

基于学位论文主题挖掘的医学信息专业跨学科研究特征分析*

王桐 王伟 车宏鑫 吴可凡 陈祥

(吉林大学公共卫生学院医学信息学系 长春 130021)

〔摘要〕 以学位论文为数据基础,采用共现聚类及文本挖掘方法,从知识交叉和研究内容方面分析医学信息研究生教育的跨学科特征,指出对新技术及应用的研究较为薄弱、各学科研究内容侧重明显不同等问题,建议增加融合多学科的科研思维培养及训练。

〔关键词〕 医学信息学;研究生教育;学科特征;主题识别;文本聚类

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.05.017

Analysis of Interdisciplinary Research Characteristics of Medical Information Specialty Based on Topics Mining of Postgraduate

Dissertation WANG Tong, WANG Wei, CHE Hongxin, WU Kefan, CHEN Xiang, Department of Medical Informatics, School of Public Health, Jilin University, Changchun 130021, China

〔Abstract〕 Taking the postgraduate dissertation as the research data, through co-occurrence clustering and text mining methods, the interdisciplinary characteristics of medical informatics postgraduate education are analyzed from the aspects of knowledge intersection and research content. It is pointed out that the research on new technologies and applications is relatively weak, the emphasis of research content in each discipline is obviously different. It is recommended to increase the cultivation and training of multidisciplinary scientific research thinking.

〔Keywords〕 medical informatics; postgraduate education; disciplinary characteristics; topic identification; text clustering

1 引言

医学信息学于 20 世纪 70 年代初由国际信息处理协会首次提出^[1],是以信息学、信息管理和信息

技术为依托,研究医学领域中的信息现象和信息规律,用于医学决策和管理的一门交叉学科^[2]。随着全球发达国家对医疗健康信息化日益重视,我国也积极建设医疗卫生信息化服务体系^[3]。社会对医学信息相关专业人才的需求随之增加,然而医学信息专业本科生就业岗位与专业匹配度不高^[4],这对医学信息教育提出更高要求。将中美两国医学信息教育进行对比发现^[5],我国医学信息教育以本科教育为主,开展硕博研究生培养的高校数量较少。研究生培养较本科生培养在科研问题识别、理论和方法创新方面要求更高^[6],但目前缺少对医学信息相关专业研究生培养特征的分析。同时,医学信息学是交叉学科,强调学科知识整合、共享和交融^[7],具

〔修回日期〕 2022-11-15

〔作者简介〕 王桐,博士研究生,发表论文 9 篇;通信作者:王伟,博士,教授,博士生导师,发表论文 70 篇。

〔基金项目〕 吉林大学研究生教育教学改革建设项目“医学信息学研究生跨学科培养实践探索”(项目编号:2021JGY45)。

体指利用多个学科的理论、方法和技术解决医学信息领域科学问题。分析我国医学信息相关专业的跨学科特征对完善培养方案具有重要指导作用。因此本研究以研究生学位论文为数据基础,探究医学信息学科交叉现状,分析医学信息相关学科研究生培养内容特征,总结当前我国医学信息高层次人才培养特点,以期为完善医学信息高层次人才培养体系提供参考借鉴。

2 资料与方法

2.1 数据来源

参考中国医学发展系列报告中医学信息学进展^[8]相关内容选取研究对象,确定医学信息研究生教育二级学科专业名称;通过中国研究生招生信息网筛选具体专业名称,包括医学信息学、医学/医药/智能/卫生信息管理、医学信息工程/系统/技术、中医药/药学信息学、医学生物信息学/生物医学信息处理/生物医学信息与仪器、计算医学/健康数据科学。由于本研究目的是探究医学信息专业研究生的跨学科研究特征,而管理学门类下相关专业多为本科专业,如信息管理与信息系统(医学信息方向),因此未纳入分析,最终确定具有相关专业的39所高校。通过中国知网硕博学位论文数据库进行查询,共有18所高校相关专业可检索到公开学位论文,涉及13个学科,已毕业博士70名和硕士188名,见表1。提取其论文所标注的中国图书馆分类代码^[9]及论文摘要。检索时间为2021年12月1—5日。

表1 研究对象基本情况统计分析

学科名称	高校数量(所)	博士(人)	硕士(人)
中医药信息学	6	0	21
医学信息学	4	19	44
医学生物信息学	2	0	21
生物信息技术	1	36	31
生物医学信息技术	1	9	7
卫生信息管理	1	2	26
医学信息工程	1	2	4
生物医学信息处理	1	1	0

续表1

学科名称	高校数量(所)	博士(人)	硕士(人)
计算医学	1	1	2
医学信息管理	1	0	16
医疗信息技术	1	0	10
智能医学信息管理	1	0	6

注:调研高校包括安徽中医药大学、北京协和医学院、成都中医药大学、重庆医科大学、复旦大学、贵州大学、广西医科大学、广州中医药大学、华北理工大学、华中科技大学、吉林大学、江西中医药大学、昆明理工大学、南京中医药大学、宁波大学、宁夏医科大学、山东大学、浙江中医药大学。

2.2 数据分析

2.2.1 学科分布基本统计分析 对学位论文中标注的二级学科分类代码进行描述性统计分析,了解医学信息相关专业研究领域的基本分布。按专业名称分别统计代码分布情况,对比不同专业在研究内容上的学科侧重。

2.2.2 基于VOSviewer的论文分类代码共现分析 以中国图书馆分类代码^[9]作为学位论文学科归属的测量指标。同一篇论文被赋予多个分类代码体现医学信息领域同一个研究课题的多学科交叉。分类代码共现可以揭示医学信息专业不同研究方向中各细支学科知识互换和交叉的模式。因此借助可视化分析工具VOSviewer对学位论文标注的分类代码进行共现与聚类。

2.2.3 基于隐含狄利克雷分布模型的学位论文研究主题分析 以学位论文摘要为数据源,构建语料库,基于隐含狄利克雷分布模型主题建模对摘要内容进行语义分析,挖掘医学信息研究的主题分类。对所有毕业论文进行主题识别,分析医学信息宏观研究内容的主题分布;根据专业名称划分3个类别,分别进行主题识别,对比分析不同专业研究内容上的异同。中文摘要的自然语言处理通过Python的jieba库实现。

3 结果

3.1 学科分类

根据学位论文所属二级学科的分布(仅显示频

次大于等于 3 的学科分类), 当前医学信息专业使用最多的方法和技术源于计算机科学领域, 围绕基础医学、肿瘤医学和预防医学相关问题展开研究。根据名称将专业进一步划分为 3 类: 生物医学信息学、医学信息学和中医药信息学。对 3 类专业分别进行分析, 医学信息学和中医药信息学仍以计算机科学为主要方法, 但医学信息学以预防医学和肿瘤学为主要研究问题, 中医药信息学主要解决中医学和临床医学中的科学问题。生物医学信息学则以生物化学、分子生物学、肿瘤学等为主要研究核心, 侧重于基础研究, 见表 2。

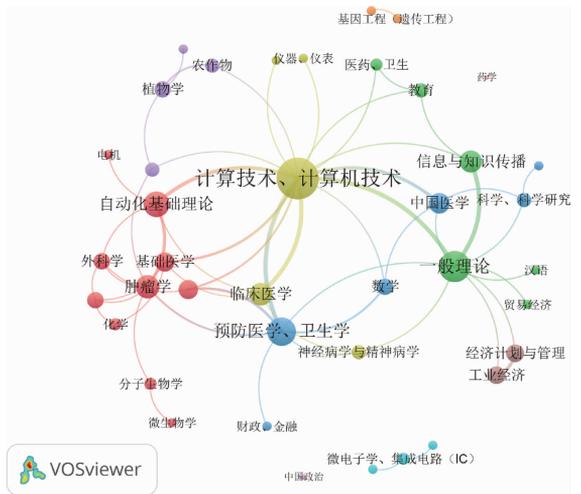


表 2 各类别专业研究生选题所属的二级学科 (部分)

学科类别	二级分类号	分类号含义	频次	
生物医学信息学/ 技术/处理类	Q5	生物化学	15	
	Q7	分子生物学	12	
	Q94	植物学	12	
	R73	肿瘤学	12	
	S5	农作物	11	
	TP	自动化技术、计算机技术	10	
	R3	基础医学	9	
	Q81	生物工程学 (生物技术)	8	
	R74	神经病学与精神病学	5	
	R9	药学	4	
医学信息学/管理/ 工程/技术类	TN	无线电电子学、电信技术	4	
	TP	自动化技术、计算机技术	48	
	R-0	一般理论	31	
	R1	预防医学、卫生学	27	
	R73	肿瘤学	17	
	R3	基础医学	9	
	R4	临床医学	9	
	G2	信息与知识传播	8	
	R5	内科学	7	
	R9	药学	7	
	G3	科学、科学研究	5	
	R74	神经病学与精神病学	5	
	F2	经济计划与管理	4	
	F4	工业经济	4	
	中医药信息学类	TP	自动化技术、计算机技术	14
		R2	中国医学	8
R-0		一般理论	3	
R4		临床医学	3	

图 1 医学信息专业的学科分类共现

表 3 医学信息专业研究生选题共现分类及所属学科

类别	学科	类别	学科	
分类 1	分子生物学	分类 4	财政、金融	
	内科学		预防医学、卫生学	
	化学		临床医学	
	基础医学		仪器、仪表	
	分类 2	外科学	分类 5	生物物理学
		微生物学		神经病学与精神病学
		肿瘤学		计算技术、计算机技术
		自动化基础理论	分类 6	农作物
		自动化技术及设备		园艺
		一般理论		植物学
信息与知识传播		生物工程学 (生物技术)		
医药、卫生		微电子学、集成电路		
教育		生理学		
汉语		通信		
分类 3	贸易经济	分类 7	基因工程 (遗传工程)	
	中国医学		生物化学	
	口腔医学	分类 8	工业经济	
	数学		经济计划与管理	
	科学、科学研究		分类 9	中国政治
		分类 10	药学	

其中分类 4 是所有医学信息相关专业研究内容的核心, 包含计算技术、计算机技术、临床医学、生物物理学、神经病学与精神病学、仪器、仪表, 以解决医学相关问题为目的, 以计算机技术为手段, 应用于医疗健康领域。

3.3 研究内容主题分布

医学信息专业研究生选题内容主题建模可视化结果, 见图 2; 医学信息相关专业学位论文研究主题, 可归纳为 7 个, 见表 4。3 个专业类别下研究内

3.2 学科交叉现状

根据学科代码共现关系, 所有学科分类号可聚为 10 类, 见图 1、表 3。

容的主题分布，见表 5。

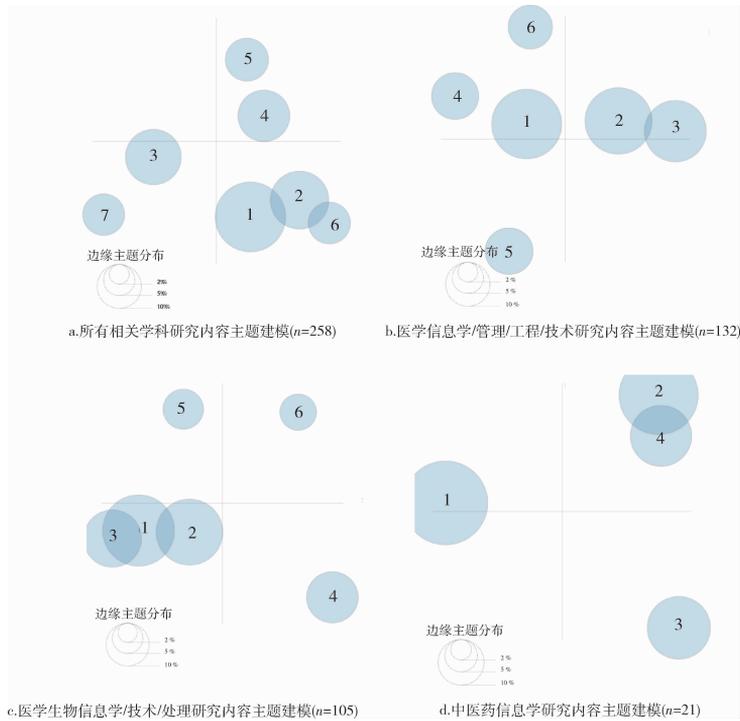


图 2 研究内容主题建模可视化

表 4 医学信息相关学科研究生毕业论文摘要主题及关键词

主题	关键词
生物基因转录及表达类研究（生物医学信息学方向）	基因，认知，蛋白质，转移，健康，位点，预测，生物，转录，表达
数据驱动医院业务管理及患者管理类研究	患者，医疗，模型，医院，服务，数据，因素，疾病，管理，业务，决策
临床图像处理及诊断类研究	图像增强，成像，发育，校正，扫描，算法，超声，噪声，诊断，临床
医药健康产业专利威胁预警类研究	技术，专利，模型，分析，识别，前沿，企业，威胁，预测，构建，产品
公共卫生事件网络舆情管控及健康信息学研究	信息，舆情，运行，网络，特征，情感，时间，主题，公共卫生，突发，引导
数据驱动的药物不良反应类研究	药品，数据，知识，信息，不良反应，框架，临床，干预，指南，个体
基于影像和临床数据的肿瘤检测和治疗类研究	细胞，图像，算法，分割，检测，表达，作用，肿瘤，治疗，实验，阈值

表 5 不同专业类别研究生毕业论文摘要主题及关键词

学科类别	主题	关键词
医学信息学/管理/工程/技术类	医药健康产业专利威胁预警类研究	专利，技术，企业，前沿，识别，竞争对手，威胁，研发，预警，医学
	数据驱动医院业务管理及患者管理类研究	医疗，医院，服务，管理，数据，知识，临床，诊断，过程，因素，模式
	用户的健康信息行为类研究	信息，健康，认知，用户，情感，特征，风险，搜寻，网络，舆情，事件
	基于图像和临床数据的疾病诊断预测	算法，框架，图像，干预，报告，模型，诊断，检测，噪声，文本，降噪
生物医学信息学/技术/处理类	疾病的应用类型	细胞，视神经，医疗，产检，脑血管病，老年人，妊娠，胎儿发育，医生
	数据驱动的药物不良反应类研究	药品，不良反应，患者，数据，基因，信号，临床，筛查，差异，分析
	微生物基因的转录相关研究	基因，表达，数据，微生物，分析，功能，转录，差异，调控，衰老，癌症
	基因的结构测序等研究	基因，基因组，结构，预测，序列，同源，分析，进化，染色体，特征
	蛋白质和基因组的修饰等研究	蛋白质，修饰，基因，自噬，位点，数据，调控，磷酸化，代谢，翻译
	研究目的（为患者疾病进行预测和检验）	模型，转换器，患者，疾病，服务，信息，医疗，建立，预测，慢性病，分析

续表 5

学科类别	主题	关键词
中医药信息学类	研究方法	细胞, 骨钙素, 表达, 作用, 实验, 治疗, 机制, 检测, 快速, 荧光
	应用场景 (探究由遗传等因素产生的疾病)	芯片, 儿童, 家庭, 川崎, 遗传, 环境, 临床, 诊断, 刺激
	中医病历及文本数据的术语研究	标注, 中医, 数据, 实体, 命名, 训练, 病历, 规范, 语料库
	中医药医疗及健康档案的数字化	中药, 免疫, 功能, 信息, 档案, 数据, 数字化, 标准, 健康
	基于临床影像及文本数据的诊断	中医, 图像, 中医药, 挖掘, 临床, 数据, 文本, 症状, 诊断
中医药处方和方剂的数据挖掘	处方, 用药, 治疗, 数据, 方剂, 中药, 分析, 数据挖掘, 规则	

4 讨论

4.1 我国医学信息专业学科建设基本特征

当前我国高校医学信息相关专业建设并不统一, 名称上仅中医药信息学达到一致。生物医学信息学相关专业名称差异较大, 存在医学生物信息学、生物医学信息技术及生物医学信息处理多个表述, 医学信息学相关专业也存在同样现象。不同表述揭示了学科侧重方向, 但不利于整个学科体系的规范管理, 仍需进一步统一。

不同专业类别在选题上具有明显差异, 生物医学信息学以生物化学、分子生物学和肿瘤学为主要组成, 以计算机技术与生物技术为研究手段, 解决细胞及更细化层面的生物医学问题。中医药信息学利用计算机技术处理中医领域的科学问题。医学信息学关注信息处理方法和理论的创新及改进, 以提出更高效率和精度的数据和信息处理技术、方法及模型理论为导向。虽然医学信息的 3 个专业类别均涵盖多个学科, 但在信息与知识传播等方面探究较少。

4.2 医学信息专业研究生选题学科交叉特征

当前医学信息专业的研究以医药卫生、工业技术及生物医学为核心。以解决医学相关问题为目的, 借助信息科学及计算机科学的理论和方法解决科学问题是医学信息专业跨学科研究的主要模式, 因此培养单位应以医学及计算机科学为培养核心, 开设理论及方法实践课程。可根据研究生本科教育经历, 提供薄弱环节的选修课程。为非医学生提供医学技能培训, 如开设疾病与健康检测方法、诊断和治疗干预、疾病预防等课程。为医学相关专业学生提供信息科学及计算机科学相关课程, 例如数据库开发、编程技术、数据挖掘、信息检索、信息组

织等, 提高数据收集及处理组织能力。随着数据时代的快速发展, 要不断更新课程内容。计算机科学、数学与计算生物学、保健科学、统计与概率等是全球医学信息主要学科组成^[10]。相比国际医学信息学科发展, 我国在数学等学科方面较为薄弱, 缺少数学及建模方面相关研究。

4.3 医学信息专业研究生研究内容主题分布特点

4.3.1 前沿主题研究尚待加强 医学信息学前沿主题可分为 5 个: 基于网络的认知行为疗法、可穿戴智能运动检测设备、移动医疗及适用于健康医疗的智能移动端应用、基于网络的干预治疗研究和移动医疗及数字医疗^[11-13]。我国研究生所开展的研究内容较少涉及上述选题, 在探知前沿方面仍需锻炼。国际医学信息学近 10 年研究热点可分为 5 类: 生物医学信息学、卫生信息系统的开发与管理、卫生信息分析方法与实际应用、人工智能在临床诊疗中的应用、医学信息学新技术的应用与发展^[14-15]。我国研究生对前 3 个热点均有探究, 但对新技术, 如移动健康、远程医疗及数字疗法等方面探究较少, 应增加相关课程或讲座, 培养研究生对前沿信息的快速掌握能力。

4.3.2 生物医学信息学与中医药信息学研究特征具有明显差异 生物医学信息学以基因、蛋白质和细胞层面的科学问题为核心, 利用计算机和信息技术及方法解决问题。因此该学科应保证研究生具备基础医学相关知识, 关注组织及以下水平的研究前沿问题, 进行计算机及信息分析相关技术培训。中医药信息学是我国特色学科, 目前处于起步发展阶段, 应多关注中医结构化及非结构化数据特征的识别, 融合医学信息学解决问题的科学方法处理中医相关问题。除基本知识相关课程外, 也应尝试开设以研究为基础的教学课程, 规范研究生科研流

程, 培养科研思维以及处理未知问题的科研能力。

5 结语

通过对我国医学信息相关专业研究生学位论文进行学科分布及研究内容分析, 探索医学信息专业的跨学科特征, 研究结果对医学信息高层次复合型人才培养方案的完善及更新具有指导价值。对 258 位研究生的学位论文进行分析发现: 首先, 医学信息专业的研究以医药卫生、工业技术及生物医学的跨学科交叉为主; 其次, 与当前国际医学信息学研究热点及研究前沿相比, 新技术及应用的相关研究较为薄弱, 应增加获取前沿研究信息的渠道及训练。最后, 各专业应根据研究内容侧重, 提供符合学科特色的核心课程, 并以科学研究为基础进行科研思维、逻辑和能力的培养及训练。当前我国医学信息相关专业在专业名称、培养目标、骨干课程和学科隶属关系等方面存在问题, 建议加强规范化管理和标准化建设, 各院校形成合力, 推动医学信息专业和学科的高质量发展和可持续发展。

参考文献

- 1 代涛. 医学信息学的发展与思考 [J]. 医学信息学杂志, 2011, 32 (6): 2-16.
- 2 王伟. 医学信息学 [M]. 北京: 高等教育出版社, 2006.
- 3 汤学军, 董方杰, 张黎黎, 等. 我国医疗健康信息标准体系建设实践与思考 [J]. 中国卫生信息管理杂志, 2016, 13 (1): 31-36.
- 4 何四维, 兰小筠, 李忠民. 医学信息学毕业生就业问题分

- 析及对策研究 [J]. 图书馆学研究, 2012 (12): 27-30.
- 5 张雪, 张志强, 陈秀娟. 中美医学信息学本硕教育现状对比分析与启示 [J]. 图书情报工作, 2019, 63 (12): 12-21.
- 6 王晓东, 王伟, 王呼生, 等. 医学信息学实验教学三维体系构建 [J]. 现代情报, 2016, 36 (6): 124-127.
- 7 张琳, 孙蓓蓓, 黄颖. 交叉科学研究: 内涵、测度与影响 [J]. 科研管理, 2020, 41 (7): 279-288.
- 8 代涛. 医学信息学教育与人才培养 [M]. 北京: 中华医学电子音像出版社, 2020.
- 9 《中国图书馆分类法》编委会. 《中国图书馆分类法》第五版使用手册 [M]. 北京: 国家图书馆出版社, 2012.
- 10 亢川博, 王伟, 张世玉, 等. 国内外医学信息学研究现状的可视化分析 [J]. 中华医学图书情报杂志, 2016, 25 (8): 24-30.
- 11 王菲菲, 刘明. Altmetrics 视角下的交叉学科研究前沿探测——以医学信息学领域为例 [J]. 情报学报, 2020, 39 (10): 1011-1020.
- 12 FORNAZIN M, PENTEADO B E, CASTRO L, et al. From medical informatics to digital health: abibliometric analysis of the research field [C]. Atlanta: Assoc Information Systems, 2021.
- 13 HUSSEIN R. Medical informatics in the digital personalized health and medicine era: a SWOT analysis and actionable strategies [J]. Studies in health technology and informatics, 2020 (6): 869-873.
- 14 张雪, 陈秀娟, 张志强. 近十年国际医学信息学发展趋势与热点研究——基于 10 种高影响力外文期刊的文献计量分析 [J]. 现代情报, 2018, 38 (12): 151-163.
- 15 GÜLKESEN K H, HAUX R. Research subjects and research trends in medical informatics [J]. Methods of information in medicine, 2019, 58 (S1): e1-e13.

(上接第 70 页)

- 5 刘国庆, 韩婧怡, 宋文静, 等. 虚拟现实在护理领域应用研究关键词的可视化分析 [J]. 护理学报, 2022, 29 (3): 29-34.
- 6 杨耿, 萧裕红, 黄惠怡, 等. VR 在体育教学与活动中的应用分析 [J]. 电子技术与软件工程, 2021 (23): 21-23.
- 7 高昊辰, 韩煦, 朱强, 等. VR 联合 AR 融入式教学系统在临床医学专业口腔通识医学教育的应用研究 [J]. 中华医学教育探索杂志, 2021, 20 (11): 1279-1282.
- 8 李建, 张杰. 基于 VR 的虚拟仿真实验教学项目设计与探索——以“景观视觉空间分析虚拟仿真实验教学项目”为例 [J]. 中国教育信息化, 2021 (22): 42-48.
- 9 朱高红, 尹竹萍, 陈礼林, 等. 虚拟现实 (VR) 模式

- 立体医学影像解剖教学实践的初步研究 [J]. 中国高等医学教育, 2021 (10): 59-60.
- 10 李嘉祺, 罗光丽, 陈星宇, 等. 信息技术对抑郁症心理调适的模式开发创新研究——基于 VR 沉浸技术的应用 [J]. 电子质量, 2022 (1): 62-66.
- 11 毛润坤, 张莹, 董雪婷, 等. 基于心理声学的 VR 音频舒适度的主观评测和客观量化 [J]. 电声技术, 2019, 43 (3): 74-77.
- 12 PHAN H L, KIM J P, KIM K, et al. Wrist rehabilitation system using augmented reality for hemiplegic stroke patient rehabilitation: a feasibility study [J]. Applied sciences, 2019, 9 (14): 2892.