统具有与病原体第一次遭遇就能消灭它们的能力，它还能够识别自体和非自体组织结构，参与到自体与非自体组织识别中，并起到促进适应性免疫的重要作用。

②适应性免疫系统使用两种类型的淋巴细胞(T 细胞和 B 细胞)。适应性免疫系统能完成固有免疫系统不能完成的免疫功能，清除后者不能清除的病原体。

一旦病原体进入身体，固有免疫系统和适应性免疫系统就开始处理，此时两个系统的细胞都由多种细胞和分子以复杂的方式交互作用来检测和消除病原体。检测和消除功能都依赖化学结合：免疫细胞表面都覆盖不同受体，其中的一些结合病原体，而另一些结合其他免疫系统细胞或者分子，使系统发出信号触发免疫应答。

1.1.2 生物免疫系统的功能

我们知道，生物免疫学以生物机体免疫系统的组织结构和生理功能为研究主体；以免疫系统的种系发生与个体发生，免疫细胞的起源、分化、特征与功能，淋巴细胞的识别、活化与效应机制和机体免疫反应的调节等为研究内容；以细胞学、分子学和生物化学等为主要研究手段，着力揭示生物机体对“自体”(Self)和“非自体”(Non-Self)抗原的识别与应答、排斥异己和维持自身耐受过程中的奥秘。因此，免疫功能就是免疫系统在识别和清除“非己”抗原的过程中所产生的各种生物学作用的总称。生物免疫系统的功能包括免疫防御、免疫自稳和免疫监视，可以维持机体免疫系统的稳定，并降低肿瘤、持续性感染等疾病发生的可能性。

1. 免疫功能的分类及生理和病理反应

免疫功能的分类及生理和病理反应如下：

(1)免疫防御(immune defence)

抗感染免疫是机体排斥外来抗原性异物的一种免疫保护功能。正常时可产生抗感染免疫的作用，能够抵御和清除病原微生物及其他抗原的侵袭；防御功能过强会产生超敏反应，不足时则产生免疫缺陷。

(2)免疫自稳(immune homeostasis)

这是机体免疫系统维持内环境相对稳定的一种生理功能。

正常时，机体可及时清除体内损伤、衰老、变性的血细胞和抗原-抗体复合物，而对自身成分保持免疫耐受；异常时，发生生理功能紊乱，出现自身免疫疾病等。

(3)免疫监视(immune surveillance)

这是机体免疫系统及时识别、清除体内突变、畸变和病毒干扰细胞的一种生理保护作用。如丧失免疫监视，机体突变细胞失控，有可能导致肿瘤发生，或出现病毒的持续感染。

2. 免疫系统的免疫应答机制及功能

免疫系统的免疫应答机制及其许多重要的功能包括免疫识别、免疫学习、免疫记忆、免疫宽容和免疫自适应调节等，它们是免疫学研究的重要内容。

(1)免疫应答

免疫应答是指病原体入侵生物体后，免疫系统作出反应。主要的类型有两种：固有性免疫应答(innate immune response)[2]和适应性免疫应答(adaptive immune response)[3]。

①固有性免疫应答，即遇病原体后，固有免疫系统首先产生的迅速而短时间的应答。固有免疫应答是固有免疫系统执行固有免疫功能的结果，适应性免疫应答的执行者是前面介绍的T淋巴细胞及B淋巴细胞，故又称为抗原特异性免疫应答(antigen-specific immune response)。

②适应性免疫应答由适应性免疫系统产生。适应性免疫应答分为两种过程：初次免疫应答及二次免疫应答。应答过程如图1-2所示。

对图1-2分析可知，初次免疫应答发生在某种病原体第一次入侵时。此时免疫系统产生大量抗体，清除生物体内的抗原。初次应答学习过程较慢，通常发生在身体初次感染的前几天，要用几周时间才能清除抗原。初次免疫应答后，免疫系统首次遭遇异体物质且该物质已经被清除到体外，但免疫系统中仍保留一定数量的B细胞，称为免疫记忆细胞。免疫记忆细胞保持对

初次遭遇的抗原特性的记忆，即免疫记忆。免疫记忆使得免疫系统能够在再次遭遇同样抗原甚至其变异种类后仍能快速反应并反击抗原，这个过程称为二次免疫应答。对引起初始免疫反应及造成免疫系统 B 细胞数量迅速增加的抗原而言，二次免疫应答是特异的。当抗原或类似抗原再次入侵时，免疫记忆细胞使免疫系统不用重新生成抗体，效率自然有很大提高。二次应答不仅可以由病原体重新引发，也可以由类似病原体引发，即免疫记忆是联想性的。

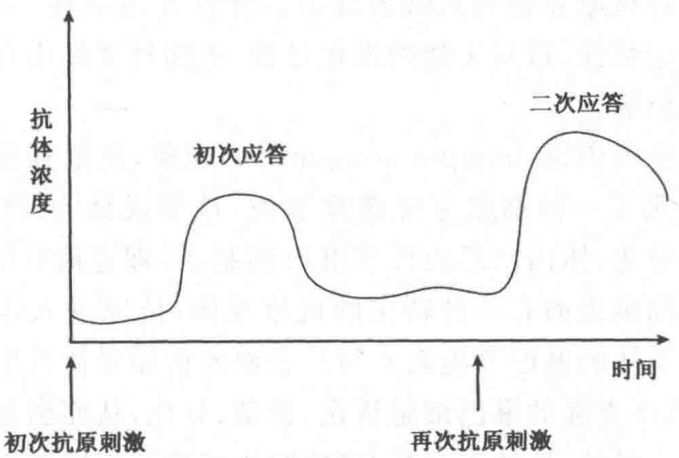

图 1-2　免疫应答过程

出现在二次应答中的抗体一般比初次应答出现的抗体对抗原具有更高的亲和力。这种现象称为免疫应答成熟[4,5]。从初次应答到二次应答的过程中，抗体亲和力的增加表明免疫应答成熟是一个连续过程(强化学习)。这些亲和力更高的抗体被选入记忆细胞池中。该现象已经用于设计人工免疫算法，解决优化问题。初次应答过程可以看作免疫系统学习、识别、记忆外部入侵物质特性的过程，二次应答过程则可看作免疫系统利用免疫记忆杀死入侵物质的过程。免疫应答及记忆是许多人工免疫系统模型的基础。

(2)免疫识别

现代免疫学认为，机体免疫功能是对抗原刺激的应答，而免

疫应答又表现为免疫系统识别自己和排除非己的能力。免疫系统在发挥免疫功能的过程中，识别是重要的前提，一切生物都具有这种能力。单细胞生物只具有分辨食物、入侵微生物和本身细胞成分等低级的识别功能。脊椎动物的机体免疫系统逐渐完善，不仅具有完整的免疫器官和免疫细胞，而且免疫活性细胞还能产生特异性抗体和淋巴因子，从而准确地识别自己、排除异物，以达到机体内环境的相对稳定，这对保护自己、延续种族和生物进化都有重大意义。高等生物充分发展的免疫系统对内外环境的各种抗原异物刺激既表现出多样性和适应性，又表现出特异性和记忆性，这对生物的进化过程、生物种系的生存和适应具有重大影响。

对于免疫识别(immune recognition)现象，奥地利免疫学家Burnet提出了一种细胞克隆选择学说，该学说认为：胚胎期由于细胞的分化，体内已形成许多淋巴细胞系，即克隆(clone)，每一个淋巴细胞表面有一种特定的抗原受体；抗原进入体内选择具有相应受体的淋巴细胞系并与其表面的抗原受体发生特异性结合，导致该克隆的淋巴细胞活化、繁殖、分化，从而引起特异性免疫应答。另外，抗体本身具有“抗原决定簇”，它能够被机体本身产生的其他抗体所识别并引起反应。这样，抗体便具有识别抗原而又被其他抗体所识别的双重性[6]。

(3)免疫耐受

免疫耐受(immune tolerance)是指免疫活性细胞接触抗原性物质时所表现的一种特异性的无应答状态。它是免疫应答的另一种重要类型，也是机体免疫调节的内容之一，其表现与正向免疫应答相反，也与各种非特异性的免疫抑制不同，后者无抗原特异性，对各种抗原均呈现无应答或低应答。

免疫耐受现象是抗原诱导的专一性淋巴细胞功能缺失或死亡，从而导致的机体对该抗原反应功能丧失或无应答的现象。抗原侵入机体后可能导致淋巴细胞的活化，也可能产生免疫耐受，这是淋巴细胞对抗原的识别和应答的两种可能结果。诱导

免疫耐受的抗原称耐受原(tolerogen),而诱导产生正常免疫反应的抗原称为免疫原(immunogen)。正常生理状况下,机体对自身组织抗原是耐受的,这是免疫系统的基本性质,称为自身免疫耐受。如果破坏了这种自身免疫耐受,就会导致自身免疫疾病。诱导对特异性抗原的免疫耐受对自身免疫疾病的治疗、消除器官移植的排异现象及对过敏反应的治疗均有重要意义。因而,免疫耐受与免疫活化是免疫应答过程同一个问题的两个不同侧面[7]。

依据免疫耐受形成的特点,它可以分为天然耐受与获得耐受两种。其中,机体对自身成分不发生免疫应答为天然耐受现象,并称作自身耐受性;而通过人工诱导对机体形成的免疫耐受则为获得耐受。免疫耐受的一般特性主要表现在以下几个方面:①免疫耐受因抗原特异性T或B淋巴细胞被排斥或抑制而具有特异性;②未成熟淋巴细胞比成熟淋巴细胞诱导耐受性容易得多;③诱导与维持耐受性需要耐受原的持续存在[6]。

(4)免疫记忆

免疫记忆(immune memory)是免疫系统的另一个重要特点。当机体接触过某种抗原后再次接触相同抗原时,则抗体出现的潜伏期较初次应答明显缩短,抗体含量大幅度上升,而且维持时间长。这种当同一种抗原再次入侵机体时,引起的比初次免疫更强、更高亲和度的抗体产生的现象就称为免疫记忆。在体液免疫和细胞免疫中均可发生免疫记忆现象。免疫记忆现象可以解释成对特异性抗原应答的淋巴细胞数量增加的现象[6]。

(5)免疫调节

免疫调节(immune regulation)是指在遗传基因控制下具有增强作用和抑制作用的免疫细胞和分子的相互制约及相互调节,共同调控免疫应答的强度和正、负方向[8]。免疫调节存在于免疫应答的全过程,控制着免疫应答的发生、发展和消退。免疫调节机制能将免疫应答的强度限定在一定水平上,它通过对抗体的抑制和促进作用控制机体内抗体的浓度,自我调节(由T

淋巴细胞进行调节)产生适当数量的必要抗体,避免正常细胞受到损害,使免疫应答过程协调进行,从而维持机体内环境的稳定性。

1.2 人工免疫系统原理及模型

目前人工免疫系统的主要算法和模型有阴性选择算法、克隆选择算法和免疫网络模型等,它们可以为免疫应答过程中的各种免疫现象和免疫机制提供理论解释。其中,阴性选择算法由于其不需要先验知识,能够利用少量的正常样本检测出无限的异常样本的独特特点,因此已经成为人工免疫系统的核心。

1.2.1 阴性选择理论

阴性选择算法由免疫系统的阴性选择机制启发,最初由Forrest于1994年提出[9]。生物免疫系统具有非常复杂的防御机制,能够检测外来物质,产生抗体,攻击抗原。这种能力主要由免疫系统的B淋巴细胞和T淋巴细胞实现。淋巴细胞的阴性选择描述这样一个过程:淋巴细胞与抗原相互作用导致该淋巴细胞的死亡(或无反应能力),从系统中清除这种T细胞或者B细胞,其目的是提高自体细胞的耐受性。

未成熟的T细胞在胸腺中发展,如果在发展期间被激活,它们通过程序性细胞死亡(programmed cell death)而死亡。而更多的自身蛋白质在胸腺中表达,所以存活的T细胞继续发育成熟并离开胸腺,对所有那些自体蛋白质耐受,该过程称为T细胞阴性选择,因为未激活的T细胞才会存活。在阴性选择过程中,95%的T细胞会被清除,被清除的T细胞会强烈地对自体反应或根本不对自体反应。换句话说,选择机制有效地加速了系统融合新类型的免疫细胞。但阴性选择并不完善,一些自

体反应 T 细胞逃脱并进入胸腺外围成为完全的免疫活性细胞，增加了自身免疫疾病的威胁。

除了淋巴细胞受体结构中大量的随机成分，淋巴细胞抗体和抗原决定基之间的遭遇不可避免地激活淋巴细胞，但可能造成其死亡或者钝化(无反应力)。因此，阴性选择可以防止自身特异淋巴细胞成为攻击细胞。

阴性选择算法可分为 3 个阶段：定义自体、产生检测器、监测异常的发生。但一般将定义自体与产生检测器作为一个阶段，由此大致可分为产生和检测两个阶段[10]。在产生阶段，给定一个称为自体集合 S 的已知模式集合，测试 T 细胞受体识别和结合自体模式的能力。如果 T 细胞受体识别来自自体集合的字符串，则忽略它；否则，作为一个竞争细胞进入检测器集合。检测器可用于监测系统的异常变化[11]。阴性选择算法的具体步骤(如图 1-3 所示)如下[12]：

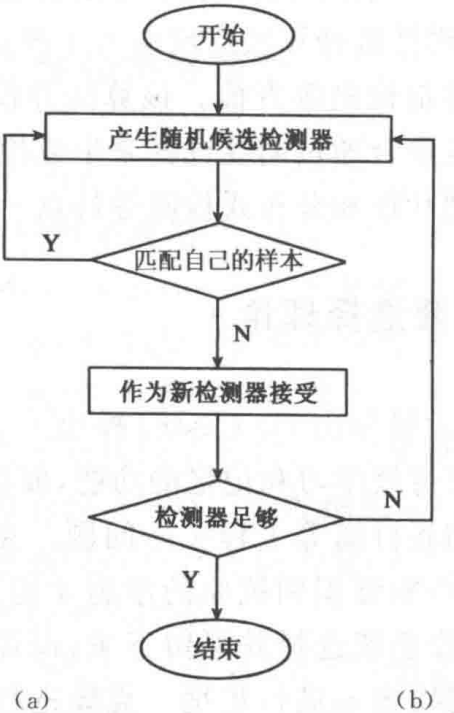

图 1-3　阴性选择算法示意图

(a)检测器集合产生；(b)新情况检测