

中华博士 园地

这是本刊特为海内外正在就读和学成立业的博士、博士后青年学者们开辟的一片科普园地。深学浅著,是一门德识、慧学、素质修养的学问。你们的新知识、新调研、新观察、新目光、新展望,能够用尽可能深入浅出、通俗流畅的语言,汇报给祖国、人民、家乡父老子弟乡亲们吗?中华博士园地,乃耕耘忠孝之地、科教兴国、民族昌盛之地。要用慈母听得懂的语言,写出你们的心声!

人工生命与基因组可视化

刘健勤

1 源于自然的计算

光阴似箭,日月如梭,随着时间之箭,自然界这个宏大的系统饱经了沧海桑田般的变迁,其间最为奇妙和壮观的现象莫过于生命。人类探索生命奥秘的努力包含了生命科学、系统科学、信息科学等许多学科的工作。1987年美国科学家兰顿提出了人工生命的概念,由此这个被称为人工生命的新兴学

科得以萌发。时至今日,从整个学科的众多浪花中不难体会其源头的潺潺涓流。生命现象的奥秘和生物的自适应行为一直被科学家们所探求,遗传和进化是两个重要因素,达尔文、华莱士、魏茨曼、孟德尔等在历史长河中留下了足迹,当世界进入20世纪时,又出现了杜布赞斯基、莫诺、迈尔、摩尔根等大师,沃森和克里克对DNA双螺旋结构的发现,使得人类对生命的视野获得了巨大的拓展,分子生物学成为当代生命科学的必备工具,令世人瞩目的“人类基因组计划”是一项面向21世纪的、由全世界各国科学家协作研究的巨大工程项目。直接来自生命科学的启示与人类建造自动化机器的实践相结合,孕育并诞生了人工生命的世界。人工生命是继人工智能、神经网络之后的又一门新兴的交叉性前沿学科,在美国、日本、西欧等发达国家中该领域的研究已形成高潮,被认为是跨世纪的重大研究课题。人工生命这个学科历经10年的发展,已在国际学术界获得了独立学科的地位。

人工生命是通过人工方式和相应技术手段,去认识、描述、建模、分析乃至构造抽象意义下具有生命形式化特征的机器系统的科学。目前该学科在

刘健勤 副教授,博士。1994年9月至1995年9月为日本理光(RICOH)公司研究开发本部信息与通信研究所客座研究员,现为中南工业大学信息工程学院智能控制研究所副所长、中国有色金属工业总公司跨世纪学术带头人。代表性的学术专著为《人工生命理论及其应用》(冶金工业出版社,1997年),并在国内外发表论文70余篇。研究方向为人工生命、人类基因组、生物信息工程。



理论、方法和技术研究的同时，已逐渐走向针对实际对象的应用领域。相对于地球上出现原始生命以来漫长的历史进程，人工生命学科显得是如此地年轻；但是在面临世纪之交这个时刻，它又是那样地充满着活力和希望。

2 从“生命之如其所能”到“生命之如吾所识”

人工生命中有两句著名的格言：“生命之如其所能”和“生命之如吾所识”。人工生命的核心工作就是实现将前者与后者相连接的桥梁。就象自然界的生命系统是动态的那样，人工生命的体系框架是适于放在空间、时间、逻辑这三维空间中加以考查的，如图 1 所示。鉴于目前人工生命理论的全部体系还未建立，在图 1 中给出有代表性的几个理论逻辑点，以表明学科的内在联系。

到目前为止，人工生命学术界的 6 次里程碑式

的国际学术会议，构成了该学科发展轨迹在时间维上的重要坐标点。人工生命国际会议的名称采用了 ALIFE 加上罗马数字的表示，这个简称已成惯例，沿用至今（第一届的名称不加 I 的标记）。现将这 6 次学术会议简要评述如下：

(1) ALIFE (1987 年 9 月在美国新墨西哥州洛斯阿拉莫斯召开)：本次会议的论文集共收录了 24 篇论文，内容分布在人工生命研究的理论、生命现象的仿真、细胞自动机、遗传算法、进化仿真等 5 个方面。兰顿发表了题为“人工生命”的开拓性论文，他在文中提出了人工生命的概念，并讨论了它作为一门新兴学科存在的意义。兰顿被公认为人工生命学科的创立者。该论文集也包括了理查德·道金斯的“可进化能力的进化”这篇该学科的经典研究论文，理查德·道金斯是英国的生物学家，著有《自私的基因》和《盲目的钟表匠》，这两本著作在国际学术界产生了很大的影响。这次会议标志着人工生命学科的诞生。

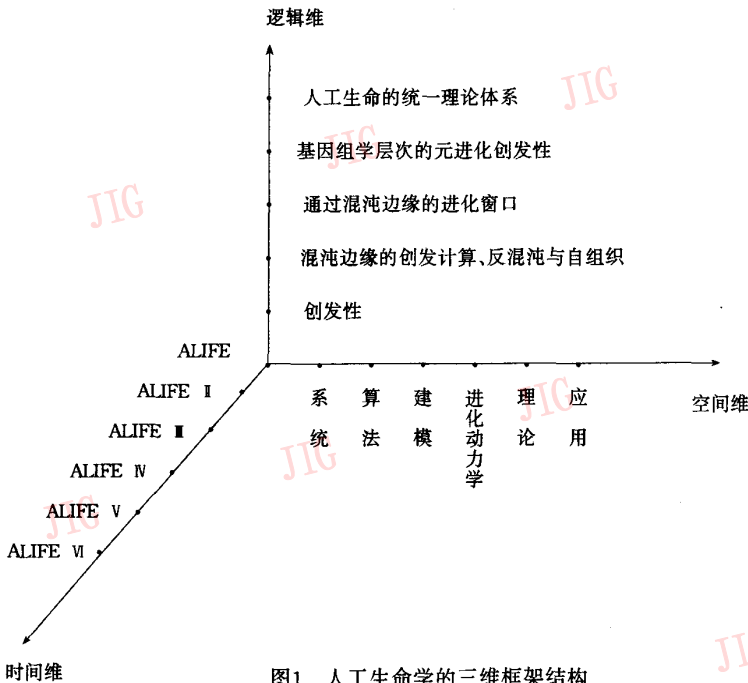


图1 人工生命学的三维框架结构

(2) ALIFE II (1990 年 2 月在美国新墨西哥州圣菲市举行)：本次会议论文集共收录了 31 篇论文，内容分为概貌、来源/自组织、进化动力学、开发、学习与进化、计算、哲学/创发、未来这 8 个部分。该论文集的扉页上登载了阿里斯蒂德·林登梅耶(1925~1989)生前的照片和怀念他的文字。在照片中他手握一株植物；众所周知，他所创立的 L-

系统已成为人工生命研究的一种成功的工具，并对植物形态的计算机模拟工作产生了深远的影响，在计算机图形学方面具有实际应用价值。经典性的论文主要有兰顿的“混沌边缘的生命”、约翰·R·科赞的“计算机程序的遗传进化与共同进化”等，戴维·阿克雷和迈克尔·利特曼的“学习与进化之间的交互”对从计算机算法角度理解鲍德温效应具有参

考价值。1996年时逢鲍德温效应发表100周年,国际《进化计算》曾组织了专辑加以讨论这个至今仍有争议的问题。

(3) ALIFE III (1992年6月在美国新墨西哥州圣菲市举行): 本次会议的论文集共收录了26篇论文, 内容涉及遗传算法、进化仿真、创发行为、适应度概貌图、群体动力学和混沌机制等。科赞的“人工生命: 自复制和改进计算机程序的自发创发性”和金子邦彦、铃木俊二的“一个模仿游戏中通向混沌边缘的进化”这两篇论文可堪称杰作, 分别从遗传编程算法和混沌学两个方面探讨了在人工生命研究中关键的创发机理。

(4) ALIFE IV (1994年7月在美国马萨诸塞州坎布里奇市麻省理工学院举行): 本次会议的论文集共收录了56篇论文, 内容分为特邀报告、长文、短文3个部分, 它覆盖了共同进化、遗传算子、进化与其它技术的综合、人工生命算法、关于混沌边缘和分岔的研究、人工生命建模、鲍德温效应、学习能力、进化动力学、细胞自动机、DNA非均衡学说研究、人工生命在字符识别、机器人等方面的应用等较为广泛的内容。其中土居洋文、和田健之介、古泽满的“源于半保守性DNA复制的非对称突变: 双束DNA型遗传算法”一文介绍了“DNA非均衡假说”方面的最新进展。

(5) ALIFE V (1996年5月在日本古城奈良举行): 这次会议的承办单位是日本著名的研究机构“国际电气通信基础技术研究所”(简称ATR), 该研究所于1985年开始建设、预计21世纪初完成的关西文化学术研究都市, 与位于日本东部的筑波科学城遥相呼应。ATR的进化系统研究室在人工生命研究方面取得了引人注目的成果, 著名学者有下原胜宪、雨果·德·加里斯、托马斯·雷等。下原胜宪曾在ALIFE IV上作了题为“针对脑通信的进化系统——走向人工脑”的专题报告, 该报告简要概括了在进化系统研究室所进行的主要研究工作。这次会议是人工生命学科在第一个10年时间坐标点上的里程碑, ALIFE V是该学科逐步走向成熟的重要标志, 这次是第一次、也是目前唯一的在美国以外的地方召开的ALIFE系列会议。

(6) ALIFE VI (将于1998年6月26日至6月29日在美国加利福尼亚州洛杉矶市举行): 目前该会议的稿件工作已基本完成, 主办单位是美国加州大学洛杉矶分校(UCLA)。ALIFE VI最为显著的特征是人工生命与生命科学的结合, 努力探索学科前

沿的重大课题, 会议的主题是: “生命与计算: 变化中的疆界”, 主要内容包括下列7个方面: DNA计算、基因组数据采掘、计算化学、生物技术中的组合与进化方法、关于遗传、发育、行为、生态的建模、外星生物学(例如火星陨石ALH84001中生物活动所产生的踪迹)和基于代理(agent)的WWW媒体、通信和经济学。有兴趣的读者可以参看网址: <http://alife6.alife.org/>。

3 基因组生存

生命科学已进入了被称为“基因组学”或“功能基因组学”的时代, 与之相对应人工机器系统研究的思想方面也有达尔文机、细胞自动机机器以及莱昂纳德·阿德勒曼的DNA计算机理论。在国际互联网络使人们得以跨越时空的尺度共享巨大的信息资源时, 尼葛洛庞地这位麻省理工学院媒体实验室的学者在传播他关于“数字化生存”的思想时特别强调信息DNA的观点。人工生命作为新兴学科就应该努力尝试去解决那些传统(现存)技术不能或不能很好解决的问题。人类基因组包括了30亿个碱基对序列的信息, 其结构与功能的复杂关系是决定生命活动的关键因素之一, 人类基因组计划的意义是巨大和深远的, 而水稻基因组计划则属于对国民经济发展具有直接作用的科学实践活动。由于基因组学是所有生命科学的基础之一, 所以生物信息学是多学科交叉的非常前沿的新领域。

众所周知, 基因组学在很大程度上改变了人类认识自己的方式, 特别是人类基因组计划不仅在探索自然规律方面, 而且在医疗、制药、生物技术方面具有广阔的应用前景和社会经济效益。就国际人类基因组计划在2003年完成对30亿核苷酸对序列分析这个任务来说, 其所需的高科技手段对自动化的程度要求很高, 例如美国麻省理工学院人类基因组中心所实现的系统所针对的是从加样操作到结果输出的数据处理过程, 每次所完成的分析可达15万个。国际上Affynretrix公司开发了面向大规律并行基因组学的DNA序列分析技术, 美国的Sequatron系统和相关的High-Throughput技术也值得注意。就基因组学工作的核心内容来说, 生物信息学的手段是关键。1997年1月17日国际权威刊物《科学》就以“生物信息学”为题安排了3篇文章介绍该方面的几项重要进展。人工生命由于在数学和计算机技术方面的作用, 已在生物学和生态系统方

面取得了一些应用, 人工生命模型化计算形式具有对生命现象进行探索的辅助工具作用。

完整基因组的获得、致病基因和相关基因组病理学分析等环节均依赖于基因组功能辨识, 连锁图、物理图、转录图、序列图诸层次的有机联系为工作的展开提供了逻辑关系, 就遗传学意义下的分散性而言, 基因型与表现型的非线性高维模式关系又是定性定量计算相综合的途径, 显然, 基因组序列测定与功能分析中, 对象信息处理的非线性非确定性又要求相应的定位、克隆等工具具有很大的经验性, 如何采用人工生命建模和辨识的计算机高新技术手段来进行人类基因组的工作, 是很有发展前途的技术手段和开发途径。

4 可视化的基因组学结构

就基因组功能而言, 蛋白组 (proteome) 是重要的结构环节, 蛋白质由于分子生化作用而形成的一级、二级、三级、四级结构, 具有很强的非线性编码能力, 与此对应的数学工具可采用 Knot 理论, 以充分描述相应符号化形式系统中的抽象代数结构。这样所需处理的信息属于高维模式空间中的复杂时空非线性系统, 粗粒度的符号动力学描述是基因组结构可视化的必要途径之一。因为它是动态的过程, 所以必须选用能有效地体现生命特征的形式。人工生命在发展初期就与图形学的视觉手段有紧密联系。如果读者看过河口洋一郎的计算机图形学作品, 一定会留下深刻的印象。如果说人工生命中

著名的鸟类群体行为演示系统不够复杂的话, 那么美国麻省理工学院媒体实验室永久性安装的 ALIVE 系统, 则是一种能为人类参加者和动物化自律智能代理体所处的虚拟世界之间提供无线全身体交互的虚拟环境系统。由于使用了视频摄像机和投影屏幕, 参加者不需要眼镜、手套、导线等工具就可与虚拟世界中的对象进行交互操作。该系统形象生动、直观有趣, 不仅是多功能可视化的, 而且是多自由度可“动”(操作)的。(美国伊利诺依大学国立超级计算应用中心 (NCSA) 在 1996 年 6 月开发出被称为“生物学工作台”的 Java 方式下交互式软件, 用户可通过 Internet 访问。其网址为: <http://biology.ncsa.uiuc.edu/BW/>。美国《科学》在 1996 年 8 月曾专门刊载文章加以介绍。从计算机信息技术的角度看, 这是目前同类网络环境下的一个出色的基因组可视化软件。

基因组与蛋白组所耦合的生命过程和结构中, 最重要的动态核心属性就是创发性。到目前为止, 人工生命创发性研究方面的主要理论有兰顿的混沌边缘理论、金子邦彦的混沌边缘进化窗口理论、考夫曼的反混沌 (Anti-chaos) 理论、刘健勤的基因组学元进化创发性理论。这些努力均是对人工生命统一理论建立所进行的有益尝试。基因组学元进化创发性研究工作的目的在于实现针对连锁图、物理图、序列图、转录图高维关联信息中元进化创发性的计算模型和基因组学高维模式结构的计算机辅助分析软件, 相关的符号动力学工具与粗糙集数学理论相结合, 以构造弱混沌约束下的文法模式发现。

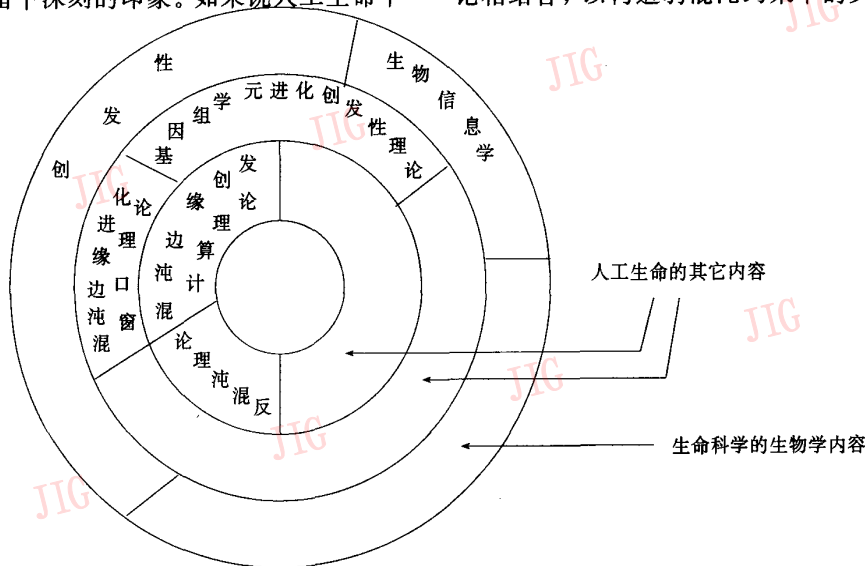


图2 理论要素之间的关系

科学事实表明创发性的根源可由混沌机制予以解释和分析,并可立足于进化的算法层次进而涉及到元进化机制的基因组学层次的实质。有关理论的联系可用图2的环状3层嵌套结构来示意,基因组学元进化创发性理论中基因组分子不对称时空耦合结构的功能蛋白组环模L——符号反演是主要形式之一,符号化模式表达是解决高维空间下可视化表达的有效手段,其描述的逻辑性强、信息压缩程度也较高。

5 结 语

包括图象图形的视觉信息处理,是与生命科学有着一定关系的,视觉信息的生物神经网络机制为计算机视觉和模式识别工作提供了借鉴,人工智能系统的建造也仿真了某些认知和感知(例如视觉)的功能;采用模式识别方法识别染色体形状已是图象处理领域内被人们所熟知的事例。人类在接收信息时视觉形式占了很大比重,对一台机器的完整说明中应包括一张(或一套)图纸;生物确实不能简单地同机器相比,但是30亿个符号组成的基因组序列着实为一部“天书”,对其“破译”的工作是很艰巨的,基因组可视化的计算机辅助工具这个思路就应运而产生了,这既是很自然的,也是必需的。

在生物学研究中,象人类(Homo sapiens)的生理和神经机理认识,经常通过相对简单的生命形式(如果蝇等)来对照,其主要依据就是所分析的生物与模式生物之间在基因方面的同源性等属性。在与

中风相关的CADASIL的Notch3变异研究中,人和老鼠的部分序列对比就可提供有用的信息。在基因组数据的计算机分析工作中,连锁基因、跳动基因、甚至Junk基因等因素均应在数据库构造时加以考虑。1997年7月和9月分别获得了幽门螺杆菌(Helicobacter pylori)和大肠感菌(Escherichia coli)的完整基因组序列,说明了某些计算机方法是有一定的启发和借鉴作用的,例如模式匹配、训练、数据库、预测、数学的马尔科夫模型等,国际上相关的软件有GeneMark、GeneSmith、MacPattern、FASTA、Signal-P、RepScan、TopPred等。从本质上来讲,这里整个计算方法的核心就是模式识别和系统辨识。

视觉不仅是我们人类而且是多数动物所具有的本能之一,它的存在使得我们在认识自然、探索自然规律时有了一种形象化的信息方式,同样借助于人工生命的理论和图象图形学的可视化技术,涉及生命、信息、智能等因素的多学科交叉前沿领域的最新技术,将在人类实现探究自然奥秘宏愿的征途上发挥其特有的科学魅力。

6 后 记

本课题得到国家教委留学回国人员科研启动基金、国家教委图象信息处理与智能控制开放实验室基金、中国有色金属工业总公司跨世纪人才基金和湖南省自然科学基金资助。

柯达与英特尔共促数字摄影技术

柯达和英特尔在营销和开发上主动联合,目标是拓宽客户使用数字技术进行摄影的途径。作为协议的一部分,柯达计划年末使用新的扫描设备来改善它的Qualex处理实验室,这样客户就可以把他们的照相底片放在光盘驱动器上处理。Kodak Picture CDs使客户能在家中的PC机上处理照片,但相应的软件和文件格式还没有最后确定。两家公司计划

于一两年内推出新的数字相机,这种相机将能支持英特尔的互补金属氧化物半导体(CMOS)传感技术,英特尔希望该技术能取代当今数字相机所采用的电荷耦合器件(CCD)技术标准。

柯达公司总裁兼首席执行官George Fisher认为,通过这样一种新方法使用照片,将开创该产业的新纪元。