

大语言模型在智慧医院数据挖掘中的实践

柴国丽

山西医科大学第二医院

DOI:10.32629/pe.v3i6.18029

[摘要] 在智慧医院发展进程中,积累了海量的医疗数据,并呈现多样化特征,而传统数据分析技术在提取语义表达以及不同模式之间的关系方面已经不能满足需求。因此,作为具有强大自然语言处理、语义表示学习以及知识推理能力的人工智能大语言模型,为智能数据分析提供了一条新路径。本文围绕大语言模型参与智慧医院数据挖掘的实践展开研究,从底层技术、 workflow 和技术路线三个层面介绍了其在多源异构数据整合、非结构化文档处理、医疗服务关系发现及医院管理方面的能力。在此基础上,通过效果评估探讨大语言模型对提高医师诊断精准度及改善医院运营的作用。

[关键词] 大语言模型; 智慧医院; 数据挖掘; 医疗信息化

中图分类号: TU246.1 **文献标识码:** A

The Practice of Large Language Models in Smart Hospital Data Mining

Guoli Chai

The Second Hospital of Shanxi Medical University

[Abstract] In the process of developing smart hospitals, a vast amount of medical data has been accumulated, exhibiting diverse characteristics. Traditional data analysis techniques can no longer meet the demands for extracting semantic expressions and relationships between different patterns. Therefore, as artificial intelligence large language models with powerful natural language processing, semantic representation learning, and knowledge reasoning capabilities, they provide a new path for intelligent data analysis. This paper studies the practice of large language models participating in data mining in smart hospitals, introducing their capabilities in multi-source heterogeneous data integration, unstructured document processing, medical service relationship discovery, and hospital management from three levels: underlying technology, workflow, and technical approach. On this basis, through performance evaluation, the paper explores the role of large language models in improving physician diagnostic accuracy and enhancing hospital operations.

[Key words] Large Language Model; Smart Hospital; Data Mining; Medical Informatization

引言

大语言模型 (LLM) 是一种基于海量文本数据训练的深度学习模型,其核心在于通过无监督学习从数据中自动提取语言模式和结构。该模型不仅能够生成连贯、自然的文本,还能在一定程度上理解输入文本的含义,并进行逻辑推理和信息检索,相较于目前的静态分析工具 FlowDroid 与 Soot,其精确度可提升 10%~400%。基于此,本文从智慧医院应用场景出发,系统探讨大语言模型参与医疗数据挖掘的实现方式与应用成效,以为智慧医院数据治理与应用创新提供参考。

1 大语言模型参与智慧医院数据挖掘的技术基础与应用逻辑

1.1 智慧医院数据类型特征与挖掘需求分析

在信息化、数字化背景下,智慧医院形成庞大的异构数据体

系,其包含的数据类型复杂多样,既有电子病历、检验检查结果、生命体征等结构化数据,又有症状、影像报告、护理记录等半结构化、非结构化数据;同时包含医疗器械使用记录、 workflow 信息、患者行为路径等多种类型的信息。由于每类信息所含数据的格式、抽象程度及更新频率不同,难以用传统单一规则或单一模板进行数据分析处理。结合实际需求,医生需要对患者的全貌以及疾病的发展过程有深入认识,医院管理人员则需要关注各个部门之间的数据关联及未来发展趋势,这对数据挖掘提出了从“数据统计”向“语义理解”和“知识发现”转变的要求。

1.2 大语言模型在医疗数据理解与语义建模中的技术优势

基于海量语料库进行深度学习的大规模语言模型,可以强建模复杂语义关系,突破传统基于人工编写或内置标记系统等特征工程实现手段,在海量训练材料基础上形成关于医学领域

知识及临床表述和上下文背景信息的认知,能够对非结构化医学文本有更高的理解程度和归纳能力^[1]。该类模型不仅能够识别出疾病、症状以及检验结果等显性信息,同时也能挖掘隐性的治疗方法及因果关系,并进行高层语义表示,从而更好地处理医生自由撰写的文本、表述形式及各个部门之间用词差异等问题,并为后续研究提供技术基础。

1.3 大语言模型嵌入智慧医院数据挖掘流程的基本模式

在智慧医院应用场景中,大语言模型通常以“能力模块”形式嵌入既有数据挖掘流程,并不涉及对原有系统架构的颠覆性改造。在数据采集与预处理阶段,模型可以帮助整理、标注口语化表达;在信息分析阶段,使用标准化数据构建跨源关联和语义对应关系;在最终的产出环节,可以将研究结果转化为便于医生或管理者理解的方式呈现出来,在不破坏现有信息系统的工作原理的情况下,引入了自然语言理解和演绎能力,实现“数据一知识一决定”的过程链接,并结合规定引擎、统计学模型等多项技术联合应用后,能够保证安全性以及可靠性的情况下提升大数据挖掘效果^[2]。

2 大语言模型在智慧医院数据挖掘中的实践路径

2.1 面向多源异构医疗数据的统一理解与整合挖掘路径

在智慧医院运作过程中,不同工作系统的产出信息具有不同来源、组成及意义层面,如电子病历系统以描述医学文献为主导,实验系统则突出数值化以及时间序列属性,管理程序则呈现过程性与资源配置的信息。为达成多元信息融合探究,可以借助大语言模型作为“统一语义中介”,不需要更改原有数据架构便可将多来源数据做语义化转译及场景对应。做法上就是在资料导入时,仍维持各系统原本的字段,再用这个模型标注出字段所代表的语义以及商业情境,还有其隐含关联性;在合并过程中基于语义一致性而非字段一致性来创建数据链接,降低异构系统间集成成本;在此基础上,可以构建单个研究视角用于整合不同系统间的数据进行分析研究。这样很好地克服了传统数据仓库范式中先结构后语义的问题,对异构数据融合更加现实的认知。

2.2 基于大语言模型的非结构化医疗文本深层挖掘实践

在智慧医院实际运行中,大量关键信息以非结构化的文字形式存在于病历、病程记录、影像诊断报告和出院小结当中,其描述严重依赖医生个体笔迹并具备诸如隐含意义、多义性和跨域关联等特点。针对该问题,实践中引入大语言模型对原始医疗文本进行逐条解析,通过上下文构建模式化手段,实现自动抽取疾病的描述、诊断推理、治疗方法等语义信息,以及对原文字进行改动即可生成结构化语义表征的目的。在具体流程上,模型对文本进行词句级别的切分及语义编码,再根据上下文关系消除人称指代歧义性,最终得到可计算分析的文本向量化表示结果^[3]。为支持病例相似性分析与文本聚类挖掘,通常采用向量相似度度量方式评估不同医疗文本之间的语义接近程度,其计算形式如下:

$$Sim(T_i, T_j) = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|} \quad (1)$$

其中, v_i 和 v_j 为大语言模型生成的文本语义向量。基于该方法,可在不依赖人工规则的情况下实现病情描述模式挖掘、异常文本识别及病例分组分析,为后续临床决策支持与科研数据整理提供可靠基础。

2.3 大语言模型驱动的临床业务关联关系挖掘实践

智慧医院应用过程中,医疗业务流是连续进行的,在诊疗环节中存在大量隐性的关联性,单纯的数据统计分析或单项评价难以发现其内在关联。实践中,运用人工智能对病历文书、医嘱单以及化验单等进行处理,从事件识别到关系抽取,构建“诊疗行为一时间线一业务结果”的关联体系。具体方法是将模型从非结构化文本中抽取疾病诊断、检查结论及治疗计划等关键业务信息,并进行事件排序后建立可供分析挖掘的临床业务关系图谱;通过对高频关系及特异性连接模式的挖掘,能发现过多检查项目、延迟步骤或者偏离治疗路径等问题,为流程优化提供依据。为便于说明大语言模型在临床业务关联挖掘中的典型应用形式,其主要挖掘对象与分析价值如表1所示。

表1 大语言模型驱动的临床业务关联关系挖掘内容示例

挖掘对象	关联关系类型	主要数据来源	实际应用价值
诊断行为与检查项目	因果关联	病历文本、检查记录	优化检查决策路径
治疗措施与病情变化	时间关联	病程记录	支持疗效评估
医嘱执行与住院时长	过程关联	医嘱系统	改进流程效率
科室协作与转诊行为	业务关联	会诊与转诊记录	提升协同水平

2.4 面向医院精细化管理的数据挖掘应用实践

大数据时代下,智慧医疗服务管理的重点由单纯的统计分析转变为对服务过程及服务模式进行描述。在实践中,可借助AI技术将各系统中收集的管理数据、业务记录以及文字评价纳入其中,进而将分散在不同方面的管理信息归入同一范畴的概念体系当中。例如,在门诊管理流程中,模型可以将挂号信息表、排班表及患者评价文档进行整合,挖掘高峰期时段内资源配置情况和顾客满意度之间的相关性特征,为优化工作提供依据。针对病房管理,利用深度学习方法,可挖掘影响患者住院天数最关键的工作环节,这为医院有针对性地改进提供了可能。该模型得到的研究结果还可以转化为结构化指标及自然语言解释,降低管理者理解应用数据的门槛。这种语义驱动的大数据分析方法,可以将管理经验由感性转向理性,从而达到医院运行效率以及管理水平同步提升的效果^[4]。

3 大语言模型智慧医院数据挖掘应用的效果评估

3.1 数据挖掘结果在临床决策支持中的应用效果评估

尽管数据挖掘的应用效果不能立刻替代医疗实践中医生对诊断的结论,但其更侧重于呈现方式及决策过程的变化。在评估中将模型的解读结果整合至医师原有的工作台界面,并着重考察其整合患者信息、强调关键检查项目、归纳治疗意见的程度。

对比模型加入前后医生浏览病历花费的时间以及信息遗漏情况发现,当相关信息被结构化及语义汇总呈现后,医生对患者的整体认识更加深刻。在部分疑难病例中,甚至有模型输出结果提示医生关注先前被忽略的既往检查或用药记录,但最终判断仍由医生完成。

3.2 数据挖掘实践对医院管理效率提升的影响分析

对于医院管理部门而言,数据挖掘结果的好坏不是看某个具体指标数值变化了多少,而是看管理部门采取了哪些措施进行处理。因此,选取门诊运行情况、床位利用率、人力资源配置等情况作为效果评价指标,对比应用该模型前后管理部门获取相关数据的方式及频率。根据实验的结果,通过对业务资料及文字记录的研究,一些管理问题能够提前预判出来。如高峰期资源不足的原因是什么?流程衔接不通畅的环节在哪里?这样管理者就有了更加明确的决策依据,就不会出现反复查看多个系统的情况。这种影响分析着眼于数据挖掘是否改变了管理思路,而不是只关注短期绩效的变化。

3.3 大语言模型参与数据挖掘的准确性与稳定性评估

在大数据环境下,对大语言模型的可靠性进行评估时,强调该技术在应用过程中的稳定性和可控性。准确性评估主要采用抽样比对的方式,即根据部分语料库结果进行比较和确认,保证重要字段如症状、体征、用药等方面的抽取完整性和正确率均较高。为了测试鲁棒性,对不同时间段的数据样本以及不同部门的文件进行多次操作,以观察模型输出的形式和构建变化情况^[5]。实验表明,在较为规范的文本格式下,模型表现更为稳定;在风格差异较大的情况下,仍需人工再次核查。这也表明大语言模型在大数据挖掘中更适合作为辅助分析工具存在,其使用效果与数据质量和应用场景密切相关。

3.4 智慧医院场景下数据挖掘应用的综合价值分析

在智慧医院运行环境中,数据挖掘更加关注对常规工作细节上的改进,而非单个参数的大幅提升。从实践角度来说,经过语义化处理的数据,可以帮助医生更快检索到当前治疗有关的信息,省去了在各个平台上查找信息的时间成本。在管理层面,

基于业务数据和文字信息分析发现的一些隐含问题以及资源冲突,可以帮助管理人员进行调整和参考。但是以上成果往往以补充的形式应用于业务中,并未影响原有决策边界,这导致其在真实工作场景中的可接受性更高一些。归根结底,其作用是提高信息整理的方式和利用效率,在反复应用的过程中,慢慢对医院运行稳定性和管理水平产生一定的促进作用。

4 结语

本文结合具体实践角度出发,对大语言模型应用于医疗数据挖掘的方法及效果进行研究。在智慧医院建设过程中面临的数据应用以及信息碎片化等问题,可以合理将大语言模型应用到现有信息系统中,进而更好地实现无序信息管理优化。并且能够帮助完成不同来源数据间的关联分析,并能稳定支持临床指引以及管理评价等工作。需要指出的是,大语言模型更适合作为资料整理和解读的一种手段来使用,其效果很大程度上取决于与某一具体流程的匹配程度。在今后还应结合不同类型医院及应用场景,进一步探讨模型的应用场景、运行稳定性及后期维护方法等内容,以期对智慧医院中的数据分析应用给出更具说服力的操作指导建议。

[参考文献]

- [1]李冠华,迟锐,孙思伟,等.大语言模型GPT-4在精神专科智慧医院建设中的应用探析[J].中国现代医生,2024(33):12-17.
- [2]白培发,黄宗浩,王奕.大模型在智慧医院的应用研究综述[J].计算机应用与软件,2024,41(7):1-5.
- [3]唐荣高,周全.基于数据挖掘技术的医院智慧能源管理平台设计与应用[J].中国医疗设备,2023,38(8):86-91.
- [4]冯羽.“IoT+数据挖掘”下医院智慧医疗健康管理系统设计[J].中国新技术新产品,2024(10):143-145.
- [5]那晓红.基于数据挖掘的智慧财务决策支持方法设计及应用[J].中国总会计师,2022(4):36-40.

作者简介:

柴国丽(1974--),女,汉族,山西洪洞人,硕士,高级工程师,研究方向:医院信息管理,网络数据安全。