

智慧医院建设背景下的 电子病历分析利用框架

徐良辰, 郭崇慧

大连理工大学系统工程研究所, 辽宁 大连 116024

摘要

作为核心的医疗大数据, 电子病历成为智慧医院建设的基础, 对电子病历的分析利用进行研究对于智慧医院建设具有重要意义。提出了电子病历的生成、分析及利用一体化研究框架, 阐述了电子病历的内涵, 分析了电子病历系统与其他医疗信息系统的关系, 梳理了电子病历的分析挖掘流程, 并从计算机辅助诊断、治疗推荐以及管理支持3个角度总结了电子病历分析的应用; 讨论了电子病历分析对智慧医院相关分级的影响, 指出电子病历分析挖掘是智慧医院较高等级的要求, 将助力高等级智慧医院建设; 最后, 从数据层面、模型层面以及应用层面对当前电子病历分析利用存在的问题进行分析, 并给出相应意见和建议, 为电子病历分析利用以及智慧医院的建设提供参考。

关键词

电子病历; 数据挖掘; 智慧医院; 互联网+医疗健康

中图分类号: R197.323

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2021044

Analysis and utilization framework of electronic medical records under the background of smart hospital construction

XU Liangchen, GUO Chonghui

Institute of System Engineering, Dalian University of Technology, Dalian 116024, China

Abstract

As the core medical big data, electronic medical record data has become the core foundation of smart hospital construction, and the research on the analysis and utilization of electronic medical records is of great significance. In order to promote the analysis and utilization of electronic medical records, an integrated research framework for the generation, analysis and utilization of electronic medical records was proposed, the connotation of electronic medical records was expounded, the relationship between electronic medical record systems and other medical information systems was analyzed, and the analysis and mining process of electronic medical records were sorted out. The application of electronic medical record analysis was also summarized from three perspectives: computer-aided diagnosis, treatment recommendation and

management support. The impact of electronic medical record analysis on the relevant classification of smart hospital was also discussed. Finally, the existing problems in the analysis and utilization of electronic medical records were analyzed from the data level, model level and application level, and corresponding opinions and suggestions were given to provide references for the analysis and utilization of electronic medical records and the construction of smart hospital.

Key words

electronic medical record, data mining, smart hospital, Internet plus healthcare

1 引言

在经济和社会发展的现实需求拉动以及政策和技术因素的驱动下,人们对医院信息化水平的要求越来越高,这促使“互联网+医疗健康”不断深入发展。传统医院信息系统难以满足当今人们对医疗服务水平的要求,国家政策、经济、社会和技术等多重因素推动着医院向智慧化方向转变。以智慧服务、智慧诊疗、智慧管理为核心的智慧医院进入快速建设时期^[1],其建设基本框架如图1所示。

医疗健康大数据泛指所有与医疗和生命健康相关的信息,其贯穿个人生命的全周期^[2]。国家已对医疗健康大数据管理进行了统筹设计,2014年国家卫生和计划生育委员会发布了“46312”工程,其中

“3”是指三大数据库,分别为电子健康档案数据库、电子病历(electronic medical record, EMR)数据库和全员人口个案数据库。医疗健康大数据治理也受到学者的广泛关注^[3-4]。电子病历数据库以患者为中心,将患者医疗信息及其相关处理过程综合集成,促进了工作流程的优化、医疗质量的提高以及服务水平的提升^[5]。电子病历数据成为核心的医疗大数据,也是智慧医院建设的核心基础。发达国家对电子病历有较深入的研究和应用,并成立了相关机构来组织医疗单位实施和普及电子病历^[6]。2010年卫生部印发《电子病历基本规范(试行)》,规范了医疗机构电子病历的管理及应用;2018年,国家卫生健康委员会制定了《电子病历系统功能应用水平分级评价方法及标准》,为电子病历系统的功能应用水平分级提供了具体评价指标和方法,这标志着我国电子病历建设迈入了一个新的阶段。电子病历的建设、分析及利用对于智慧医院的建设和发展具有基础支持作用。

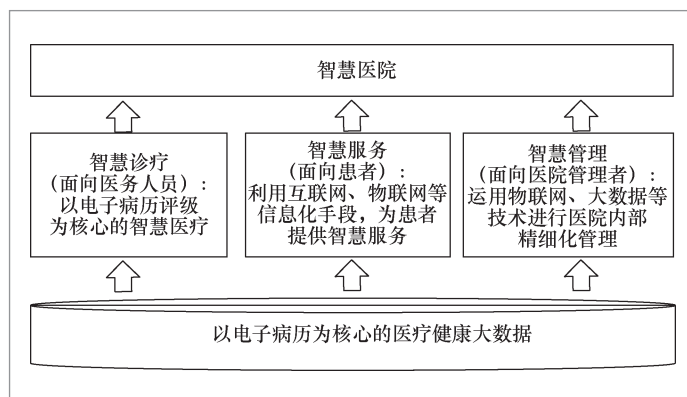


图1 智慧医院建设基本框架

2 电子病历的内涵及生成

电子病历是病历的一种信息化记录形式,在当前医疗信息系统建设中得到迅速发展。电子病历的定义可分为狭义和广义两种:狭义的电子病历是指纸质病历的电子化;广义的电子病历是指医务人员在医疗活动过程中形成的文字、符号、图表、影

像、切片等资料的总和,包括门(急)诊病历和住院病历。电子病历不仅包括静态病历信息,还包括医院提供的相关服务,涉及病人信息的采集、存储、传输、处理和利用等多个过程。美国医学研究所对电子病历做了如下定义:以电子化方式管理的有关个人终生健康状态和医疗保健的信息,可在医疗中作为主要的信息源取代纸张病历,满足所有的诊疗、法律和管理需求。综合来看,电子病历以患者数据为中心,是患者综合、完整、详尽的诊断治疗全过程的原始记录,并具有完整性、标准化、规范化、易检索等特点。

作为医院信息系统的一个子系统,电子病历系统以患者为中心,综合集成相关医疗信息系统数据及其相关处理过程,更强调患者信息的原始性和完整性,在医院信息管理系统中占有重要地位。电子病历系统与其他医院信息系统的关系如图2所示。

由图2可知,电子病历数据是以患者为中心的其他医疗信息系统数据的综合

集成,且表示形式多样化。电子病历数据既包含患者情况的文本描述等非结构化数据,也包含检查、检验报告等结构化数据。针对具体的患者,电子病历主要包括入院记录、病程记录(包括首次病程记录和其他病程记录等)、检查报告、检验报告、治疗记录(如手术记录等)以及出院小结等。电子病历记录了患者的重要诊疗信息,具有为临床决策、临床科研以及运用管理提供巨大支持的潜力,因此有必要对电子病历数据进行分析挖掘,并赋能实践。

3 电子病历分析及利用

随着新一代信息技术的发展,医疗数据的采集、抽取、存储以及处理能力增长迅速。面对海量医疗信息,数据挖掘等理论方法将助力电子病历的分析及利用,更好地为相关决策提供支持。基于数据挖掘的电子病历分析及利用框架如图3所示。

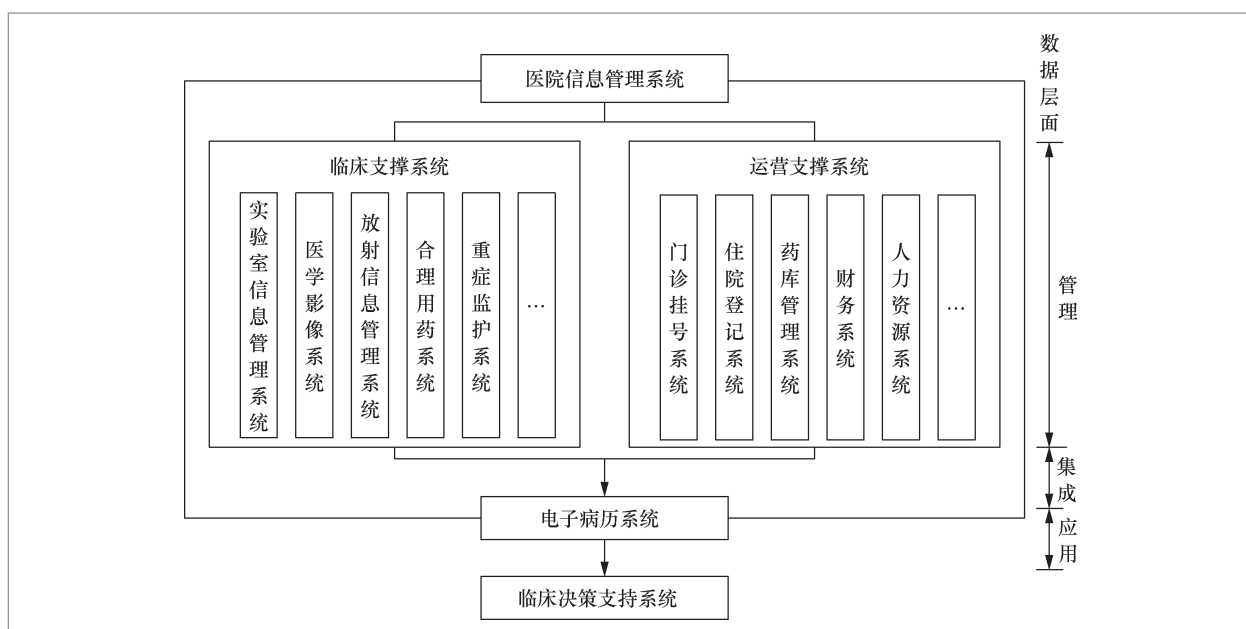


图2 电子病历系统与其他医院信息系统的关系

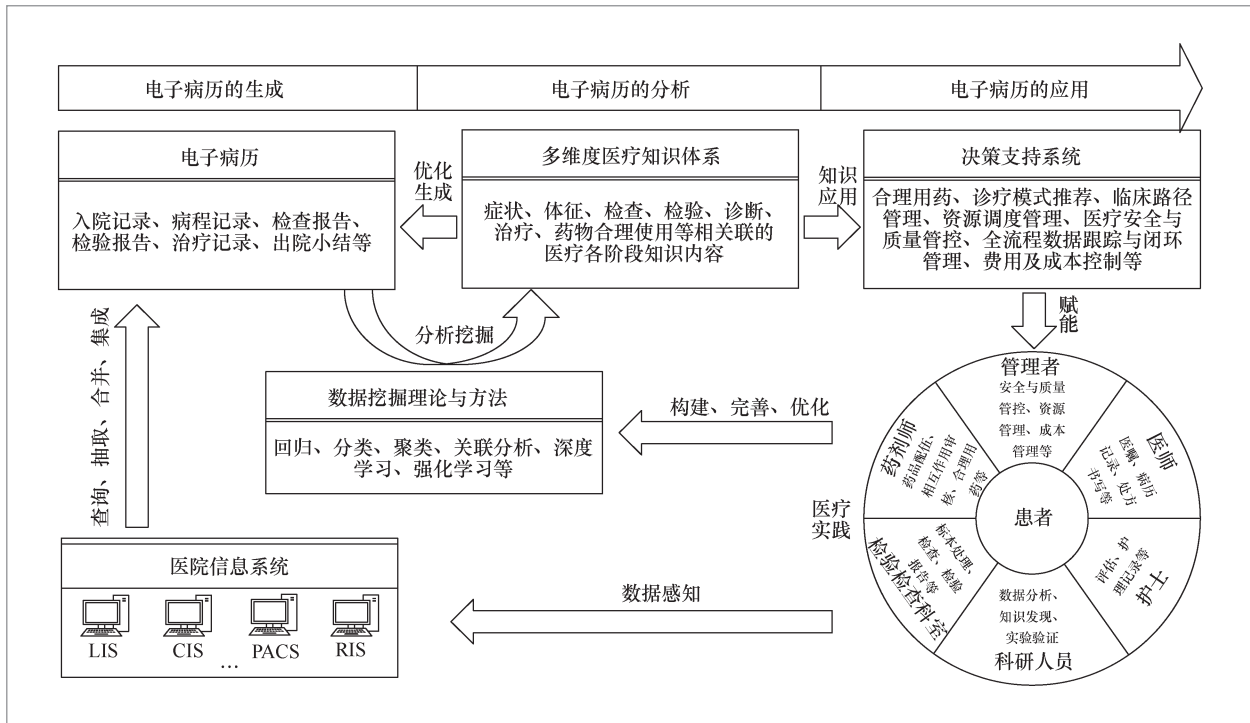


图3 基于数据挖掘的电子病历分析及利用框架

3.1 电子病历分析挖掘

在当前的临床实践中，许多治疗手段缺乏有效的证据支持。尽管临床医生利用参考指南来辅助决策，但参考指南通常建立在随机对照试验（randomized controlled trial, RCT）的基础上，而RCT的患者队列有限、标准严格，且经过了精心设计，普适性较低。传统RCT研究面临很多局限，如对于罕见病症的研究，临床上病例较少，而且可能包含其他难以控制的因素。此外，受到伦理道德的限制，一些干预手段难以实施。由于限制繁多，RCT只支持部分临床决策，大多数临床决策无据可循^[7]。

电子病历数据蕴含大量的信息和知识。相比于RCT以及前瞻性队列研究，充分利用电子病历可以极大地降低研究成本

和研究难度，为循证医学的发展提供新的思路。合理分析和挖掘电子病历可以极大地促进医护人员对临床实践的认知、优化诊疗业务流程，帮助医护人员制定符合患者特点的个性化临床决策，进而提高医疗服务质量和水平，为精准医疗的发展奠定基础。

海量的医疗大数据意味着机遇与挑战并存。电子病历集成了多部门、多类型、多项目的数据，具有多种数据特征。电子病历数据具有多样性，其包含多种数据类型，如结构化数据（如体温、血压、脉搏、检验结果等数据）、半结构化数据（如病史信息、病程记录、出院小结等）以及非结构化数据（如心电图、CT报告单等），而且数据量较大。随着医疗信息化平台的建设和深入应用，越来越多的医疗流程依托于医院信息系统，这促进了医疗平台整合及医疗设备互联，以便为患者提供便捷的医疗

服务。与此同时,患者的预约、诊断、治疗等活动产生的海量数据都会被记录,形成信息丰富的电子病历数据。电子病历数据是对患者诊疗过程和病情进展的记录,具有动态性;电子病历数据记录过程可能因为医疗人员的理解不同而存在误差或者缺失,产生很多异常值和缺失值,具有不完整性和不规范性;电子病历数据包括大量患者隐私信息,具有高隐私性。电子病历数据的特性为电子病历数据挖掘带来了困难和挑战。基于已有研究,本节梳理了基于数据挖掘的电子病历分析流程,并针对分析挖掘结果阐述了基于知识图谱的多维度医疗知识管理。

(1) 基于数据挖掘的电子病历分析挖掘流程

基于数据挖掘的电子病历分析挖掘指通过数据挖掘的理论方法对电子病历进行处理,将其中蕴含的知识进行挖掘和呈现。电子病历分析挖掘离不开实践需求,因此解决实践问题成为电子病历分析挖掘的重要目标。基于医疗实践活动中存在的相关问题进行分析提炼,形成科学研究问题,依据关注的科学问题设计实验。首先

基于对研究问题的理解,选择合适的研究队列,随后对定义好的研究队列的数据进行抽取,通过数据预处理来清洗完善数据,并通过探索性分析对数据有一个大致的了解。基于研究问题以及数据特点选择合适的模型方法进行实验,并进行评估和改进。达到预期后,可在大规模真实数据集上进行验证,最终发现知识,为医疗实践提供支持。基于数据挖掘的电子病历分析挖掘流程如图4所示。

(2) 基于知识图谱的多维度医疗知识管理

2012年Google公司提出的知识图谱改变了传统的搜索引擎模式,知识图谱可以智能化地理解用户的输入,并返回最相关的知识。因此,可采用知识图谱的形式组织医疗知识,构建多维度医疗知识库管理系统,实现多维度的医疗知识管理。基于知识图谱的多维度医疗知识管理主要包括医疗知识图谱构建、知识库管理系统设计和实现以及知识应用。

在医疗知识图谱的构建中,实体识别和实体关系抽取是基本任务,也是关键问题。实体识别的主要任务是从电子病历中

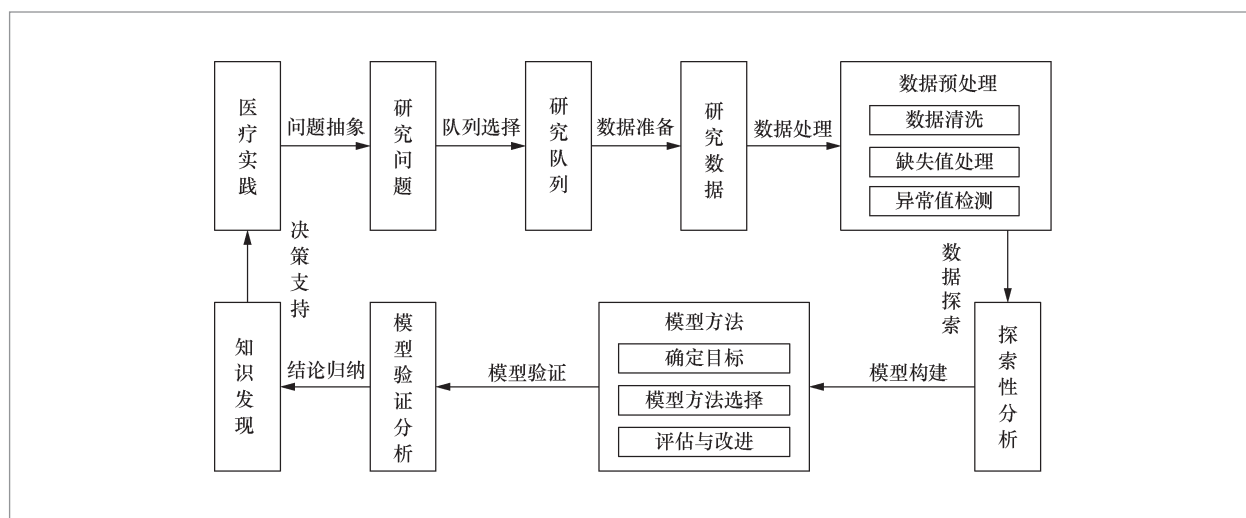


图4 基于数据挖掘的电子病历分析挖掘流程

找出相关实体(已存在的概念词语),包括疾病、症状、药物、检测以及治疗等;关系抽取的主要任务是识别抽取出的实体之间的关系,包括同质实体之间的关系(如药物与药物之间的关系)和异质实体之间的关系(如疾病与症状之间的关系)。在医疗知识图谱中,以实体为节点、实体之间的关系为边,从而对医疗知识进行表示。基于自然语言处理的电子病历的实体识别和知识图谱构建方法成为较流行的知识图谱构建方式^[8]。阮彤等人^[9]利用互联网数据与百科数据构造了医疗知识图谱,为中文医疗知识图谱的构建提供了思路。

3.2 电子病历分析的应用

3.2.1 基于数据挖掘的电子病历分析促进电子病历的智能化生成

当前电子病历很大程度上靠医生输入计算机,这占用了医生日常工作的大量时间。据统计,在工作日医生要花费6~11个小时在电子病历的输入上,这缩短了医生为患者服务的有效时间,降低了医疗人员的服务能力。随着信息技术的不断发展,电子病历的智能化生成将为降低医生的电子病历输入压力提供思路。电子病历的智能化生成将提高电子病历的生成效率,减少医生的电子病历输入任务,有助于提高医疗人员的服务能力和水平。基于数据挖掘的电子病历智能化生成包括电子病历医患共创、个性化电子病历模板推荐以及结构化数据推理生成等重要部分。电子病历的智能化生成框架如图5所示。

电子病历的智能化生成可依托互联网交互的特性,引导患者参与自身电子病历的生成,将一些信息收集阶段移至患者端,实现医患共创,降低医疗人员的电子病历输入压力。作为电子病历的重要信息标

准,电子病历模板确定了电子病历的风格和基本结构,可提高医疗数据记录的规范性和完整性。不同病种对电子病历模板有不同的要求,基于数据挖掘与分析,依据病人、病种、病情等实际需求,自动推荐契合的个性化电子病历模板将进一步提高电子病历的生成效率。电子病历中除了结构化数据,还包括医生的诊断与总结等非结构化数据,非结构化数据的推理生成也是电子病历智能化生成的重要方式,已有相关研究对中文出院小结自动生成的方法进行了探讨^[10]。基于医生的思维逻辑,依据可参考的数据信息推理生成诊断、摘要等非结构化数据,医生可进一步对生成内容进行完善,人机结合,进一步提高电子病历的生成效率和质量。

3.2.2 基于多维度医疗知识体系的决策支持

当前的医疗知识主要来源于专家知识(包括显性专家知识、隐性专家知识等)。随着医疗健康数据的积累以及数据挖掘技术的发展,基于数据挖掘的模式发现正逐渐成为医疗知识的重要来源之一,与专家知识一起构成多维度的医疗知识体系,进而为医疗决策提供支持。本文基于“互联网+医疗健康”的特点,结合数据挖掘、循证医学、医患共同决策、知识管理以及决策科学等理论方法,构建了基于多维度医疗知识体系的决策框架,如图6所示。

基于多维度医疗知识体系的支持,借助医生经验和患者知识,医患共同决策,采取契合患者的临床诊疗模式、护理模式以及康复模式等,利用积累的诊疗数据对已有数据进行补充和更新,以进一步完善医疗知识体系,形成治疗管理循环。

(1) 基于多维度医疗知识的辅助诊断

在临床实践中,疾病诊断很大程度上依赖于医生的专业能力和经验知识,这可

能导致医患矛盾加剧等问题。基于机器学习的电子病历分析可以基于历史数据挖掘医疗知识,模拟医生诊断推理,得出较为可靠的诊断预测。

电子病历的分析挖掘可赋能疾病预测与病情评估。基于患者的人口统计学信息、症状、检查结果等相关数据,对患者可能患有的疾病进行预测,对患者病情轻重缓急进行评估,可以为医务人员的相关诊断提供支持和参考,对于后续的治疗措施也有重要意义,如基于人工智能对肺结节病理类型的良恶性进行预测^[11]、基于电子病历数据挖掘对疾病危重度进行动态预测^[12]、基于混合模型对ICU病人死亡风险进行预测^[13]。

电子病历医学影像数据挖掘通过实现医学影像自动分析来辅助诊断。电子病历数据中包含大量的影像数据,这些数据是疾病诊疗决策的重要依据。医学影像数据量占医院数据存储总量的90%以上,积累量巨大^[14]。医学影像数据的分析将花费医生大量的精力,且对专业知识要求比较高。基于数据挖掘的影像数据分析有助于提高诊断效率和水平,如基于裂隙灯影像对小儿白内障进行自动诊断和预测^[15]。

(2) 基于多维度医疗知识的治疗指导

从电子病历数据库中挖掘潜在的医疗知识,识别大多数患者在治疗过程中的最佳实践,可为临床治疗决策提供参考,主要包括临床用药指导和临床路径管理。

临床用药一方面需要考虑药品的成分和药效,另一方面要考虑药品的毒副作用和配伍问题。基于电子病历中的用药等数据,采用数据挖掘方法对用药效益和药物治疗方案进行分析,可为临床用药决策提供参考,如基于数据挖掘的糖尿病肾病证候及用药规律研究^[16]、基于电子病历挖掘对中药方剂配伍规律进行研究^[17]。

在典型的治疗模式挖掘中,临床路径

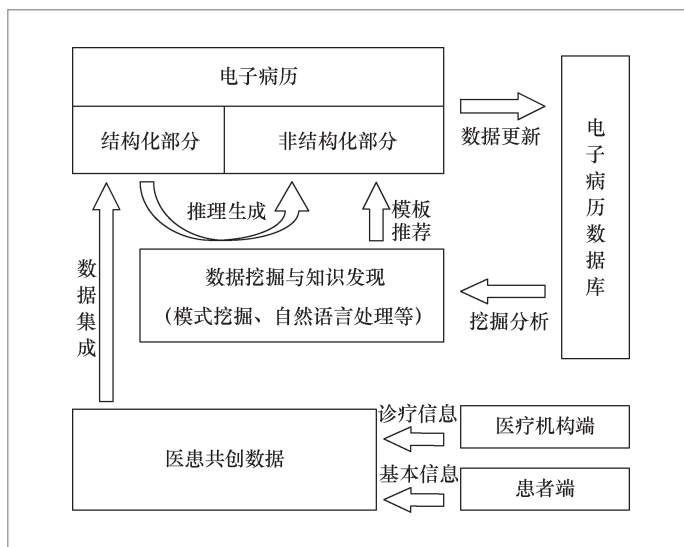


图5 电子病历智能化生成框架

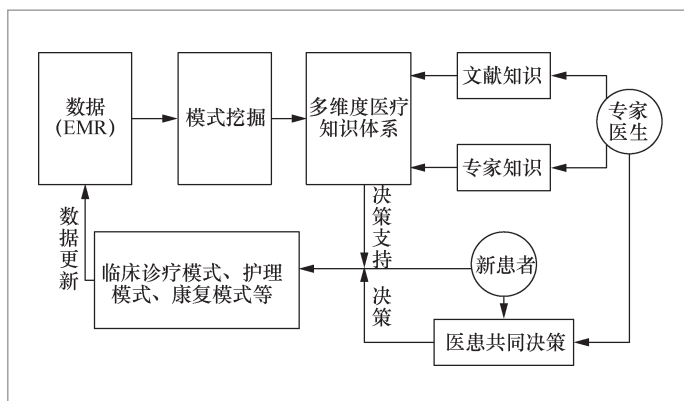


图6 基于多维度医疗知识体系的决策框架

挖掘是一个重要的方向。作为临床重要决策问题,临床路径管理体现了循证医学、质量管理以及流程管理等现代管理思想,对于保障医疗质量、优化服务流程以及控制医疗成本具有重要作用^[18]。传统临床路径主要依靠专家知识针对特定疾病构建的标准流程,耗时长、效率低,且需要很长时间进行完善,通用性不强,临床变异较大。基于医疗大数据特别是电子病历,采用数据挖掘技术对临床路径模式进行发现,形成临床路径知识体系,为患者的临床路径

的纳入、时序化管理以及变异管理等提供支持,将突破传统临床路径的制约瓶颈,促进临床路径的应用和推广。基于多维度医疗知识的临床路径管理框架如图7所示。

(3) 基于电子病历数据的运营管理

在医疗系统运作中,将运营管理与信息技术结合可有效地实现成本控制和精细化管理。

电子病历分析挖掘可为疾病费用管理提供支持。运用数据挖掘技术,对疾病诊疗过程中的药品、卫生材料使用等不同的诊疗行为规律进行分析识别,探索不同诊疗行为规律下影响患者费用的主要因素,基于患者信息对疾病费用进行预测,为疾病费用管理和控制提供了新的思路,如基于数据挖掘技术的冠心病费用管理^[19]、基于数据挖掘算法的慢性病住院费用预测^[20]、基于K-means聚类与支持向量机的大病患者住院费用影响因素与控制策略分析^[21]。

疾病诊断相关分组 (diagnosis related

group, DRG) 是医疗成本控制和精细化管理的重要举措,DRG将传统的实报实销的后付制转变为预付制,在控制不合理费用增长和过度医疗需求方面发挥了积极作用。在我国,对DRG的研究还处于探索阶段,DRG的应用程度还不够,分组的合理性对DRG的应用有重要影响。电子病历数据蕴含患者诊疗及费用等详细信息,数据挖掘在属性筛选、样本分类等方面具有突出优势,因此电子病历挖掘将为DRG的研究和应用提供思路,对于费用控制具有重要意义^[22]。

4 电子病历分析对智慧医院相关分级的影响

随着以电子病历为核心的医院信息化建设的不断推进,智慧医院创新发展,特别是在新型冠状病毒肺炎疫情期期间,智慧医院改善了医疗服务成果,在应对疫情、

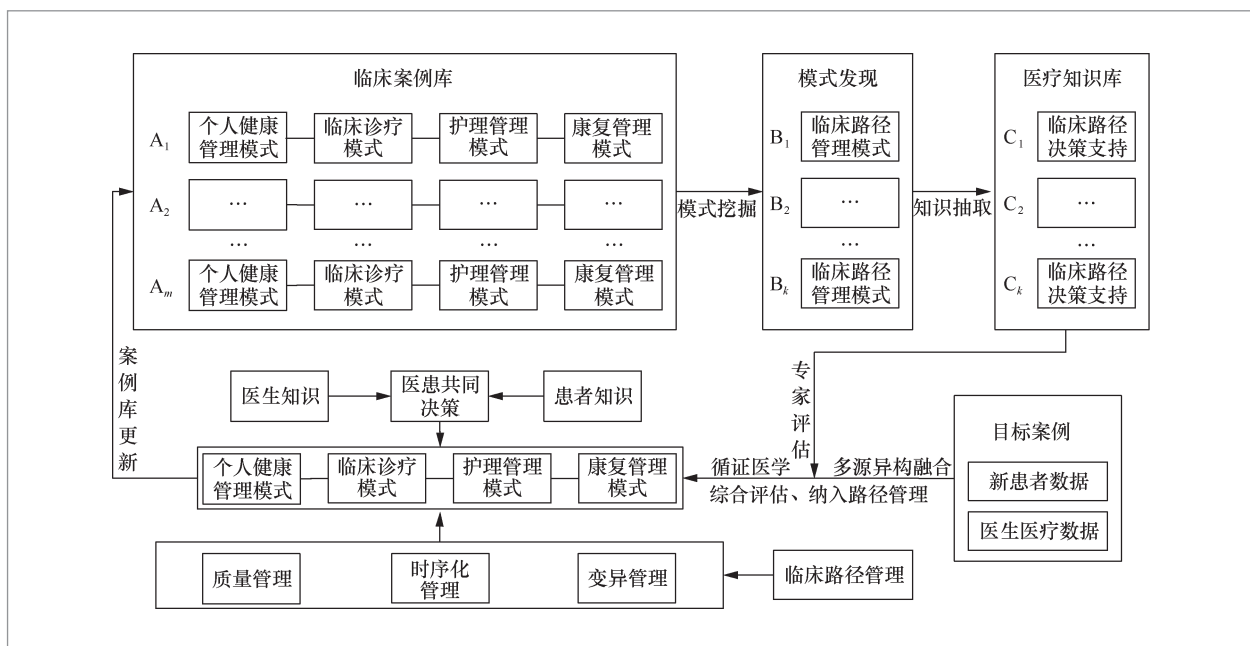


图7 基于多维度医疗知识的临床路径管理框架

满足人民群众就医需求等方面发挥了积极作用。为了进一步加强智慧医院建设,国家卫生健康委员会发布加强智慧医院建设的工作通知,创新建设完善医疗、服务、管理“三位一体”的智慧医院系统,提升医护人员工作效率,提高医院管理水平,为患者提供高质量、高效率的个性化医疗服务。

相关部门已针对服务和医疗分别制定《医院智慧服务分级评估标准体系(试行)》和《电子病历系统功能应用水平分级评价方法及标准(试行)》,二者对决策支持有一定的要求。电子病历是智慧医院建设的基础,对智慧医院的智慧服务以及电子病历系统功能应用水平分级有重要的影响。

(1) 电子病历分析对医院智慧服务分级的影响

智慧服务指应用信息技术改善患者的就医体验,提升医疗服务水平,是智慧医院建设的重要内容。电子病历分析有助于个性化医疗服务的展开。表1是《医院智慧服务分级评估标准体系(试行)》中对智慧服务的分级划分总体要求。

将医院智慧服务分级评估6个级别

表1 医院智慧服务评估分级

级别	内容
0	医院没有或极少应用信息化手段为患者提供服务
1	医院应用信息化手段为门(急)诊或住院患者提供部分服务
2	医院内部的智慧服务体系初步建立
3	联通医院内外的智慧服务体系初步建立
4	全医院智慧服务体系基本建立
5	基于医院的智慧医疗健康服务体系基本建立

的描述文本进行聚合。为了展示每个级别的特点,采用TF-IDF(term frequency-inverse document frequency)方法进行特征词权重的计算,并将6个级别的关键词以词云图形式进行展示,如图8所示。

由图8可以大致看出医院智慧服务每个级别的要求,也可以大致看出不同级别之间的差异。0级的主要特征词有“手工”“登记”等,1级的主要特征词有“预约”“信息系统”等,2级出现“自助”“共享”等特征词,3级和4级凸显“移动”“在线”等特征词,5级主要有“监测”“记录管理”等特征词。特征词的演化展现了不同级别要求的变化。

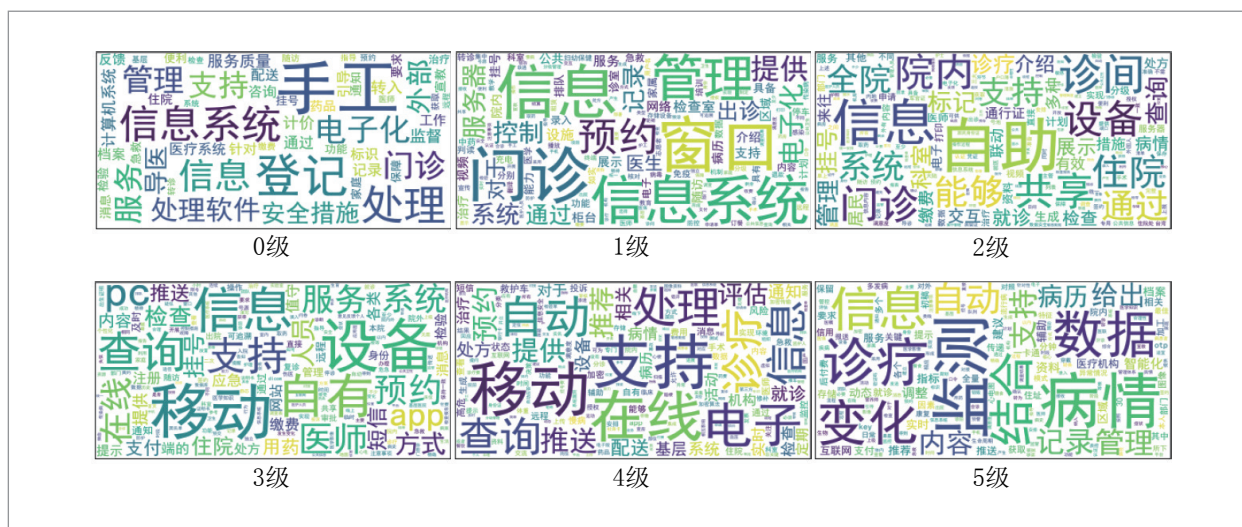


图8 医院智慧服务分级评估具体要求词云图

电子病历分析将赋能实践,助力医院智慧服务的提升。本文参照《医院智慧服务分级评估标准体系(试行)》中的医院智慧服务分级评估具体要求,分析、归纳电子病历分析可能影响的医院智慧服务及相关任务方法,见表2。从表2可以看出,电子病历分析有助于更高级别的医院智慧服务建设,依托分类、推荐、关联规则、文本挖掘以及自然语言处理等数据挖掘方法,可进一步提升医院的智慧服务水平。

(2) 电子病历分析对医院电子病历系统功能应用水平分级的影响

《电子病历系统功能应用水平分级评价方法及标准(试行)》将电子病历系统功能应用水平划分为0~8共9个等级,见表3。除了信息检索等基本功能,对数据

的分析挖掘和临床决策支持也提出了较高的要求。作为支持临床决策的一种重要方法,基于数据挖掘的电子病历分析对医院电子病历系统功能应用水平分级有重要影响。

对电子病历系统功能应用水平分级评价整体和局部要求的9个级别的描述文本进行抽取。为了展示每个级别的特点,本文同样采用TF-IDF方法进行特征词权重的计算,并将9个级别的关键词以词云图的形式进行展示,如图9所示。

由图9可以大致看出,0~3级主要是基础信息化建设,出现“计算机系统”“数据交换”“信息系统”等特征词;4级以上出现“审核”“集成”“知识库”等特征词,8级强调“区域”“质量”“整合”等特征词。

仔细分析表3中的电子病历系统功能应

表2 电子病历分析可能影响的医院智慧服务及相关任务方法

类别	业务项目	级别	具体要求	数据挖掘相关任务方法
诊前服务	转诊服务	5	可根据健康档案或监测得到的患者病情变化情况,给出诊疗或转诊建议	预测、推荐等
诊中服务	信息推送	5	(1)对于出院签约管理患者,可根据其健康情况自动调整消息通知内容; (2)根据患者病情和诊疗阶段,自动为患者、患者家属推送注意事项及宣教内容	分类、推荐等
	标识导航	5	可获取患者院内或医联体内多个科室的诊疗活动安排,并为患者规划最佳的诊疗路径	规划、推荐等
	便利保障	5	支持管理部门根据患者诊疗情况,结合营养师所下膳食医嘱自动向患者推荐适宜餐食	关联规则、推荐等
诊后服务	患者反馈	5	结合医院信息系统数据、患者满意度调查结果、舆情监测等信息,对医疗服务进行综合评估	舆情监测、文本挖掘等
	患者管理	4	支持患者提问的自动应答功能	自动问答、自然语言处理等
		5	根据患者病情变化,动态调整康复计划	模式挖掘、推荐等
	调剂配送	5	根据患者日常健康记录,动态检查患者用药合理性,并向患者及管理医师发送提示	药品配伍、异常检测等
	家庭服务	5	可依据患者病情、住址等内容,向患者推荐家庭医生团队	推荐等
	基层指导	5	可监控基层医疗机构的主要疾病情况,给出相应指导	疾病监测和支持
	智能导医	5	(1)根据患者病情及区域多发病、流行病情况等,给出患者分诊建议; (2)患者在诊前录入的症状、病史等信息可自动转为病历记录初稿	推荐、自然语言处理等
全程服务	健康宣教	4	可根据患者病历资料自动完成风险评估,并将结果推送给患者或者监护人	风险评估、预测等
		5	根据患者健康记录、监测信息、病情变化,有针对性地推送医学知识	预测、推荐等

用水平划分以及图9可知,3级及以下级别主要强调基础信息化的建设,强调信息系统的建设和基础的数据交换,传统信息管理系统可以满足3级及以下级别的基本需求。4级及以上级别强调数据共享和医疗决策支持,超出了传统信息管理系统信息存储与信息检索的功能范围。基于数据挖掘的电子病历分析可以为4级及以上级别的电子病历系统的建设提供理论以及技术方法支持。表4分析了电子病历评级主要决策支持要求及数据挖掘相关任务方法。

表3 电子病历系统功能应用水平级别划分

级别	整体要求
0	未形成电子病历系统
1	独立医疗信息系统建立
2	医疗信息部门内部交换
3	部门间数据交换
4	全院信息共享,初级医疗决策支持
5	统一数据管理,中级医疗决策支持
6	全流程医疗数据闭环管理,高级医疗决策支持
7	医疗安全质量管控,区域医疗信息共享
8	健康信息整合,医疗安全质量持续提升



图9 电子病历系统功能应用水平级别具体要求词云图

表4 电子病历评级主要决策支持要求及数据挖掘相关任务方法

级别	具体要求	数据挖掘相关任务方法
4	实现药品配伍、相互作用自动审核、合理用药监测等功能	关联规则挖掘等
5	提供临床诊疗规范、合理用药、临床路径等统一的知识库 病历、报告等的结构化、智能化生成	模式挖掘、知识图谱等 文本分析、摘要抽取等
6	检查、检验、治疗、手术、输血、护理等实现全流程数据跟踪与闭环管理,并依据知识库实现全流程实时数据核查与管控 形成全院级多维度医疗知识库体系(包括症状、体征、检查、检验、诊断、治疗、药物合理使用等相关联的医疗各阶段知识内容),能够提供高级别医疗决策支持	流程挖掘、异常检测等 知识图谱等
7	医疗质量与效率监控	异常检测、风险预测等
8	整合跨机构的医疗记录、健康记录、体征检测、随访信息用于本部门医疗活动	多源异构数据融合等

《电子病历系统功能应用水平分级评价方法及标准(试行)》中多次强调知识库的应用,特别是知识库为4级及以上级别电子病历应用提供的支撑。以知识图谱为代表的知识库的构建将助力全院级多维度医疗知识库体系建设。关联规则、自然语言处理、流程挖掘等数据挖掘理论和方法将助力电子病历系统功能应用水平的提升。

5 电子病历建设存在的挑战及相关建议

5.1 电子病历建设存在的挑战

电子病历的使用和开发利用将成为一种趋势。但当前电子病历还处于建设阶段,电子病历的应用水平比较低。据中国医院协会信息专业委员会(CHIMA)发布的《2017—2018年度中国医院信息化状况调查》,目前三级医院的功能应用普遍集中在3级、4级,三级以下医院主要在2级、3级。电子病历建设过程中仍然面临较多问题,具体如下。

(1) 电子病历建设标准不统一

电子病历是对患者诊疗过程的详细记录,电子病历数据的共享可以更好地发挥其价值。电子病历系统是逐渐发展起来的,各部门、科室、医院以及医疗服务机构的电子病历建设标准不尽相同,缺乏顶层设计和统筹推进,出现“数字孤岛”等现象,严重阻碍了电子病历数据跨部门、跨机构的数据共享,限制了电子病历系统的互联互通。

(2) 数据共享与隐私保护

电子病历开放和共享过程中的数据共享和隐私保护是不可回避的问题,数据如何共享、数据开放到什么级别以及如何保护患者隐私需要重点研究。我国

在电子病历开放和共享方面处于初级阶段,很多情况下电子病历数据仅供内部或者科研使用,甚至很多数据静置于数据库中,无法发挥其巨大价值。因此,如何在保护好患者隐私的基础上开放和共享电子病历以发挥其价值是需要探讨的问题。

(3) 挖掘结果的可解释性

近年来,数据挖掘理论和方法发展迅速,尤其是深度学习算法的提出和发展,使得数据挖掘方法在多领域广泛应用,模型的表示学习能力大幅度提高。但随着模型复杂度的增加,算法可解释性越来越差。尽管关于数据挖掘的研究已经很多,但挖掘结果的可解释性仍是一个难题,并且已经严重影响其在特定领域的应用,尤其是医疗领域。如果一个决策是可以解释的,那么决策者可以更好地评估其优缺点。因此,结果的可解释性是需要关注和解决的重要问题。

(4) 挖掘结果应用责任主体及关系

基于挖掘结果的临床决策支持的辅助诊断在医疗责任认定方面存在巨大挑战,明确责任主体及其关系能更好地促进数据挖掘结果的应用。当前法律标准还未涉及此问题,因此需要进一步完善和健全法律法规体系,以保障电子病历的分析应用。

5.2 相关建议

基于当前电子病历分析、应用存在的问题,本文给出以下建议。

(1) 加强顶层设计和统筹推进

电子病历建设融合临床、信息工程、管理等多领域,涉及管理、技术以及方法等多方面,是较复杂的问题。可以借助元决策以及系统工程理论方法,加强顶层设计,强调整体性和系统性,统筹推进电子

病历建设,实现提升医疗服务效率和水平的目标。

(2) 诊疗服务环节全覆盖

梳理优化业务流程,实现医嘱、用药等环节的闭环管理,并基于物联网、互联网等技术实现数据感知和存储,实现数字追踪及可追溯,积累的数据也可为业务控制与优化提供支持。

(3) 探索隐私保护与数据共享的机制与方法

首先要建立健全隐私保护机制和风险审查机制,建立相关机构对隐私保护进行监察和审查。其次要在保护隐私的前提下促进数据的共享,使其产生更大的价值。一方面可以从机制上完善数据共享流程,另一方面也可从模型算法层面探索如何在保障数据隐私的前提下训练和共享模型。

(4) 综合算法性能和可解释性

随着机器学习理论算法以及计算机软硬件的不断发展,算法性能得到很大提升。特别是基于深度学习的广泛应用,模型拟合能力更强。但是复杂模型的可解释性大大降低,这限制了其在特殊领域的应用,特别是医疗领域。需要综合算法性能和可解释性来提高算法的综合效用。建模前主要对数据进行预处理,可以通过数据探索性分析以及可视化等方法对问题有更深入的认识。建立解释性程度比较高的模型(如决策树、逻辑回归等)可以使得到的结果更具有解释性。针对一些“黑箱”算法,可以在建模后采用隐层分析、敏感度分析以及知识蒸馏等方法来提高可解释性^[23]。

(5) 发挥临床诊疗决策支持的功能

面对临床医护人员相对不足以及人们对医疗水平的要求不断提高等现实问题,基于电子病历数据挖掘的决策支持将在辅助临床诊疗决策方面发挥巨大价值,已

有电子病历数据脱敏、集成及二次利用等平台在医院医疗能力、质量、效率的分析中被采用^[24]。随着电子病历应用的不断加深,各地医院尝试应用临床决策支持系统,但是由于信息系统性能、责任主体以及观念等原因,实际应用不足。临床决策支持系统有助于减少医疗差错和医疗纠纷情况的发生,通过综合分析多模态信息(如人口统计学信息、临床表现、检查结果以及既往史等),进一步保障了诊疗过程(问、检、诊、疗等)的全面性、科学性以及规范性,提升了临床决策和医疗服务的质量。因此,需要克服当前临床决策支持系统的应用阻碍,发挥其在临床决策中的重要作用。

6 结束语

在政府大力推进智慧医院建设的背景下,本文对电子病历的分析利用框架进行探讨。本文首先介绍了电子病历的内涵及其生成,之后对其分析利用框架进行梳理,对其分析流程进行总结,从诊断、治疗以及管理3个角度对其应用进行阐述。然后,本文分析了电子病历分析对智慧医院相关评级的影响,通过对评级整体和具体要求分析,指出电子病历分析挖掘是智慧医院较高等级的要求,将助力高等级智慧医院建设。最后,本文从数据、模型以及应用层面探讨当前电子病历分析利用中存在的问题,并据此给出相关建议。在未来的电子病历分析利用建设中,可参考智慧医院相关分析的具体要求,重点关注电子病历建设的顶层设计、电子病历数据共享机制、电子病历数据挖掘模型算法和解释性以及电子病历分析挖掘结果的临床应用等方面的问题。

参考文献:

- [1] 胥婷, 崔文彬, 于广军. 我国智慧医院建设现状及发展路径[J]. 中国医院, 2020, 24(3): 1-3, 10.
XU T, CUI W B, YU G J. Construction status and development path of smart hospitals in China[J]. Chinese Hospitals, 2020, 24(3): 1-3, 10.
- [2] 戴明锋, 孟群. 医疗健康大数据挖掘和分析面临的机遇与挑战[J]. 中国卫生信息管理杂志, 2017, 14(2): 126-130.
DAI M F, MENG Q. Opportunities and challenges in data mining and data analysis of health care big data[J]. Chinese Journal of Health Informatics and Management, 2017, 14(2): 126-130.
- [3] 阮彤, 邱加辉, 张知行, 等. 医疗数据治理——构建高质量医疗大数据智能分析数据基础[J]. 大数据, 2019, 5(1): 12-24.
RUAN T, QIU J H, ZHANG Z X, et al. Medical data governance: building the data foundation for intelligent analysis of high quality medical big data[J]. Big Data Research, 2019, 5(1): 12-24.
- [4] 董诚, 林立, 金海, 等. 医疗健康大数据: 应用实例与系统分析[J]. 大数据, 2015, 1(2): 78-89.
DONG C, LIN L, JIN H, et al. Big data in healthcare: applications and system analytics[J]. Big Data Research, 2015, 1(2): 78-89.
- [5] 马锡坤, 杨国斌, 于京杰. 国内电子病历发展与应用现状分析[J]. 计算机应用与软件, 2015, 32(1): 10-12, 38.
MA X K, YANG G B, YU J J. Analysing the development and application status of electronic medical records in China[J]. Computer Applications and Software, 2015, 32(1): 10-12, 38.
- [6] 王秋霞, 刘利, 杜晓莉. 美国医疗信息化建设特点及其经验启示[J]. 卫生经济研究, 2019, 36(12): 50-52.
WANG Q X, LIU L, DU X L. The characteristics of American medical information construction and its enlightenment[J]. Health Economics Research, 2019, 36(12): 50-52.
- [7] MILLS E J, THORLUND K, IOANNIDIS J P A. Demystifying trial networks and network meta-analysis[J]. British Medical Journal, 2013, 346: f2914.
- [8] 黄梦醒, 李梦龙, 韩惠蕊. 基于电子病历的实体识别和知识图谱构建的研究[J]. 计算机应用研究, 2019, 36(12): 3735-3739.
HUANG M X, LI M L, HAN H R. Research on entity recognition and knowledge graph construction based on electronic medical records[J]. Application Research of Computers, 2019, 36(12): 3735-3739.
- [9] 阮彤, 高炬, 冯东雷, 等. 基于电子病历的临床医疗大数据挖掘流程与方法[J]. 大数据, 2017, 3(5): 83-98.
RUAN T, GAO J, FENG D L, et al. Process and methods of clinical big data mining based on electronic medical records[J]. Big Data Research, 2017, 3(5): 83-98.
- [10] 熊英. 中文出院小结自动生成方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
XIONG Y. Research on automatic generation of Chinese discharge summary[D]. Harbin: Harbin Institute of Technology, 2019.
- [11] 金叶. 基于人工智能构建诊断肺结节的良恶性多种病理类型的分类模型研究[D]. 济南: 山东大学, 2020.
JIN Y. Research on the classification model of diagnosing multiple pathological types of benign and malignant pulmonary nodules based on artificial intelligence[D]. Jinan: Shandong University, 2020.
- [12] 李季, 丁凤一, 李翔宇. 基于电子病历数据挖掘的疾病危重度动态预测研究[J]. 信息资源管理学报, 2017, 7(4): 38-43.
LI J, DING F Y, LI X Y. Study of dynamic forecasting model based on EMR data mining for disease severity[J]. Journal of Information Resources Management, 2017, 7(4): 38-43.

- [13] EL-RASHIDY N, EL-SAPPAGH S, ABUHMED T, et al. Intensive care unit mortality prediction: an improved patient-specific stacking ensemble model[J]. IEEE Access, 2020, 8: 133541-133564.
- [14] 韩冬, 李其花, 蔡巍, 等. 人工智能在医学影像中的研究与应用[J]. 大数据, 2019, 5(1): 39-67.
HAN D, LI Q H, CAI W, et al. Research and application of artificial intelligence in medical imaging[J]. Big Data Research, 2019, 5(1): 39-67.
- [15] 蒋杰伟. 基于裂隙灯影像的小儿白内障自动诊断和预测方法研究[D]. 西安: 西安电子科技大学, 2018.
JIANG J W. Research on automatic diagnosis for predictic cataract based on slit-lamp image[D]. Xi'an: Xidian University, 2018.
- [16] 文玉敏, 董兴鲁, 李平. 糖尿病肾病证候及用药规律的数据挖掘研究[J]. 中华中医药杂志, 2015, 30(10): 3665-3670.
WEN Y M, DONG X L, LI P. Data mining research on syndromes distribution and prescription regularity of diabetic nephropathy based on periodicals[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2015, 30(10): 3665-3670.
- [17] 任建业, 许鸣, 陆嘉惠. 基于数据挖掘的中医临床用药规律和证型研究进展[J]. 中华中医药杂志, 2017, 32(10): 4579-4582.
REN J Y, XU M, LU J H. Research progress on clinical medication regularity and syndrome type based on data mining[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2017, 32(10): 4579-4582.
- [18] 郭崇慧, 陈静锋, 魏伟. 元决策下的临床路径管理模式研究[J]. 医学与哲学, 2017, 38(14): 6-9.
GUO C H, CHEN J F, WEI W. Research on clinical pathway management mode with meta decision-making[J]. Medicine & Philosophy, 2017, 38(14): 6-9.
- [19] 赵璇. 基于数据挖掘技术的冠心病费用研究[D]. 北京: 北京中医药大学, 2018.
ZHAO X. Research on the cost of coronary heart disease based on data mining technology[D]. Beijing: Beijing University of Chinese Medicine, 2018.
- [20] 汤羽, 袁炜, 芮磊, 等. 基于数据挖掘算法的慢病住院费用预测模型分析[J]. 现代医药卫生, 2019, 35(4): 481-485.
TANG Y, YUAN W, RUI L, et al. Analysis of hospitalization cost prediction model of chronic disease based on data mining algorithm[J]. Journal of Modern Medicine & Health, 2019, 35(4): 481-485.
- [21] 陈默, 蔡苗, 黄阿红, 等. 基于K-means聚类与支持向量机的大病患者住院费用影响因素与控制策略研究[J]. 中国医院管理, 2019, 39(5): 45-47, 53.
CHEN M, CAI M, HUANG A H, et al. Analysis of influencing factors of hospitalization expenditure of patients with serious disease based on K-means clustering and support vector machine[J]. Chinese Hospital Management, 2019, 39(5): 45-47, 53.
- [22] 郭珉江. 数据挖掘技术在疾病诊断相关分组中的应用[D]. 长沙: 中南大学, 2009.
GUO M J. Research on the application of data mining technology in disease related groups[D]. Changsha: Central South University, 2009.
- [23] 陈珂锐, 孟小峰. 机器学习的可解释性[J]. 计算机研究与发展, 2020, 57(9): 1971-1986.
CHEN K R, MENG X F. Interpretation and understanding in machine learning[J]. Journal of Computer Research and Development, 2020, 57(9): 1971-1986.
- [24] 包小源, 张凯, 金梦, 等. 基于数据空间的电子病历数据融合与应用平台[J]. 大数据, 2019, 5(6): 47-61.
BAO X Y, ZHANG K, JIN M, et al. A data-space based platform for the integration and application of electronic health records[J]. Big Data Research, 2019, 5(6): 47-61.

作者简介



徐良辰(1992-),男,大连理工大学系统工程研究所博士生,主要研究方向为数据挖掘与知识发现。



郭崇慧(1973-),男,博士,大连理工大学系统工程研究所教授、博士生导师,主要研究方向为大数据分析
与挖掘、健康医疗大数据、商务智能与商务分析。

收稿日期: 2020-10-25

基金项目: 国家自然科学基金资助项目(No.71771034)

Foundation Item: The National Natural Science Foundation of China(No.71771034)