

R 语言在医学院校生物信息学实验教学中的应用与探索 *

阎 洁 杨俊丽 王建文 杨 帆

(山西医科大学计算机教学部 太原 030001)

〔摘要〕 分析生物信息学及其研究工具 R 语言的特点, 以山西医科大学为例阐述生物信息学实验教学现状以及以学生为中心的实验教学方案设计, 实践表明该方案对提高学生成绩具有一定作用。

〔关键词〕 生物信息学; R 语言; 实验教学

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.01.018

Application and Exploration of R Language in Experimental Teaching of Bioinformatics in Medical Colleges YAN Jie, YANG Junli, WANG Jianwen, YANG Fan, Computer Teaching Department of Shanxi Medical University, Taiyuan 030001, China

〔Abstract〕 The paper analyzes the characteristics of bioinformatics and the study tool R language. Taking Shanxi Medical University as an example, it expounds on the current status of experimental teaching of bioinformatics and the design of student-centered experimental teaching scheme. Practice shows that the scheme has a certain effect on improving students' performance.

〔Keywords〕 bioinformatics; R language; experimental teaching

1 引言

生物信息学是综合计算机科学、数学、生物学等多门学科理论及方法, 通过对研究对象相关数据的采集、处理、计算模拟和分析认识其本质, 揭示其内在机理的学科^[1-2]。目前生物信息学主要应用

于生物科学研究中蛋白质组、代谢组、基因组、转录组等各种组学数据的分析和处理。随着各学科数据的积累, 运用数据揭示学科中某一类共性问题的本质具有重要意义。医学作为以人体复杂系统为研究对象的科学, 在解决具体病症时常需要综合考虑和分析引发某一病症的多维数据^[3]。在具体实践中生物信息学综合运用计算机科学、数学对已有的诊断和治疗数据进行深入分析, 对相关病症的准确诊断及治疗方案的合理制定和优化具有重要意义^[4-5]。生物信息学技术其具有的通过对数据采集、处理、分析以确定研究对象本质的特点, 在医学领域中的应用得到逐步推广, 各大医学院校相继开设生物信息学课程作为医学专业本科生的必修课或选修课, 以培养学生运用现代新技术, 提高其在医学实践中准确诊断病症和合理制定治疗方案的能力^[6-7]。

〔收稿日期〕 2019-06-25

〔作者简介〕 阎洁, 讲师, 发表论文 4 篇。

〔基金项目〕 山西省高等学校教学改革创新项目“医学院校交叉学科生物信息学课程建设”(项目编号: J2017043); 山西医科大学教学改革基金资助项目“基于计算思维的 R 语言生物医学数据处理教学研究与实践”(项目编号: XJ2018102)。

随着生物信息学所处理数据的数据量和复杂性增加,需要一种功能全面的专业语言和软件工具,R 语言较传统统计软件具有更好的可编程性、体积小、功能强大及易于扩展的特性,在现代生物信息学大数据分析中得到广泛应用^[8-9]。目前各医学院校为适应大数据时代医学生基于已有诊疗数据进行诊断分析等需求,所开设的生物信息学课程大多以 R 语言作为相关实验教学工具。因此如何有效地结合 R 语言设计生物信息学课程实验内容并有效开展,对提高生物信息学课程实验教学效果具有重要意义。本文在借助文献分析生物信息学及 R 语言特点的基础上,结合医学院校学生在实际工作和研究中对生物信息学的应用需求特点,探讨基于 R 语言的生物信息学课程实验设计并结合教学实践评价实验设计效果。

2 生物信息学特点^[3-6]

2.1 生物医学数据类型复杂,数据量庞大

生物医学数据特点由生命本质所决定,生命体内部运行机理、生理生化反应以及病理变化等是其复杂性的内因。得益于高通量生物技术如二代测序技术的发展,生物医学数据呈现出爆炸性增长趋势。

2.2 学科交叉性强,覆盖面广,综合应用难度大

生物信息学是基于分子生物学和多学科交叉而形成的一门前沿学科,在生物和医学领域有较广泛的研究和应用,国内外众多高校相继开设生物信息学课程。医学院校学生虽然对生物化学、分子生物学、生理学、遗传学、细胞生物学等学科基础知识的掌握较为全面和扎实,但生物信息学课程理论性较强、覆盖面较广,学生认为其与其他生物医学课程结合不够紧密,较难将生物信息学的知识和原理融入到现有的知识体系当中。

2.3 以计算技术为主要工具,算法抽象

生物信息学主要借助于统计算法分析生物大数据所包含的生物学意义,要求学生有一定的计算机

技术和数学基础,能够结合计算机编程和算法实现对具体问题的分析。目前在生物信息学相关数据统计和处理中常用的工具有 Matlab 和开源的 R 语言,由于 Matlab 属于商业收费软件,在实际应用尤其是高校教学中受到一定限制;R 语言因其自由、免费及源代码开放等优势得到了广泛应用。

3 R 语言特点

3.1 概述

生物信息学主要借助计算机,通过统计学方法分析生物学数据以揭示其中包含的内在信息,因此需要一种功能强大的数据处理和分析软件作为工具。R 语言具有强大的数据分析能力、简捷高效的编程及用户自定义功能。

3.2 主要特点^[8-9]

一是自由。R 语言编译软件免费且源代码开放,学习者可在其镜像或学习网站免费下载安装包、源代码及相关学习资料。二是可编程性。R 语言与其他统计学软件如 SAS, Matlab 等相比语法更加通俗易懂且允许使用者编制个人函数以扩展现有语言。三是更新快。R 语言软件标准安装包中包含基本常用的程序包,且随着新算法的出现,新版本中会及时更新新算法对应的程序包。四是互动性强。R 语言可以实现同窗口输入和输出,且兼容多种图片格式的图形文件输出。五是跨平台性。R 语言软件包可在 Windows、Linux 及 Mac OS 等系统中兼容运行。此外 R 语言具有强大的数据分析和处理、整合及图形图像显示功能,对强调训练学生统计、数据分析等能力的专业优势明显。

3.3 应用

R 语言应用于生物信息学教学实践中不仅可以很好地结合理论、案例分析,还可极大地调动学生学习积极性,因此 R 语言在生物信息学相关课程的实践教学中得到广泛应用。目前国外众多大学统计及生物医学和环境科学专业相关课程都将 R 语言作为教学工具软件。在结合生物信息学课程及 R 语言

特点的基础上，本研究开展基于医学生的生物信息学实验设计和教学工作，通过以学生为中心的实验教学

4 基于 R 语言的生物信息学实验教学设计

4.1 现状

一是教学内容设置宽泛。由于生物信息学理论知识覆盖面广等特点，目前在实验教学的开展中重点不聚焦、不突出，实验课教学内容、实施方式、教学大纲及实验指导书内容不完善，难以形成具有针对性的教学案例。二是教学方式单一，缺乏培养学生解决问题能力的理念。实验课内容设置较为简单或抽象，常演变为验证性实验或单纯执行教师提供的代码，缺少分析、解决问题的思路

4.2 设计

4.2.1 教学计划 山西医科大学从 2016 年开始为生物制药专业本科生开设生物信息学课程，针对开课专业学生的知识背景及培养目标，结合生物信息学课程特点，课程团队结合参考教材、教学经验、学术科研和学生专业背景等因素对生物信息学实验进行综合设计。根据实验教学的 12 个学时进行教学内容的划分和整合，设计生物信息学教学计划，见表 1。

表 1 生物信息学实验教学计划

序号	核心知识内容	实验类型	学时
1	R 语言基础	演示 + 验证型	2
2	R 语言应用案例	演示 + 验证型	2
3	生物医学数据获取、序列分析之序列比对	验证 + 设计型	2
4	系统发生分析之系统发育树、蛋白质结果预测与分析	验证 + 设计型	2
5	疾病相关蛋白质的结构分析、针对性药物的模拟设计	综合型	2
6	某药物的生物信息分析	综合型	2

4.2.2 特点 一是以学生为中心的教学设计。使学生充分利用文献检索、网络教学视频、多媒体和技术等学习方式，在掌握基本理论知识的基础上通

过自主探索、分析解决实际问题，进一步培养学生学习积极性和主动性，加深对所学内容的理解和应用。二是多种教学模式相结合下以 R 语言为工具的实验模式设计。根据生物信息学课程特点和 R 语言功能，采用“教师讲授 + 案例演示 + 实验实践 + 基于问题学习 + 个性化自学”的生物信息学实验教学模式，以医学相关专业学生在具体工作实践中的需求为导向，通过实验培养学生自主能动性和团队协作能力。三是基于网络平台的拓展训练。针对实验教学学时有限，教学内容从广度和深度均不能满足具体专业学生解决实际问题需求的情况，教学团队在课堂实验的基础上设计基于 E-learning 教学平台的实验拓展训练，平台整合课程教学课件、相关专业实验数据、实验手册供学生课后练习，进一步巩固其对相关内容的掌握。

4.2.3 实验实例 采用 R 语言扩展包 ggplot2 中的函数对生物数据鸢尾花数据集进行分类。所分析数据包含 150 个数据集，分为 3 类，每类含 50 条数据，每条数据包含 4 个属性。可通过花萼长度和宽度属性预测鸢尾花卉的具体类别，分类结果，见图 1。

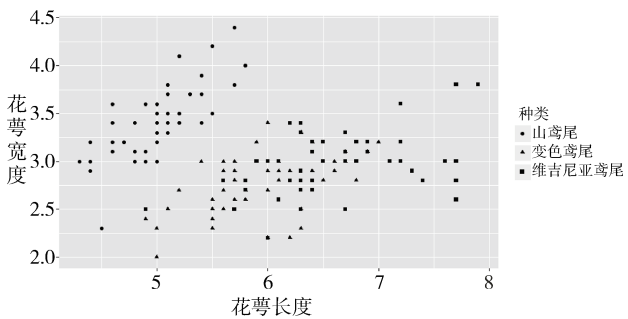


图 1 鸢尾花数据集分类结果

4.2.4 成效 基于上述实验教学方案的改革和设计，经 2017、2018 级生物制造专业生物信息学实验教学实践，表明两届学生期末实践能力考核成绩较 2016 级分别提高 3.5% 和 4.25%，进一步说明生物信息学实验实践训练及合理设计实验的重要性。

5 结语

生物研究领域 中基因测序技术的发展极大地促
(下转第 86 页)

的阅读兴趣。在内容上要扩展范围,不仅局限于专业医学信息,可以更贴近生活甚至扩大推送范围——面向患者及家属群体,与门诊或病区联合,发布简单的患者教育信息。全面利用微信平台的新媒体技术,开发本馆微信小程序,增加使用微信定位服务,通过“附近的人”、“附近的小程序”功能,扩大推广、宣传效果。

4.3 实现信息互动共享

微信平台打破时间、区域的限制,通过微信公众平台不仅可以搭建本馆读者虚拟的交流和学习平台,还可以跨区域组织用户开展各类交流与讨论^[4]。建立信息推送服务反馈与评价机制,实时跟踪读者对信息内容的反馈与评价并及时分析总结^[5]。同时还可以举办线上活动、交流论坛等,从单方面推送信息扩展到与读者群体的双边互动,实现与读者的实时交流,完成信息互动共享。

5 结语

在新媒体、新技术环境下,信息资源千变万化,读者的阅读习惯多种多样,医院图书馆作为读者获取医学信息的重要部门,必须紧跟信息时代潮流,深刻认识全媒体时代的挑战和机遇,用新的理念改进自身服务内容和方式。微信公众平台作为传播媒介,医院图书馆应予以充分利用,为读者提供更丰富的个性化服务,在图书馆知识服务中保持活力,推动医院图书馆的发展。

参考文献

- 1 陈新颜,刘敏.以人为本与图书馆服务及管理创新[J].图书馆学研究,2002(2):12-14.
- 2 张晓东,解金兰.新技术驱动下图书馆服务模式变革与启示[J].图书馆工作与研究,2018(7):97-101.
- 3 张玲.新媒体技术环境下图书馆宣传推广策略[J].图书情报工作,2015,59(1):78-81.
- 4 冯留燕,冯双燕,刘青.微信公众平台在医院图书馆的应用现状调查[J].中国中医药图书情报杂志,2016,40(6):26-30.
- 5 肖频.网络环境下医院图书馆的个性化信息服务[J].中华医学图书情报杂志,2013,22(8):57-58,74.

(上接第 89 页)

进生物信息学的发展,推动在分子生物水平逐步对生命的起源、进化、遗传和发育的本质进行更加深入和广泛的探索与研究。本文基于生物信息学在具体研究中应用特点及其广泛采用的工具 R 语言特点,结合山西医科大学生物信息学课程开展实验教学,探讨生物信息学课程实践教学部分实施方案的设计,对提高所学专业学生成绩取得一定效果。

参考文献

- 1 张春霆.生物信息学的现状与展望[J].世界科技研究与发展,2000,22(6):17-20.
- 2 赵屹,谷瑞升,杜生明.生物信息学研究现状及发展趋势[J].医学信息学杂志,2012,33(5):2-6.
- 3 胡德华,方平.生物信息学在医学上的应用及其社会意义

[J].湖南医科大学学报(社会科学版),1999(1):75-79.

- 4 王波,吕筠,李立明.生物医学大数据:现状与展望[J].中华流行病学杂志,2014,35(6):617-620.
- 5 方福德,程书钧,田玲.建设研究型医院促进转化医学发展[J].中国卫生政策研究,2009,2(7):16.
- 6 郭丽,赵杨,柏建岭,等.医学院校生物统计学专业生物信息学教学探索[J].南京医科大学学报:社会科学版,2013,13(5):101-104.
- 7 倪青山,胡福泉,饶贤才,等.医学院校生物信息学实践教学初探[J].基础医学教育,2011,13(6):538-540.
- 8 张哲,张豪.浅谈 R 语言在生物统计学教学中的应用[J].教育教学论坛,2013(27):62-63.
- 9 熊筱晶.R 语言在 Pubmed 数据库文献检索方面的应用[J].医学信息:下旬刊,2009,22(1):42-45.