

医疗大数据在精准医疗领域的要素挖掘与应用

任银华

杭州卓健信息科技股份有限公司 浙江杭州 312000

摘要: 在个性化医疗需求攀升与医疗数字化转型的浪潮中, 医疗大数据作为精准医疗的“数字基石”, 其要素挖掘与应用决定着诊疗模式变革的深度。文章剖析医疗大数据在精准医疗领域的战略价值与发展现状, 详述数据清洗整合、特征提取等挖掘方法, 探讨疾病诊断预测、个性化治疗等应用场景, 直面数据安全、标准缺失等挑战并提出应对策略, 为推动医疗大数据赋能精准医疗提供理论与实践路径。

关键词: 医疗大数据; 精准医疗; 数据要素挖掘; 数据应用; 医疗信息化

引言

伴随医疗信息化建设的持续推进, 电子病历、医学影像、基因测序等多源数据呈爆发式增长, 构筑起庞大的医疗大数据资源体系。精准医疗以患者个体差异为核心, 旨在实现疾病的精确诊断、个性化治疗与有效预防, 而医疗大数据恰好为其提供了海量、多维的数据支撑。但医疗数据存在结构复杂、质量参差不齐等问题, 如何高效挖掘其中的数据要素, 并将其转化为精准医疗的实际应用, 已成为医疗领域亟待攻克的关键课题, 对提升医疗服务质量、推动医疗模式革新意义深远。

一、医疗大数据在精准医疗领域数据要素挖掘与应用的重要性及现状

(一) 医疗大数据对精准医疗发展的意义

医疗大数据是精准医疗发展的核心引擎。其涵盖的电子病历记录了患者的症状、病史、治疗过程等信息, 医学影像包含病灶形态等直观数据, 基因数据则揭示了个体疾病易感性的内在根源。通过对这些数据的深度挖掘与分析, 能够精准识别患者的疾病特征与个体差异, 为疾病的早期诊断、个性化治疗方案制定提供科学依据。例如, 利用大量患者的基因数据与临床治疗效果数据进行关联分析, 可以发现特定基因与药物疗效的关系, 从而为患者选择最有效的治疗药物和剂量, 实现精准用药。

此外, 医疗大数据还能助力医疗研究, 通过对大规模数据的分析, 挖掘疾病发生、发展的规律, 推动医学知识的更新与创新, 为精准医疗的持续发展注入动力。

(二) 医疗大数据数据要素挖掘与应用的研究现状概述

目前, 医疗大数据的数据要素挖掘与应用研究已取得一定成果, 但仍面临诸多挑战。在研究层面, 国内外学者已开展了大量关于医疗数据挖掘算法、模型的研究, 如利用机器学习算法进行疾病预测、基于深度学习的医学影像识别等, 部分研究成果已在临床实践中得到初步应用。在应用方面, 一些大型医疗机构和科技企业开始尝试搭建医疗大数据平台, 整合多源数据, 开展精准医疗服务。然而, 整体研究与应用仍处于发展阶段。数据质量参差不齐、数据标准不统一导致数据整合困难; 数据挖掘技术在医疗领域的适应性和可靠性有待进一步验证; 医疗大数据的应用场景还不够广泛, 尚未形成成熟的应用模式, 距离实现全面精准医疗仍有较大差距。

(三) 数据要素挖掘应用对医疗模式变革的作用

数据要素挖掘应用推动着医疗模式从经验驱动向数据驱动转变。传统医疗模式主要依赖医生的临床经验和直觉进行诊断与治疗, 而医疗大数据的挖掘应用能够将海量的医疗数据转化为有价值的信息和知识, 为医生提供客观、准确的决策依据。通过对大量患者数据的分析, 医生可以了解不同疾病的发病特点、治疗效果及预后情况, 从而制定更加科学合理的治疗方案。同时, 数据要素挖掘应用还能促进医疗服务的个性化和智能化发展。基于患者的个体数据, 为其提供定制化的健康管理方案和疾病预防措施; 利用智能算法辅助医生进行诊断, 提

作者简介: 任银华 (1981-10-28), 男, 汉族, 浙江绍兴人, 硕士研究生, 副总裁, 研究方向: 互联网医疗、数字医疗、智慧医疗、全病程个案管理服务、医疗大数据数据要素应用场景。

高诊断的准确性和效率。这种转变将极大地提升医疗服务质量，改善患者的就医体验，推动医疗模式向更加精准、高效、个性化的方向发展。

二、医疗大数据的数据要素挖掘方法

（一）医疗大数据的清洗、整合与预处理

医疗大数据来源广泛、结构复杂，数据清洗、整合与预处理是挖掘数据要素的基础环节。数据清洗旨在去除数据中的噪声、重复数据和错误数据，提高数据质量。通过识别和纠正电子病历中的录入错误、去除重复的检验检查记录，确保数据的准确性和一致性。数据整合则是将来自不同系统、不同格式的医疗数据进行融合，打破数据孤岛。例如，将电子病历系统中的文本数据、影像归档系统中的图像数据以及实验室信息系统中的检验数据进行整合，构建统一的数据存储与管理平台。预处理过程包括数据标准化、归一化处理，对数据进行编码和转换，使其符合后续分析的要求，为数据要素的有效挖掘奠定基础。

（二）数据特征提取与关键要素识别方法

数据特征提取与关键要素识别是从海量医疗数据中获取有价值信息的关键步骤。特征提取是指从原始数据中提取能够反映数据本质特征的变量或指标，如从电子病历中提取患者的年龄、性别、症状、诊断结果等关键信息，从医学影像中提取病灶的大小、形状、密度等特征。关键要素识别则是在提取的特征中筛选出对研究问题具有重要影响的要素。通过统计分析、机器学习中的特征选择算法，如卡方检验、递归特征消除等方法，评估每个特征的重要性，去除冗余和无关特征，保留最具代表性和预测性的关键要素。这些关键要素能够更准确地描述疾病特征和患者个体差异，为后续的数据分析和模型构建提供核心数据支撑。

（三）基于机器学习的医疗数据知识发现技术

基于机器学习的知识发现技术是挖掘医疗大数据潜在价值的重要手段。机器学习算法能够自动从大量数据中学习模式和规律，实现对疾病的诊断、预测和治疗效果评估。在疾病诊断方面，利用深度学习算法对医学影像进行分析，如卷积神经网络（CNN）可以自动识别X光、CT、MRI等影像中的病灶特征，辅助医生进行疾病诊断，提高诊断的准确性和效率。在疾病预测领域，通过训练机器学习模型，如随机森林、支持向量机等，利用患者的历史数据预测疾病的发生风险、复发概率以及治疗效果，为医生制定个性化治疗方案提供参考。此外，

机器学习还可用于药物研发，通过分析大量的临床数据和生物信息数据，预测药物的作用机制和潜在疗效，加速药物研发进程。

三、医疗大数据在精准医疗领域的应用场景

（一）基于大数据分析的疾病精准诊断与预测

基于大数据分析的疾病精准诊断与预测是医疗大数据在精准医疗领域的重要应用。通过整合患者的临床数据、基因数据、影像数据等多源信息，利用大数据分析和机器学习算法，构建疾病诊断和预测模型。在疾病诊断时，模型可以对患者的症状、检查结果等数据进行综合分析，与大量已确诊病例的数据进行比对，快速准确地判断疾病类型和严重程度，为医生提供诊断建议。在疾病预测方面，根据患者的个体特征和历史健康数据，预测疾病的发生风险。例如，通过分析糖尿病患者的血糖、血压、血脂等数据以及家族病史，预测患者发生糖尿病并发症的概率，并提前采取干预措施，降低疾病发生风险，实现疾病的早发现、早诊断、早治疗。

（二）个性化治疗方案制定与医疗决策支持

个性化治疗方案制定与医疗决策支持是精准医疗的核心目标，医疗大数据为此提供了有力支持。利用患者的基因数据、疾病史、治疗反应等数据，结合大数据分析和人工智能技术，能够为患者量身定制个性化的治疗方案。通过分析大量相似病例的治疗效果，评估不同治疗方法对特定患者的适用性和有效性，为医生选择最佳治疗方案提供参考。同时，医疗大数据还可以实时监测患者的治疗过程和身体反应，根据数据变化及时调整治疗方案。此外，借助医疗决策支持系统，将大数据分析结果与临床指南、专家经验相结合，为医生提供全面、准确的决策信息，辅助医生做出更加科学、合理的医疗决策，提高治疗效果和患者的生存率。

（三）全生命周期健康管理与疾病预防

医疗大数据在全生命周期健康管理与疾病预防中发挥着重要作用。通过收集和分析个人从出生到老年的健康数据，包括体检数据、生活习惯数据、基因数据等，建立个人健康档案。利用大数据分析和预测模型，对个人的健康状况进行动态监测和评估，预测疾病发生的可能性，并提供个性化的健康管理建议。例如，针对高血压高危人群，根据其血压变化、饮食、运动等数据，制定个性化的饮食和运动计划，提醒患者按时服药，定期进行检查，预防高血压的发生和发展。同时，通过对人群健康数据的分析，还可以发现疾病的流行趋势和危险

因素，为公共卫生政策的制定和疾病预防措施的实施提供数据依据，实现疾病的群防群控，提高公众的整体健康水平。

四、医疗大数据在精准医疗应用中面临的挑战及对策

（一）医疗数据的安全与隐私保护难题及应对

医疗数据包含患者的个人身份信息、健康状况、疾病史等敏感信息，数据安全与隐私保护是医疗大数据应用面临的首要难题。在数据收集、存储、传输和使用过程中，存在数据泄露、篡改等风险。为应对这一挑战，首先需要加强法律法规建设，明确医疗数据的安全与隐私保护的责任和义务，加大对数据泄露等违法行为的惩处力度。其次，采用先进的数据加密技术，对敏感数据进行加密处理，确保数据在传输和存储过程中的安全性。同时，建立严格的数据访问控制机制，对数据的使用进行权限管理，只有经过授权的人员才能访问和使用数据。此外，还应加强数据安全意识教育，提高医疗从业者和患者的数据安全意识，共同维护医疗数据的安全与隐私。

（二）数据标准不统一与共享壁垒问题及解决

医疗数据来源多样，不同医疗机构、不同系统之间的数据标准不统一，形成了数据共享壁垒，严重制约了医疗大数据的整合与应用。解决这一问题，需要建立统一的医疗数据标准体系，包括数据格式、编码规则、术语定义等方面的标准，确保数据的一致性和兼容性。政府和行业协会应发挥主导作用，组织制定国家标准和行业规范，并推动医疗机构和相关企业遵循标准。同时，加强数据共享平台建设，打破数据孤岛，实现医疗数据在不同机构之间的安全共享和交换。通过建立数据共享激励机制，鼓励医疗机构和企业积极参与数据共享，促进医疗大数据的流通和应用，提高医疗资源的利用效率。

（三）技术瓶颈与专业人才短缺困境及突破

医疗大数据的挖掘与应用涉及数据科学、医学、计算机科学等多个领域，目前面临技术瓶颈和专业人才短

缺的双重困境。在技术方面，现有的数据挖掘算法和模型在处理复杂医疗数据时，存在准确性、可靠性不足等问题，且部分技术难以适应医疗领域的特殊需求。为突破技术瓶颈，需要加大科研投入，鼓励高校、科研机构和企业开展联合研发，针对医疗大数据的特点，开发更加精准、高效的挖掘算法和模型。在人才培养方面，应加强跨学科专业人才的培养，在高校设置相关专业和课程，培养既懂医学知识又掌握数据科学技术的复合型人才。同时，开展在职人员培训，提高医疗从业人员的数据处理和分析能力，为医疗大数据在精准医疗领域的应用提供人才保障。

结论

医疗大数据的数据要素挖掘与应用是推动精准医疗发展的核心路径。文章系统阐述了其在精准医疗领域的重要意义、研究现状、挖掘方法、应用场景以及面临的挑战与对策。尽管目前面临数据安全、标准统一、技术人才等诸多挑战，但随着技术的不断进步、政策的逐步完善和人才培养体系的健全，医疗大数据将在精准医疗领域发挥更大价值。未来，需持续强化数据安全防护，完善数据标准体系，突破技术瓶颈，加强人才队伍建设，推动医疗大数据与精准医疗深度融合，为提升医疗服务质量、实现全民健康目标提供强大动力。

参考文献

- [1] 李娇. 基于大数据 Logistic 回归模型的医院线上诊疗数据精准推荐 [J/OL]. 自动化技术与应用, 1-5 [2025-05-20].
- [2] 祝丽燕. 大数据在医疗耗材企业精准营销中的应用及策略优化 [J]. 张江科技评论, 2024, (11): 155-157.
- [3] 庞怡真. 医疗大数据技术在医院信息系统中的应用 [J]. 信息与电脑 (理论版), 2024, 36 (15): 213-215.
- [4] 殷明雪. 健康医疗大数据建设与隐私保护之冲突与协调 [J]. 医学与社会, 2023, 36 (02): 125-131.