

基于医院大数据的智能就诊路径实时推荐系统设计与应用

蒋宏建 陈健超

(江门市五邑中医院 江门 529099)

[摘要] **目的/意义** 优化患者就诊流程, 缩短在院就医时间与移步路径, 帮助患者预知各就诊环节时间。**方法/过程** 基于图论与机器学习算法模型, 融合医院空间布局、诊疗科室分布及患者就诊项目等数据信息, 构建智能就诊路径推荐模型; 基于医院大数据, 构建就诊时间预测模型, 预测患者在挂号、候诊、检查等各环节耗时; 设计并实现智能就诊路径实时推荐系统, 为患者提供清晰的就诊路径指引与较准确的耗时预估服务。**结果/结论** 智能就诊路径实时推荐系统融合就诊时间预测与就诊路径推荐能力, 提升了患者的就医体验和医院的服务效率。

[关键词] 就诊时间; 路径推荐; 就诊体验; 智慧医院

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2024.10.017

Design and Application of an Intelligent Real-time Treatment Path Recommendation System Based on Hospital Big Data

JIANG Hongjian, CHEN Jianchao

Wuyi Traditional Chinese Medical Hospital of Jiangmen, Jiangmen 529099, China

[Abstract] **Purpose/Significance** To optimize the patient's treatment process, to shorten the treatment time and moving path in the hospital, to help patients predict the time of each treatment process. **Method/Process** Based on graph theory and machine learning algorithm model, the intelligent treatment path recommendation model is constructed by integrating data information such as hospital spatial layout, diagnosis and treatment department distribution and patient treatment items. Based on the big data of hospital patients, the prediction model of patient treatment time is constructed to predict the time spent in registration, waiting for treatment or examination, etc. A real-time recommendation system for intelligent treatment paths is designed and implemented to provide patients with clear guidance and more accurate time-consuming estimation services. **Result/Conclusion** The intelligent real-time treatment path recommendation system integrates the capability of treatment time prediction and path recommendation, which significantly improves the treatment experience of patients and the service efficiency of hospitals.

[Keywords] treatment time; path recommendation; treatment experience; smart hospital

1 引言

[修回日期] 2024-07-21

[作者简介] 蒋宏建, 硕士, 高级工程师, 发表论文 1 篇;
通信作者: 陈健超, 高级工程师。

2023 年 5 月《关于开展改善就医感受提升患者体验主题活动的通知》^[1]指出要优化门诊流程, 缩

短患者在门诊的滞留时间;加强引导,明确当日检查检验结果回报患者的接诊流程,减少无序流动,以持续改善就医感受、提升患者体验。为解决当前患者对候诊时间缺乏预期、就诊项目组合复杂导致路径规划困难,以及人工导诊服务效率受限等问题,本研究基于就诊项目历史数据,构建就诊时间预测模型;结合医院信息系统的实时动态数据构建就诊项目路径规划模型,为患者提供最佳实时就诊项目路径推荐服务。该服务积极响应医院智慧服务分级评估标准体系^[2]中关于诊中服务四、五级的应用评估要求,融合移动互联网技术,支持患者通过移动设备即时查询诊疗科室具体位置、当前排队情况及预估等待时间,实现就诊路径的动态调整与优化,有效避免患者不必要的等待与往返,以提升医院整体服务效率与患者就医满意度。

2 研究现状

2.1 候诊时间预测研究

候诊时间预测对于改善患者体验和优化医疗资源配置至关重要^[3]。Hung G R等^[4]通过仿真软件构建排队模型预测候诊时间,模拟医院候诊场景,考虑人员流动、事件触发及病情复杂性等参数,但对建模细节的精确性要求极高,且预测结果稳定性易受参数配置准确性影响。Sun Y等^[5]运用分位数回归模型,基于分诊时的统计数据,选取第95百分位数作为候诊时间的代表值,具有一定严谨性,但其局限性在于难以全面捕捉影响候诊时间的多维度因素。相比之下,Curtis C等^[6]利用放射学信息系统丰富数据,结合神经网络、弹性网络、随机森林、线性回归及k近邻等多种机器学习模型,精准预测放射影像科候诊时间,其中弹性网络模型表现尤为突出。Hijry H等^[7]进一步对比SGD、Adam、RMSprop、AdaGrad等深度学习优化算法在候诊时间预测中的应用效果,发现SGD优化算法在减少预测误差方面表现最佳。郭琳琳等^[8]基于患者数据构建6种不同机器学习模型用于预测儿科急诊的候诊时间,确定最佳模型的决定系数为0.523,预测效果良好。

2.2 就医体验提升研究

为提升患者就医体验并缩短其在门诊的滞留时间,当前研究聚焦于多维度策略。Li N等^[9]研究依据患者就诊时间波动,动态分配预约时段,并引入基于时变优先级的排序机制,有效平衡即时预约与预约患者的服务优先级,减少预约患者的等待时间。李立等^[10]通过优化门诊“一卡通”系统,实现诊疗与结算流程无缝对接,简化了就诊流程,减少了患者非医疗等待时间。王元元等^[11]以患者为中心进行医院信息平台门诊流程优化再造,通过系统总线进行信息技术架构规划,开展业务流程梳理,从技术层面优化医院门诊就医流程。顾懂燕等^[3]研究表明,医护人员的服务态度、就诊环境的舒适度等软性服务因素是提升患者满意度不可忽视的关键。Ferreira D C等^[12]综述表明医患沟通、患者年龄也是影响患者满意度的关键因素。

2.3 研究现状评述

通过分析候诊时间预测和患者就医体验提升的研究现状,发现机器学习在提高预测准确性方面具有重要作用,技术创新和流程优化在改善患者体验方面取得显著成效,但整体上仍存在整合不足的问题,尤其是要将软性服务因素量化并整合到预测模型和流程优化中时。未来研究工作需要探索跨学科方法,将技术手段与人文关怀相结合,以实现更全面的候诊时间预测和就医体验优化。本研究基于医院实时大数据,结合各科室候诊状态、检查项目的先后依赖关系及并行执行的可能性,从患者具体诊疗项目需求出发,为患者推荐个性化、高效且合理的就诊项目流程,旨在最大限度地减少患者在门诊的滞留时间和无效移动,提升就医体验。

3 就诊路径实时推荐模型构建

3.1 就诊时间预测模型

通过整合医生门诊系统、门诊分诊系统数据可以对就诊时间预测模型进行训练,相关字段内容经过脱敏编码处理,取值时间为2023年10—12月,

主要字段内容包括医生科室编码、医生编码、医生年龄段、医生职级编码、患者性别、患者年龄段、当日预约人数、已就诊人数、就诊午别、诊断主诉编码、就诊时间，初始总样本数据为 38.5 万人次，经过数据清洗过滤异常字段样本数据，总体实验样本数据为 35.2 万人次，涉及 36 个科室，146 名医生。数据样例描述，见表 1。以 7:3 的比例将总体样本切分为训练数据集和测试数据集，训练模型采用线性回归 (linear regression, LR) 模型^[13]、随机森林 (random forest, RF)、极致梯度提升 (extreme gradient boosting, XGBoost) 模型^[14]和高速梯度提升器 (light gradient boosted machine, LightGBM) 模型^[15]，模型评价指标为均方误差 (mean

square error, MSE)，即预测值与实际值之差平方的期望值；均方根误差 (root mean square error, RMSE)，即预测值与实际值之差平方的期望值的平方根。MSE 和 RMSE 值越小，模型准确度越高； Δp 为平均偏差波动百分比范围。实验中，除了对单模型实验外，采用 Bagging 集成学习方法，将 LR 与 XGBoost，XGBoost 与 LightGBM 分为两组，取二者预测平均值进行结果融合。单模型中 XGBoost 预测结果较好；两组集成学习模型中 XGBoost + LightGBM 预测结果较好，见表 2。与单模型相比，集成学习模型可以弥补部分样本预测存在的偏差，总体预测结果较好。选择 XGBoost + LightGBM 组合模型构建系统。

表 1 数据样例描述

字段名称	字段类型	参数类型	字段解释	样例数据
医生科室编码	整数	自变量	医生科室 ID (脱敏)	13121
医生编码	整数	自变量	医生身份 ID (脱敏)	523
医生年龄段 (岁)	整数	自变量	0: (0, 20]; 1: (20, 25]; 2: (25, 35]; 3: (35, 45]; 4: (45, 55]; 5: (55, 200]	4
医生职级编码	字符串	自变量	P: 初级; M: 中级; H: 高级	M
患者性别	字符串	自变量	M: 男性; F: 女性	F
患者年龄段 (岁)	字符串	自变量	1: (0, 14]; 2: (14, 20]; 3: (20, 30]; 4: (30, 40]; 5: (40, 50]; 6: (50, 200]	4
诊断主诉编码	字符串	自变量	诊断主诉 ID	R42. X00
预约人数	整数	自变量	取值范围 0 ~ 200	24
已就诊人数	整数	自变量	取值范围 0 ~ 200	2
就诊午别	字符串	自变量	上午: M; 下午: A	A
就诊时间 (分钟)	浮点型	实际值	实际医生接诊耗时	32. 2
预测时间 (分钟)	浮点型	预测值	预测医生接诊耗时	29

表 2 模型预测评价

模型	模型表现		
	MSE	RMSE	Δp
LR	1. 427 4	1. 936 2	(-15. 2% , 35. 5%)
RF	1. 402 8	1. 817 3	(-13. 8% , 26. 6%)
XGBoost	1. 368 5	1. 758 3	(-13. 6% , 25. 8%)
LightGBM	1. 372 8	1. 762 8	(-13. 2% , 26. 2%)
LR + XGBOOST	1. 358 3	1. 735 6	(-12. 4% , 23. 5%)
XGBoost + LightGBM	1. 326 8	1. 727 3	(-11. 8% , 20. 3%)

3. 2 就诊路径实时推荐模型

3.2.1 分诊候诊实时数仓 分诊候诊实时数仓包括科室医生接诊信息和候诊患者信息，通过订阅分诊系统的预约、签到、接诊等消息实现实时数据推送，可供下游节点通过接口直接查询访问。分诊候诊实时数仓的主要数据字段为消息时间、门诊号、挂号流水号、挂号医生、流程状态 (预约、签到、接诊等)、状态发生时间、预约时间等。该实时数仓的数据生命周期为 2 天，定时清理过期数据，数

据请求响应要求达到 500 毫秒之内。分诊候诊实时

数仓构建流程，见图 1。

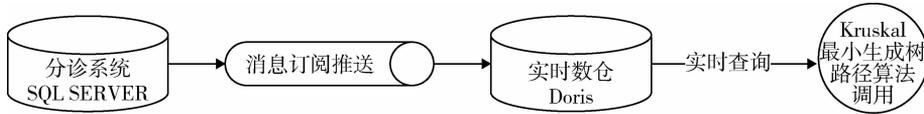


图 1 分诊候诊实时数仓构建流程

3.2.2 就诊项目路径字典 就诊项目路径字典旨在解决医院内各楼层与对应科室的定位问题，并量化各就诊点之间的步行时间。以江门市五邑中医院为例，眼科位于 18 号楼，与 1 号楼门诊楼之间步行距离约为 7~8 分钟。该字典核心字段包括科室名称、科室具体位置、诊室名称、诊室具体位置、相对于 1 号楼的步行距离、平均步行时间、就诊项目编码。鉴于患者预约挂号时通常选择科室内的具体医生，将医生编码及其出诊科室信息融入该字典。对于位于相同楼栋但不同楼层的就诊点，步行距离和步行时间则设为缺省值。

3.2.3 就诊项目知识字典 就诊项目知识字典是依据实际医院诊疗项目情况构建的重要参考工具，详细描绘业务逻辑的各种关系。其中，主要涵盖项目与项目类型之间的对应关系、项目与特定就诊点之间的对应关系，以及项目与项目之间的前置依赖关系等核心要素。例如，上腹部（肝、胆、脾、胰）1.5T 磁共振平扫属于医学影像科项目范畴，而肝功能 3 项则归类于检验科项目。医学影像科位于 1 号楼 2 层 202—205 室，而检验科则位于 1 号楼 4 层。此外，某些特定项目，如镇静检查必须在脑电图检查之前进行；B 超和胃镜项目要求患者空腹，可以集中安排。就诊项目知识字典的关键字段包括

项目编码、项目名称、项目优先级和项目就诊地点编码等。详尽的业务知识信息为路径规划提供有力依据，确保就诊项目能够按照合理高效的顺序推进。

3.2.4 基于 Kruskal 算法的路径推荐 Kruskal 算法^[16]是一种基于贪心策略的图论算法，用于解决最小生成树问题。其核心原理是从所有可用的边中，按照权重（或权值）从小到大的顺序进行选择。如果所选的边不会与已有的边形成环路，这条边就会被加入到最小生成树中。将各节点看作患者需要进行的各个环节，将边的权重理解为各就诊项目的候诊时间加就诊时间。关键字段包括节点的名称、就诊地点编码、当前候诊患者数，以及边的权重（即项目时长）等。然而，在规划就诊项目路径时，不能仅仅依赖 Kruskal 算法得出最优路径，还要结合就诊项目知识字典裁剪和优化路径。例如，患者先签到并等待接诊，医生接诊后开具医嘱处方，患者进入缴费环节后选择各就诊项目，这些业务规则和逻辑都要被考虑进路径规划中，见图 2。实黑箭头表示该路径确保患者能够按照最合理、高效的顺序完成各项就诊项目；虚线箭头路径表示经过 Kruskal 算法放弃的路径。

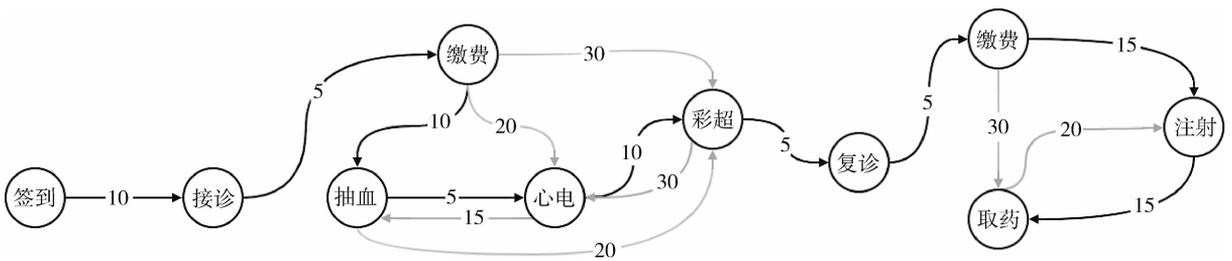


图 2 基于 Kruskal 算法的路径推荐

将为用户提供直观的就医路径指引。

4 就诊路径实时推荐系统设计

就诊路径实时推荐系统的应用入口为移动手机端，患者点击医院微信服务号“医疗服务”栏目下的“就诊路径推荐”，即可轻松访问该服务。整个推荐系统主要由两大部分构成：一是后台部署的核心数据服务部分，为系统提供关键数据支持；二是数据请求响应后，前端展示的路径推荐信息页面，

4.1 服务系统整体架构

就诊路径实时推荐系统主要分为服务请求层、安全限流缓冲层、模型数据融合层、推荐信息输出缓存层，每层独立解耦进行数据处理，整体数据请求响应平均为 800 ~ 1 500 毫秒。服务系统整体架构设计，见图 3。

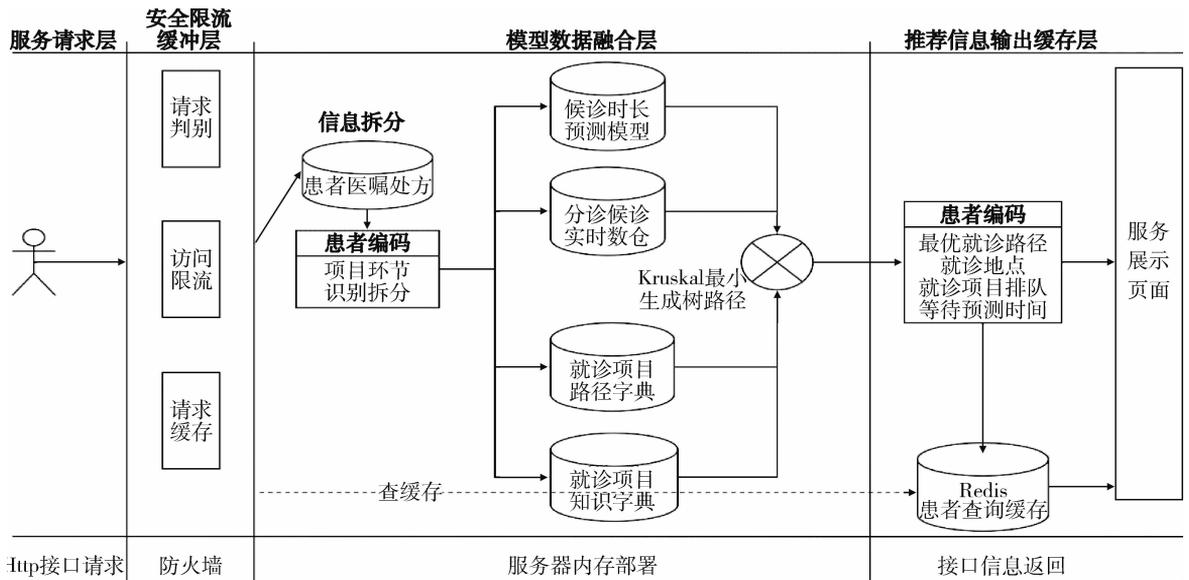


图 3 服务系统整体架构

在数据请求层，患者选择并确认就诊门诊号后，系统将患者门诊号信息通过 Http 接口请求服务发送至安全限流缓冲层。安全限流缓冲层主要用于防止恶意访问、频繁访问。患者门诊号信息到达该层，会调用就诊状态信息，包含预约状态、就诊状态等。异常信息请求会被直接拦截，频繁点击会被限制查询频率，并返回存储在输出缓存的推荐信息。在模型数据融合层，通过患者门诊号可以查询医生签署的医嘱处方内容，对其拆分归类，可以分为缴费、领药、检查检验项目等类别。针对每个类别及具体项目请求对应的服务模型，模型将会获得就诊预估时间 t_{j_c} ，实时候诊接诊取当前候诊人数 n ，可获得每个项目的耗时预计 $t = (n + 1) \times t_{j_c}$ ，整体就医过程包含 m 个项目，则总体时间 $T = \sum_{i=1}^m (n_i$

$+ 1) \times t_{j_c}$ 。结合知识模型中的地点信息，输出具体路径推荐信息。相关模型加载在服务器内存，数据访问获取速度较快。推荐信息输出缓存层主要分为两个部分，对于首次请求或者超过请求时间限制条件的情况，重新计算输出，并将结果数据以患者门诊号为主键存在 Redis 中；对于请求时间间隔较近的情况，将直接返回 Redis 缓存中的数据，Redis 数据可以通过设置淘汰时间进行周期性清理。

4.2 服务系统数据处理逻辑

就诊路径推荐系统依赖于患者的线上预约挂号信息。对于未预约患者，系统不提供签到、候诊就诊等信息的请求服务。当患者完成线上预约后尚未签到时发起访问请求，系统会返回当前已签到且预约相同时段或更前时段的患者总数，并推荐适当的

签到时间以及预测候诊时间。已接诊的患者可查看医生开具的医嘱处方项目。在检查、检验或取药前，患者需缴费。系统推荐使用自助服务终端缴费，缴费时间通常为3~5分钟。若选择人工窗口缴费，可能需要排队，排队时间则因等待人数的不确定而波动较大。缴费成功后，患者的后续项目将开始执行。系统将根据具体项目，获取各诊疗项目

的地点，计算预测候诊时间，匹配项目依赖关系，最终基于Kruskal最小生成树算法生成最优的就诊路径推荐。当患者的待处理项目中没有待缴费项目且医生开具的所有处方项目（包括取药）均已完成时，患者在医院的就医流程结束。系统数据处理逻辑，见图4。

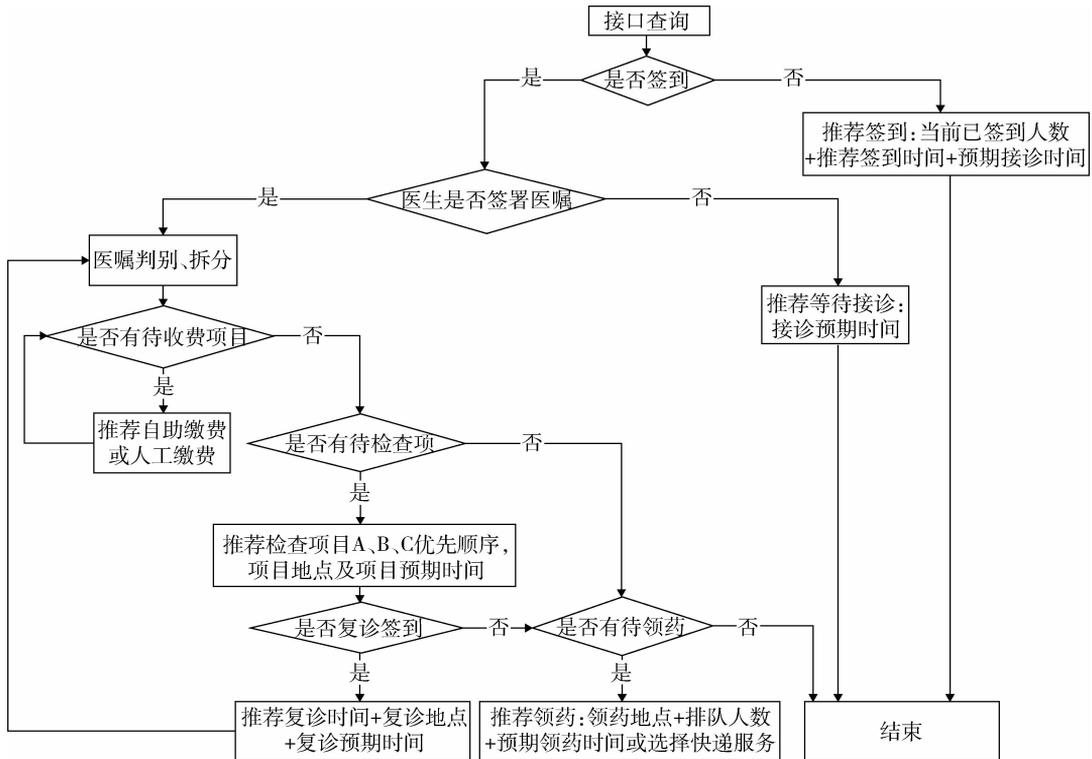


图4 服务系统数据处理逻辑

4.3 智能就诊路径推荐

当患者选择“推荐就诊路径”时，系统通过识别患者门诊号向服务器发起请求，随后在页面清晰展示查询结果，推荐结果原型图，见图5。“更换就诊人”选项便于拥有多张健康卡的患者切换就诊人信息。“就诊路径推荐”详细列出患者的就诊项目、地点、候诊人数及预计等候时间。“医院平面导航图”可以定位就诊地点。整个流程旨在为患者提供便捷、高效的就诊路径推荐服务。



图5 就诊路径推荐原型

5 相关数据的其他应用

一是基于上述系统实时采集的各门诊科室预约、签到及分诊数据,可构建门诊患者实时数据分析应用,通过门诊就诊患者实时看板大屏动态、直观展示各科室的预约情况、签到进度、候诊人数、已完成接诊人数,以及医生的接诊效率,如已完成就诊的患者人数、当前候诊人数和平均就诊时间。不仅能帮助患者了解科室医生的接诊进度,还能提升就医体验。二是基于离线数据,医院管理层可以深入分析并评估各科室和医生的繁忙程度、接诊效率,为科室医生人力资源的合理配置提供决策依据。三是通过长时间序列的数据分析,可以发现病种的时间周期性,如季节性病种的发病趋势,提前预测并合理调配医院资源。四是结合患者的位置信息和定位耗时数据,可以评估医院路标系统的指引效果,确保其高效准确。五是全面、深入地挖掘医院大数据,能够发现医院服务中的瓶颈和问题,为医院管理层提供有力的决策支持。充分展现医院大数据的潜在价值,助力医院实现更高效、更人性化的服务。

6 系统服务实效分析

智能就诊路径实时推荐系统服务于2024年3月中旬灰度上线,目前在试用阶段,服务只面向部分患者。接到请求服务后,系统会在3秒自动弹出服务1—5星评价填写框。服务体验数据统计,见图6。2024年6月服务请求有82人次,参与评价服务人次占14.63%,评价4—5星的占参评人次33.33%。考虑到服务试用阶段,患者请求服务样本不够充分,但是通过服务体验评价数据可知,患者对实时路径推荐服务的需求是客观存在的。未来服务全面上线后,可进一步评估该服务对患者就医时间指标的影响。

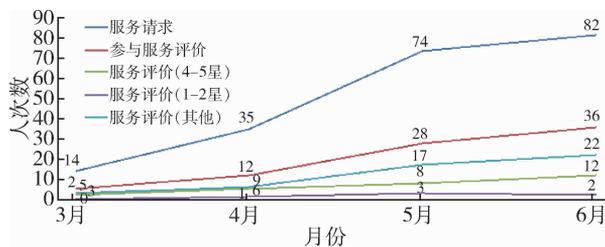


图6 系统服务体验评价统计

7 结语

本研究基于医院大数据设计并实现智能就诊路径实时推荐系统,解决患者对就诊时间缺乏预期、就诊项目组合复杂导致路径规划困难,以及人工导诊服务效率受限等问题,以提升患者就医体验和医院服务效率。该系统建设包含就诊时间预测、就诊路径规划和路径推荐系统设计,患者通过医院微信服务号请求该服务。为进一步优化该系统,未来可从以下方面尝试相关实验:在就诊时间预测方面,扩大实验样本数据,丰富特征工程,如加入患者的病种类型、初复诊状态等信息;在模型结构方面,构建基于深度学习的多元化模型结构;在细分场景方面,增加中药房排队时间预测模型;在服务体验方面,支持患者手动调整系统推荐的就诊路径节点,以更好地适应实际场景。智慧医院建设离不开对医院大数据的深入分析和挖掘,通过充分利用患者在医院产生的数据来提升就医体验和医院服务质量,能够更好地实现大数据的价值。

作者贡献: 蒋宏建负责实验数据处理、模型训练及评估、服务功能开发及测试、论文撰写与修订;陈健超负责需求分析、系统服务架构设计、服务功能开发及测试、论文修订。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 关于开展改善就医感受提升患者体验主题活动的通知[J]. 中华人民共和国国家卫生健康委员会公报, 2023, 5(5): 3-7.

- 2 国家卫生健康委发布医院智慧服务分级评估标准体系 (试行) [J]. 医学信息学杂志, 2019, 40 (4): 93.
- 3 顾懂燕, 刘宁, 董凯琴, 等. 门诊患者就诊满意度影响因素分析及护理对策 [J]. 齐鲁护理杂志, 2023, 29 (22): 127 - 130.
- 4 HUNG G R, WHITEHOUSE S R, O' NEILL C, et al. Computer modeling of patient flow in a pediatric emergency department using discrete event simulation [J]. *Pediatric emergency care*, 2007, 23 (1): 5 - 10.
- 5 SUN Y, TEOW L K, HENG H B, et al. Real - time prediction of waiting time in the emergency department, using quantile regression [J]. *Annals of emergency medicine*, 2012, 60 (3): 299 - 308.
- 6 CURTIS C, LIU C, BOLLERMAN J T, et al. Machine learning for predicting patient wait times and appointment delays [J]. *Journal of the American college of radiology*, 2017, 15 (9): 1310 - 1316.
- 7 HIJRY H, OLAWOYIN R. Predicting patient waiting time in the queue system using deep learning algorithms in the emergency room [J]. *International journal of industrial engineering*, 2021, 3 (1): 33 - 45.
- 8 郭琳琳, 魏永祥, 郭琳瑛, 等. 基于机器学习的儿科急诊患者候诊时间预测及关联分析研究 [J]. 医学信息学杂志, 2022, 43 (4): 33 - 39.
- 9 LI N, LI X, ZHANG C, et al. Integrated optimization of appointment allocation and access prioritization in patient - centred outpatient scheduling [J]. *Computers & industrial engineering*, 2021, 154 (1): 107 - 125.
- 10 李立, 陈坤福. 我院门诊流程优化解决方案分析 [J]. *中国医疗设备*, 2016, 31 (6): 103 - 105.
- 11 王元元, 牟书娟. 门诊就医流程优化实践 [J]. 医学信息学杂志, 2019, 40 (11): 48 - 51.
- 12 FERREIRA D C, VIEIRA I, PEDRO M I, et al. Patient satisfaction with healthcare services and the techniques used for its assessment: a systematic literature review and a bibliometric analysis [J]. *Healthcare*, 2023, 11 (5): 639.
- 13 MAULUD D, ABDULAZEEZ A M. A review on linear regression comprehensive in machine learning [J]. *Journal of applied science and technology trends*, 2020, 1 (2): 140 - 147.
- 14 CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system [EB/OL]. [2023 - 08 - 13]. <https://doi.org/10.1145/2939672.2939785>.
- 15 FAN J, MA X, WU L, et al. Light gradient boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data [EB/OL]. [2023 - 11 - 20]. <https://doi.org/10.1016/j.agwat.2019.105758>.
- 16 PARYATI, SALAHDDINE K. The implementation of Kruskal's algorithm for minimum spanning tree in a graph [EB/OL]. [2023 - 09 - 22]. https://www.e3s-conferences.org/articles/e3sconf/pdf/2021/73/e3sconf_iccsre21_01062.pdf.

(上接第 52 页)

- 12 李倩, 王帅. LDA 模型下我国公共图书馆微信平台阅读推广内容主题研究 [J]. *图书情报工作*, 2022, 66 (8): 72 - 83.
- 13 张世平. 健康养老亟待破解四大难题——在“健康中国 50 人论坛” 2024 年会上的主旨发言 [J]. *中国农村卫生事业管理*, 2024, 44 (5): 306 - 309, 324.
- 14 潘燕. 心理需求视角下居家社区养老服务的优化策略研究 [J]. *大众标准化*, 2021 (16): 218 - 220.
- 15 吴冬梅. 社区养老护理服务存在的问题及对策分析 [J]. *当代护士 (下旬刊)*, 2023, 30 (1): 13 - 16.
- 16 魏智慧, 柴菊童. 社会组织参与社区居家养老服务的问题研究 [J]. *国际公关*, 2021 (4): 60 - 61.
- 17 曾春艳, 付雪连, 刘丽琼, 等. 社区老年人心理健康服务需求现状研究 [J]. *岳阳职业技术学院学报*, 2023, 38 (3): 71 - 75.
- 18 李杨, 姜超. 社区养老服务协同治理思考 [J]. *合作经济与科技*, 2024 (4): 178 - 180.
- 19 顾安琪. 社会组织助力社区养老服务体系建设探索 [J]. *华章*, 2024 (3): 162 - 164.
- 20 祁月浩, 琚泽彬, 党志峰. 智慧社区养老的价值和困境分析及社会工作的嵌入路径 [J]. *经济研究导刊*, 2023 (23): 95 - 97.