·教育研究·

以知识图谱为特色的健康数据素养教育体系构建

张靓　周敏　贺培凤　吕艳华
山西医科大学管理学院，太原　030001
通信作者：吕艳华，Email：sxmu@163.com

【摘要】 在了解国内外医学院校健康数据素养课程开设情况的基础上，提出以知识图谱为特色的健康数据素养教育体系，并分别从健康数据的基础知识、知识图谱的基本技术、分析方法及工具、知识图谱的应用、健康数据伦理五方面，为本科生、研究生和医学信息学专业学生构建多层次的健康数据素养教育体系。研究结果对于提升医学生的数据处理能力，推动医学发展具有积极意义。

【关键词】 健康数据素养；　知识图谱；　医学院校；　教育体系
【中图分类号】 R-05
DOI：10.3760/cma.j.issn.2095-1485.2020.01.002

Construction of health data literacy education system characterized by knowledge map

Zhang Liang, Zhou Min, He Peifeng, Lü Yanhua
School of Management, Shanxi Medical University, Taiyuan 030001, China
Corresponding author: Lü Yanhua, Email: sxmu@163.com

【**Abstract**】 Thanks to the comprehension of health data literacy courses in medical colleges and universities at home and abroad, this paper proposed a multi-level health data literacy education system featuring knowledge maps for undergraduates, postgraduates and students majoring in medical informatics, from basic knowledge of health data, basic technology of knowledge maps, analysis methods and tools, application of knowledge maps and health data ethics. The study results have positive significances for improving students' data processing ability in medical colleges and universities, and promoting medical development fundamentally.

【**Key words**】 Health data literacy;　Knowledge map;　Medical college;　Education system
DOI：10.3760/cma.j.issn.2095-1485.2020.01.002

目前，如何享用时代快速发展带来的海量数据，已经成为科研人员提高科研能力必须面对的一个现实问题。在医学领域，临床和实验室数据使得其数据量更为丰富，尤其是“健康大数据与精准医疗”概念的提出，以及“高等学校人工智能创新行动计划”促进人机交互辅助诊断、医学知识图谱构建等技术在医疗领域的高效融合，都极大地推动了医学领域大数据与其他领域大数据的深度融合[1]。这对科研人员提出了更高要求，要求科研人员必须深入了解数据、分析数据和利用数据。由此可以看出，在医学领域，健康数据素养的紧迫性和必要性要远远超过任何一个行业。因此，医学院校构建健康数据素养教育体系，培养有健康数据素养及能力的医学专业人才是顺应当前形势、解决这一需求的有效途径。

1　健康数据素养的概念和内涵

数据素养（data literacy）最早起源于美国，关于数据素养的概念并没有统一定义。近年来，国内一些学者阐述了他们的观点，如华东师范大学的张静波[2]认为，数据素养是科研人员应具备的对数据采集、处理和分析的专业技能以及在数据生产、管理和发布过程中的道德与行为规范，是对信息素养的进一步提高

和深化；上海大学的郝媛玲和沈婷婷[3]认为数据素养是科学数据素养的重要组成部分，是对统计素养与信息素养的延伸和扩展；贵州医科大学的欧阳玲琳[4]认为数据素养侧重于对知识的构建能力，即规范地发现和获取数据、批判地选择和评估数据、科学地管理和处理数据、合理地利用和共享数据的意识和能力。

医学领域的数据量极为丰富，其广泛存在于医学数据库、基因数据、药物研发、疾病控制、电子病历、临床医疗、健康管理、患者行为和社交网络中。因此，与数据素养相比，健康数据素养更具有专业性，更能满足专业人员的需求。本研究认为：健康数据素养的内涵应该包括具备健康数据意识和健康数据敏感性，能够有效且恰当地获取、分析、处理、利用和表现数据，对数据的使用具有批判性思维，在过程中严格遵守医学数据伦理道德规范。健康数据素养教育的战略意义不仅在于掌握庞大的数据信息，更在于对这些医学数据所蕴含的知识价值进行提取。这是目前健康数据素养能力培养的重要基础[4]。

2　国内外高校健康数据素养教育现状

国外医学院校的健康数据素养教学体系已较为完善，覆盖人群有本科、硕士研究生及博士研究生等；课程内容主要围绕数理统计分析、医学领域的信息学基础、计算机及编程方法、数据分析方法、医学数据的应用等。如哈佛医学院[5]开设了统计学和流行病学概论、生物医学信息学、临床信息学、生物医学数据的深度学习、生物医学应用的数据可视化、人口健康大数据创新等；牛津大学医学部[6]开设了统计学与 SPSS、计算化学、R 语言、数据分析、思维导图等；斯坦福医学院[7]开设了生物医学信息学、R 语言、基因组学的统计和机器学习方法、基因组学和生物科学的深度学习、数学模型和医疗决策等；约翰霍普金斯大学医学院[8]开设了健康科学信息学、临床信息学、R 语言的统计分析、电子健康记录分析等。

在国内，健康数据素养教育在高校中发展较晚，且课程分布较为零散。课程内容主要包括数理统计、计算机应用、数据库应用等。如北京大学医学院[9]开设高等数学、线性代数、生物统计学基础、SPSS 软件在医学统计学中的应用、信息技术与程序设计、SPSS 统计分析、生物信息数据库、生物信息方法与工具；复旦大学上海医学院[10]开设了统计论与数理统计、程序设计、生物信息学概论数据库与知识库、人工智能等；中国医科大学[11]开设了线性代数、离散数学、计算机技能、生物信息学、生物信息与数据分析等；中南大学湘雅医学院[12]开设了医学统计学、卫生统计学、SPSS 及其医学应用、生物信息学、数据库应用等。

可见，国内外健康数据素养的课程建设都有其深刻的历史背景和发展特色。国外医学院校更注重数据分析及在医学领域中的应用，而国内大多院校偏重数学理论和编程方法，对数据分析、数据开发应用有所忽略。总体而言，我国所开设的健康数据素养课程缺乏体系性和实践性。众所周知，医学在整个科学进程中发展较为迅速[13]，在学科体系上也更加深入和细致。然而，科学研究能力的提升要求研究人员不仅要有专业能力，更要有宏观思维，这样才能做到兼顾深度和广度地思考问题。在一定程度上说，科学知识图谱可以为科研人员提供这一帮助。因此，可以尝试在目前健康数据素养所开设课程的基础上，利用知识图谱的分析方法、分析工具以及在医学领域的应用来建立其教育体系，以便更好地促进医学科研转化。

3　以知识图谱为特色的健康数据素养主要内容模块

3.1　健康数据的基础知识

健康数据种类繁杂，存储方式不一。按照不同的来源，可将健康数据分为医院医疗数据、区域卫生服务平台数据、医学研究和疾病监测数据、自我量化数据、网络数据和生物信息数据六类。其中包括临床诊疗、电子病历、健康调查、医学数据库等[14]。由于数据类型不同，在使用过程中要选取恰当的数据格式，如文本文件选择 ODF 格式，表格文件选择ASCII 格式，网页选择 XML 或 RDF 格式；医学影像数据的格式为 DICOM 国际标准通用格式；医院信息系统具有其独特的数据库结构和编码规则。

3.2　知识图谱的基本技术

知识图谱是一种显示科学知识发展进程或结构关系的知识管理工具，可以用简化的知识单元结构直观显示结果，最大限度地满足医学专业人才的数据需求。其基本技术主要包括知识抽取、知识融合、知识推理三部分。它是指从医学数据库、电子病历等数据中抽取相关知识，并将不同来源的知识进行整合加工，达到数据、方法、经验以及思想的融合；并通过知识推理，从现有知识中发现新的知识，实现不

表 1　健康数据素养教育体系内容设计

教学对象	教学模块	教学内容
医学本科生	①健康数据基础知识	·了解信息、数据、知识等基本概念特征及关系内涵
		·了解健康数据的特点，如时效性、隐私性、复杂性
		·了解健康数据的来源、类型及格式
	②知识图谱的基本概述和基本工具	·了解知识图谱的基本概念及类型特点
		·学习绘制图谱的基本工具和适用范围，如通用软件 SPSS
	③健康数据的伦理问题	·了解如何规范合理地使用数据，并遵守相关法律法规
		·了解健康数据中的伦理问题，如隐私问题、数据安全问题
医学研究生	①知识图谱的绘制流程及工具	·学习数据预处理，若样本数据过大或分析目的不同，则需要有代表性地抽取
		·掌握知识单元的构建、数据的标准化处理与简化，以及数据的可视化展示，如几何图、冲积图
		·熟悉知识图谱工具的使用方法及功能，如社会网络分析软件 Ucinet、Pajek
	②知识图谱的分析解读	·掌握知识单元的分析方法，如共现分析、共引分析
		·掌握数据标准化处理常用方法，如集合论方法、概率论方法
		·掌握数据简化分析方法，如聚类分析、多维尺度分析等
		·学习图谱解读的分析方法，如历时分析、网络分析
	③知识图谱的临床应用	·学习收集临床数据的方法并进行相关处理分析
		·运用知识图谱技术帮助诊断治疗和预防，了解患者健康状况
医学信息学学生	①健康数据的处理及分析方法	·掌握运用自动化工具如网络爬虫、深度学习、自然语言处理等方法工具对数据进行整理、鉴别、筛选、重组
		·掌握健康数据的挖掘与分析方法
	②知识图谱构建技术	·了解知识图谱的构建方法
		·熟悉构建图谱的基本流程及相关技术，如知识抽取、知识融合、知识推理
		·掌握知识图谱的绘制流程，如数据清洗、构建知识单元、数据标准化及简化、数据可视化
	③知识图谱的分析解读及工具	·掌握知识图谱的分析方法，如社会网络分析、多维尺度分析
		·掌握知识图谱的各类工具及使用方法特征，如统计软件 SPSS、社会网络分析 Ucinet 和 Pajek、图谱绘制分析工具 Citespace 和 VOSviewer
	④知识图谱的医学应用	·了解知识图谱在医学领域的应用
		·运用知识图谱进行临床决策分析、智能诊断和构建中医药知识图谱等

断迭代更新的过程[15]。

3.3　知识图谱的分析方法及工具

知识图谱的合理性和可靠性很大程度上依赖于所用数据的精确性和全面性。首先要进行数据“清洗”，剔除重复及无关数据，利用共现分析、共引分析等方法构建单元关系；再对数据进行标准化处理，用聚类分析、多维尺度分析等分析方法对样本数据进行简化；最终通过星团图、几何图等将知识可视化，并对图谱结果进行解读[16]。常用的分析工具有通用软件 SPSS、Ucinet、Pajek 和专用软件 Citespace、Bibexcel、VOSviewer 等。

3.4　知识图谱的应用

医学知识图谱为医学知识的共享与传播、诊疗手段的规范化，以及临床决策科学化提供了良好的环境和技术基础。如医疗决策支持系统可以根据患者症状描述及化验数据，给出智能诊断、治疗方案推荐及转诊指南；还可以针对医生的诊疗方案进行分析、查漏补缺[15]。谷歌的医疗知识图谱把搜索引擎、知识图谱与在线医疗进行了深度整合，对智能医疗的发展提供了重要的借鉴作用[17]。

3.5　健康数据的伦理问题

在健康数据规模不断扩大的同时，也存在着个人合法权益受侵害的风险和数据安全隐患的问题。其中主要涉及健康数据建设中的信息安全与个人隐私问题、知情同意与个人自主性问题[18]和健康数据共享中所有权、隐私权、知识产权[19]等伦理问题。

4　以知识图谱为特色构建多层次健康数据素养教育体系

根据上述内容模块，依照不同学科背景、层次阶

段,分别为医学本科生、医学研究生和医学信息学学生构建健康数据素养的教育体系。使其具备对健康数据的敏感性及数据处理分析应用等能力,遵守医学道德准则,能够在知识图谱的思维基础上,全面、迅速、高效地理解与挖掘数据背后蕴藏的有价值信息并应用于科研学习中。

表 1 中划分的三个层次分别为健康数据素养教育体系的三个阶段,即初级阶段、进阶阶段、高级阶段[20]。医学本科生培养属于基础性课程,主要用于树立健康数据道德伦理意识,同时了解关于健康数据和知识图谱的相关基础知识和工具。医学研究生培养则较为深入,主要培养其结合知识图谱对数据进行分析解读的能力。其中涉及知识图谱的绘制流程、分析方法及分析工具。通过介绍分析方法及工具,旨在培养医学研究生对数据的批判性认识,能够对数据进行进一步的处理与分析,并运用到临床学习中。医学信息学专业学生培养则更注重其专业性,在了解知识图谱构建技术的基础上,重在培养学生对数据的挖掘能力,并对数据分析解读及工具方法有进一步要求,体现健康数据素养的综合能力。

5　结语

随着计算机计算能力的提升、大数据的爆发和各种新型算法的提出,人工智能再次焕发出新的生机,知识图谱的地位也再一次得到重视。知识图谱因蕴含着人类大量先验知识而被认为是进一步提升人工智能效果的重要思路之一。在医疗卫生领域,知识图谱可以从海量数据中提炼出医疗知识,辅助医师和研究人员通过合理高效地对其进行管理、共享及应用,从而进行科学的决策。因此在医学院校构建数据素养教育体系有助于促进医学发展。本研究从不同学科背景、层次阶段,以知识图谱为核心内容,从健康数据的基础知识、知识图谱的基本技术、知识图谱的分析方法及工具、知识图谱的应用与健康数据的伦理问题五方面建立健康数据素养教育体系,旨在提升医学生的数据处理分析、科研转化能力和岗位胜任能力。通过推动从技术应用向能力素养拓展,使学生具备良好的数据思维,适应社会发展的要求,更好地解决学习生活中的问题。研究认为,在医学院校构建数据素养教育体系并开展实践是顺应时代发展的;提升健康数据素养正是当代医学院校培养学生能力的关键点和着力点,也是当今推动医学发展的必然要求。

利益冲突　所有作者均声明不存在利益冲突

作者贡献声明　张靓:文章构思,负责文章撰写、修改;周敏:负责资料的收集整理;贺培凤:指导、审核论文;吕艳华:项目负责人,提供理论指导,给出修改意见

参考文献

[1] 中华人民共和国教育部. 高等学校人工智能创新行动计划[EB/OL]. (2018-04-02)[2018-04-25]. http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html.
Ministry of Education of the People's Republic of China. University artificial intelligence innovation action plan [EB/OL]. (2018-04-02) [2018-04-25]. http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html.

[2] 张静波. 大数据时代的数据素养教育[J]. 科学, 2013, 65(4): 29-32. DOI: 10.3969/j.issn.0368-6396.2013.04.008.
Zhang JB. Data literacy in the age of big data education [J]. Science, 2013, 65(4): 29-32. DOI: 10.3969/j.issn.0368-6396.2013.04.008.

[3] 郝媛玲, 沈婷婷. 数据素养及其培养机制的构建与策略思考[J]. 情报理论与实践, 2016, 39(1): 58-63. DOI: 10.16353/j.cnki.1000-7490.2016.01.010.
Hao YL, Shen TT. Construction and strategic thinking of data literacy and cultivation mechanism [J]. Information Studies: Theory & Application, 2016, 39(1): 58-63. DOI: 10.16353/j.cnki.1000-7490.2016.01.010.

[4] 欧阳玲琳. 大数据背景下的医科院校数据素养教育[J]. 中华医学图书情报杂志, 2017, 26(6): 59-62. DOI: 10.3969/j.issn.1671-3982.2017.06.014.
Ouyang LL. Education of data literacy in medical colleges and universities in background of big data [J]. Chinese Journal of Medical Library and Information Science, 2017, 26(6): 59-62. DOI: 10.3969/j.issn.1671-3982.2017.06.014.

[5] 哈佛医学院.哈佛医学院课程[EB/OL]. (2015-04-08) [2018-04-25]. https://courses.harvard.edu.
Harvard Medical School. Harvard medical school curriculum [EB/OL]. (2015-04-08) [2018-04-25]. https://courses.harvard.edu.

[6] 牛津大学医学部.牛津大学医学部课程[EB/OL]. (2018-04-06) [2018-04-25].https://www.medsci.ox.ac.uk/study/skillstraining
Department of Medicine, University of Oxford. Oxford University department of medicine curriculum [EB/OL]. (2018-04-06) [2018-04-25]. https://www.medsci.ox.ac.uk/study/skillstraining.

[7] 斯坦福医学院. 斯坦福医学院课程[EB/OL]. (2017-09-06) [2018-04-25]. http://exploredegrees.stanford.edu/schoolofmedicine/#courseinventory.
Stanford Medical School. Stanford medical school curriculum [EB/OL].(2017-09-06). [2018-04-25]. http://exploredegrees.stanford.edu/schoolofmedicine/#courseinventory.

[8] 约翰霍普金斯大学医学院. 约翰霍普金斯大学医学院课程[EB/OL]. (2014-05-03). [2018-04-25]. https://sis.jhu.edu/classes/results.aspx.
Johns Hopkins University School of Medicine. Johns Hopkins university school of medicine curriculum [EB/OL]. (2014-05-03)

[2018-04-25]. https://sis.jhu.edu/classes/results.aspx.

[9] 北京大学医学部.北京大学医学部课程[EB/OL]. (2014-11-05) [2018-04-25]. http://cc.bjmu.edu.cn/G2S/ShowSystem/CourseAll.aspx.
Department of Medicine, Peking University. Department of medicine curriculum [EB/OL]. (2014-11-05) [2018-04-25]. http://cc.bjmu.edu.cn/G2S/ShowSystem/CourseAll.aspx.

[10] 复旦大学上海医学院.复旦大学上海医学院课程[EB/OL]. (2013-09-01) [2018-04-25]. http://sph.fudan.edu.cn/.
Shanghai Medical College of Fudan University. Shanghai Medical College of Fudan University curriculum[EB/OL]. (2013-09-01) [2018-04-25]. http://sph.fudan.edu.cn/.

[11] 中国医科大学.中国医科大学课程[EB/OL].(2015-08-08)[2018-04-25]. http://202.118.40.32/bkzs/zyjs.htm.
China Medical University. China medical university curriculum [EB/OL]. (2015-08-08) [2018-04-25]. http://202.118.40.32/bkzs/zyjs.htm.

[12] 中南大学湘雅医学院.中南大学湘雅医学院课程[EB/OL]. (2014-02-06) [2018-04-25]. http://jcyxy.csu.edu.cn/xygk1111/xyjj11.htm.
XiangYa School of Medicine, CSU. XiangYa School of Medicine curriculum[EB/OL]. (2014-02-06) [2018-04-25]. http://jcyxy.csu.edu.cn/xygk1111/xyjj11.htm.

[13] 库恩. 科学革命的结构[M]. 北京：北京大学出版社，2012.
Kuhn. Structure of scientific revolution [M]. Beijing: Peking University Press, 2012.

[14] 俞国培，包小源，黄新霆，等. 医疗健康大数据的种类、性质及有关问题[J]. 医学信息学杂志，2014，35(6)：9-12. DOI: 10.3969/j.issn.1673-6036.2014.06.002.
Yu GP, Bao XY, Huang XT, et al. Medical and health big data: types, characteristics and relevant issues [J]. Journal of Medical Informatics, 2014, 35(6): 9-12. DOI: 10.3969/j.isnn.1673-6036.2014.06.002.

[15] 袁凯琦，邓扬，陈道源，等. 医学知识图谱构建技术与研究进展[J]. 计算机应用研究，2018，35(7)：1929-1936.
Yuan KQ, Deng Y, Chen DY, et al. Construction techniques and research development of medical knowledge graph [J]. Application Research of Computers, 2018, 35(7):1929-1936.

[16] 杨思洛，韩瑞珍. 国外知识图谱绘制的方法与工具分析[J]. 图书情报知识，2012(6)：101-109. DOI: 10.13366/j.dik.2012.06.006.
Yang SL, Han RZ. Analysis of foreigh methods and tools of mapping knowledge domain [J]. Documentation, Information & Knowledge, 2012(6): 101-109. DOI: 10.13366/j.dik.2012.06.006.

[17] 张观林，欧阳纯萍，邹银凤，等. 知识图谱及其在医疗领域的应用[J]. 湖南科技学院学报，2016，37(10)：73-75. DOI: 10.3969/j.issn.1673-2219.2016.10.025.
Zhang GL, Ouyang CP, Zou YF, et al. Knowledge map and application in the medical field [J]. Journal of Hunan University of Science and Engineering, 2016, 37(10): 73-75. DOI: 10.3969/j.issn.1673-2219.2016.10.025.

[18] 刘星，王晓敏. 医疗大数据建设中的伦理问题[J]. 伦理学研究，2015(6)：119-122. DOI: 10.15995/j.cnki.llxyj.2015.06.024.
Liu X, Wang XM. Ethical issues in the construction of medical big data [J]. Studies in Ethics, 2015(6): 119-122. DOI: 10.15995/j.cnki.llxyj.2015.06.024.

[19] 陈祥鹏，谢正德. 大数据时代下生命医学伦理教育教学面临的挑战及对策[J]. 继续医学教育，2016(4)：94-95. DOI: 10.3969/j.issn.1004-6763.2016.04.055.
Chen XP, Xie ZD. Challenges and countermeasures of education teaching of biomedical ethics in the era of big data [J]. Continuing Medical Education, 2016(4): 94-95. DOI: 10.3969/j.issn.1004-6763.2016.04.055.

[20] 黄如花，林焱. 大数据背景下数据素养教育研究[J]. 数字图书馆论坛，2016(5)：19-26. DOI: 10.3772/j.issn.1673-2286.2016.5.004.
Huang RH, Lin Y. Data literacy education under the background of big data [J]. Digital Library Forum, 2016(5): 19-26. DOI: 10.3772/j.issn.1673-2286.2016.5.004.

（收稿日期：2019-06-12）
（本文编辑：唐宗顺）

欢迎订阅，敬请赐稿！