

基于大数据 Logistic 回归模型的医院线上诊疗数据精准推荐

李娇

(北京中医医院顺义医院,北京 101300)

摘要:针对当前医院线上诊疗数据精准推荐算法缺少大数据训练过程,导致平台数据承载量较低,数据推送结果异常的问题,设计一种基于大数据 Logistic 回归模型的医院线上诊疗数据精准推荐算法。应用神经网络设定电子病历大数据训练模型,并使用欧氏距离公式,构建诊疗信息关键判别函数,构建 Logistic 回归模型,选择 PageRank 方法优化诊疗资源推荐方案,完成医院线上诊疗数据精准推荐。最后设计实验测试环节,功能测试结果与性能测试结果均表明,此方法具有较高的数据承载量,病历推荐结果可靠性较高,并可以有效降低并发数据运行耗时。

关键词:数据处理与分析;互联网技术;医疗资源整合;线上诊疗;数据承载量

中图分类号:TP311.13 文献标识码:A 文章编号:1003-7241(2025)06-0175-05

Accurate Recommendation of Hospital Online Diagnosis and Treatment Data Based on Big Data Logistic Regression Model

LI Jiao

(Shunyi Hospital, Beijing Traditional Chinese Medicine Hospital, Beijing 101300, China)

Abstract: In response to the lack of big data training process in the current hospital online diagnosis and treatment data precise recommendation algorithm, which leads to low platform data carrying capacity and abnormal data push results, a hospital online diagnosis and treatment data precise recommendation based on big data Logistic regression model is designed. It applies neural networks to set up a big data training model for electronic medical records, and uses the Euclidean distance formula to construct a critical discriminant function for diagnosis and treatment information, constructs a Logistic regression model, selects the PageRank method to optimize the recommendation plan for diagnosis and treatment resources, and completes accurate recommendation of hospital online diagnosis and treatment data. Finally, the experimental testing phase is designed, and both the functional and performance test results show that this method has a high data carrying capacity, high reliability of medical record recommendation results, and can effectively reduce the running time of concurrent data.

Keywords: data processing and analysis; internet technology; integration of medical resources; online diagnosis and treatment; data carrying capacity

0 引言

最新医疗调查显示,当前医院接纳病患数量几何倍数增加,并呈现出逐年上升以及年轻化的趋势。大部分疾病的治愈过程是一个长期复杂的系统工程,在遵循医嘱进行治疗的同时,还需要患者提高自身的配合度,积极参与到治疗计划中。随着医疗资源的短缺程度不断提升,人们对智能化线上诊疗的需求不断提升^[1]。随着人工智能技术的不断发展,相关领域专家学者研发出了一系列以人工智能为基础的医院线上信息化平台,并将其应用到大型医院的日常工作。

例如,徐莉等对单病种质量管理信息化平台建设与应用进行了研究。依托平台可自动采集单病种质量管理

数据并实现网络直报,便于开展单病种质量控制与过程管理^[1];黄冰琳等设计一种基于中心化管理的药物临床试验药房信息化平台。梳理试验用药品管理工作流程,分析临床试验药房信息化需求,建立信息化平台^[2];刘倩倩等构建一种跨院帮扶肿瘤专科护理信息化平台。并通过平台应用前后专科护士管理效果、同质化护理质量、帮扶效果比较验证平台效果^[3]。

尽管目前构建了许多平台,但在以人工智能为基础的诊疗过程中应用中,还存在一定问题,如何结合医疗健康领域的专业知识,设计适合医院线上诊疗数据的推荐算法是一个重要的难点。需要研究各种算法技术,包括基于协同过滤、基于内容过滤和基于深度学习的算法等,开发出能够精准推荐医疗服务的算法。针对当前医院线上诊疗数据精准推荐算法缺少大数据训练过程,导致平台数据承载量较低,数据推送结果异常的问题。

*基金项目:北京市顺义区科技三项费项目(KS201734)

收稿日期:2023-12-14

设计一种基于大数据 Logistic 回归模型的医院线上诊疗数据精准推荐算法,力求在为病患提供更优质的服务的同时,提升医院在人工智能背景下的诊疗能力。

1 医院线上诊疗信息大数据训练模型设计

在此次研究中,线上诊疗信息化平台构建将大数据和人工智能算法作为媒介,为更好地利用大数据和人工智能技术对诊疗信息数据进行整理,需构建相应的大数据训练模型。通过文献研究发现,诊疗信息种类较多且内容较为复杂多为异构数据^[4]。因此,选择构建神经网络的方式,实现信息大数据的训练。

获取到海量病患的电子病历后,提取病历关键词将其记作维度1,而后,在此维度上构建神经网络^[5]。将此网络的隐藏层设定为A,输入层设定为a,输出层设定为b,则有:

$$\begin{cases} A < b - 1 \\ A < \sqrt{a + b} + c \\ A = \lg b \end{cases} \quad (1)$$

式中,c表示需要添加的神经元。根据此公式,将输入层到隐含层的权重设定为E,阈值设定为s,隐含层到输入层的权重设定为F,阈值设定为x。构建正向神经网络,各信息节点与误差计算公式设定如下:

隐含层信息计算式:

$$G_j = \sum_{i=1}^9 (E_{i,j} X_i) + s_j \quad (2)$$

输出层计算式:

$$G_{put} = \sum_{i=1}^9 (F_{i,j} Z_i) + x_j \quad (3)$$

误差计算式:

$$V(u) = \frac{\sum_{i=1}^9 (R_i - Y_i)}{3} \quad (4)$$

式中, Y_i 表示真实数据值; Z_i 表示关联数据值; X_i 表示输入原始数据信息; R_i 表示处理后所得数据值。通过上述处理过程,将原始数据整合为三维形式,并对其进行训练,完成数据训练过程。具体数据训练流程设定如图1所示。

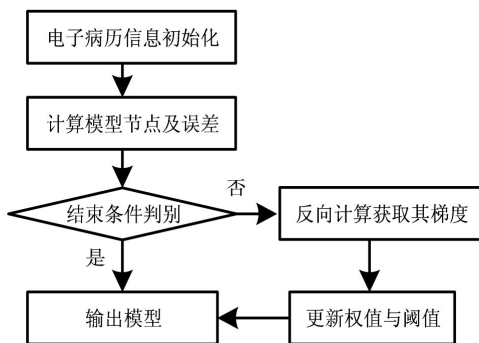


图1 原始信息数据训练流程

按照图1所示内容,对原始海量医疗大数据进行处理,将训练完成的数据存储在平台数据库中,为后续的诊疗工作提供数据基础。

应用此神经网络,确定每个数据的相配节点后,构建诊疗信息判别函数,具体如下:

$$d_i(u) = \sum_{i=1}^D (x_i - Y_{ij})^2 \quad (5)$$

应用此公式,对数据库中的数据进行激活更新,并更新数据权重值:

$$T_{i,j(x)} = \exp\left(\frac{-Y_{ij(x)}}{2\beta^2}\right) d_i(u) \quad (6)$$

式中, β 表示平台预设数据分类权重。按照上述内容,对电子病历进行类别划分,并使用下述公式,对划分结果进行评估:

$$DBI = \frac{\sum_{i=1}^c \max \left\{ \frac{\Delta(Y_i) + \Delta(Y_j)}{\alpha(Y_i, Y_j)} \right\}}{g T_{i,j(x)}} \quad (7)$$

式中, (Y_i, Y_j) 表示不同类别数据之间的距离; g 表示类别数量权重。根据以往研究结果, DBI 取值结果越小,说明数据训练效果越好^[6-7]。将上述设定内容作为此次算法构建的核心数据,为后续算法提供数据关联支撑。

2 医院线上诊疗信息推荐算法设计

2.1 Logistic 回归模型数据关联设计

为保证诊疗数据推荐过程中,算法具有智能联动性,此次研究中将数据库设定为节点式,以便于医疗数据的分析与整理。根据上文中设定的关键技术,使用 MATLAB 2016 软件包^[8-9]将医疗病例数据库整合为可调节节点式数据库,保证各个数据库节点具有连通性。

将经过训练处理后的数据按照类型分别存储在不同功能的数据库,如出现数据调取工作,根据关键词对数据进行抽取与存储。由于医疗数据存在不同的数据库中,为提升抽取的准确性,在数据库中构建 Logistic 回归模型,以此保证医疗数据的内在推荐连通性。本次研究中,将 Logistic 回归模型计算式设定如下:

$$H = \frac{1}{1 + \exp(\alpha_0 + \alpha_0 l_0 + \alpha_1 l_1 + \dots + \alpha_n l_n)} \quad (8)$$

式中, α_0 表示各医疗数据库之间的医疗数据类别距离; l 表示回归方程的自变量; l_n 表示条件改变后的医疗数据类别变化量。

通过此该公式可控制诊疗信息出现变化时,信息抽取结果的正确性。将此公式植入到数据库中,使数据存储模式更换为二进制模式,保证其在算法中正常应用。

2.2 诊疗资源推荐算法设计

本文选择PageRank方法^[10]对当前的诊疗资源推荐展开处理。在海量医疗大数据的训练过程中,如某诊疗信息被很多信息链接的话,说明此信息较为重要,且PageRank值相对较高。根据此假设,完成推荐,过程如下:

设定每个诊疗资源都存在相应的访问概率,且每个诊疗资源都由 H_1 、 H_2 、 H_3 、 H_4 四部分关联程度组成,通过上文公式(8)计算产生,如全部的资源均与 H_1 有链接,则 H_1 的PageRank值为剩余3部分的总和,则有:

$$PR(H_1) = PR(H_2) + PR(H_3) + PR(H_4) \quad (9)$$

一般情况下,每个资源不会仅存在一个链接,假设 T_1 存在两个或两个以上链接项,则有:

$$PR(H_1) = \frac{PR(H_2)}{n'} + \frac{PR(H_3)}{n'} + \frac{PR(H_4)}{n'} \quad (10)$$

式中, n' 表示链接项数量。整理上述公式,得到最终的诊疗资源推荐公式:

$$PR(H_1) = \frac{PR(H_2)}{L(H_2)} + \frac{PR(H_3)}{L(H_3)} + \frac{PR(H_4)}{L(H_4)} \quad (11)$$

式中, $L(H_i)$ 表示资源链接节点及数据库数量。应用此公式控制诊疗资源的推荐过程。对上述内容进行整合,完成诊疗资源推荐模块构建,并将文中设定的模块安装到原有平台框架中,同时融合信息化关键技术,完成基于互联网的医院线上诊疗信息化平台构建工作。

3 算法的测试实验分析

本文算法从构建到设计,均需要大量的测试证实此平台的应用性能,并在测试的过程中不断解决各类性能问题,最终得到满足日常需求的分级诊疗平台。本算法导入医疗测试平台,通过多轮测试确定各个功能模块的应用性能,并交付使用。

3.1 功能测试

本次功能模块测试采用等价类划分法,将信息化平台的应用环境罗列出来,整合场景信息,构建测试环境。对医院环境进行分析后,确定性能测试与兼容性测试两个环节对平台的基本功能进行探析,具体测试用例如表1所示。

分析上述内容后可以看出,此次研究中构建的信息化平台基础功能性较好,在应用的过程中具有较高的稳定性,可为医院的诊疗工作提供帮助。与此同时,在上述测试结果的基础上,可对平台的使用性能展开测定分析,确定其应用价值。

3.2 功能测试结果

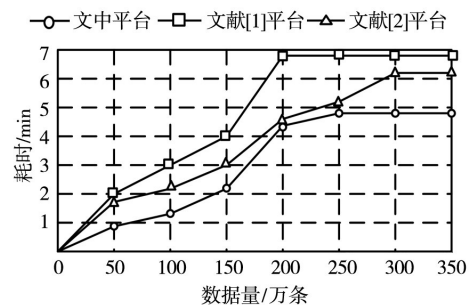
3.2.1 平台数据承载量测试

大量的案例指出,线上诊疗信息化平台中还有海量的病患信息与诊疗数据,如平台的信息承载量较差,将导

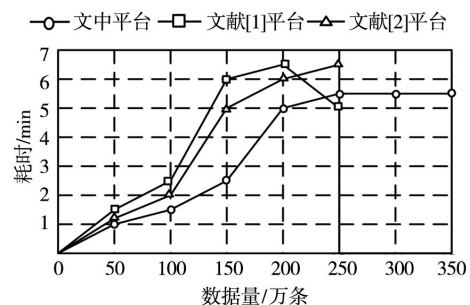
致严重的问题。因此,此次功能测试中,将首先对平台的数据承载能力进行分析。为保证测试结果的真实性,将测试环境设定为静止状态与平台信息交互状态两组环境。与此同时,将初始信息量设定为50万条,每测试一次增加50万条,直至平台停止运转。此次测试中,选取文献[1]与文献[2]平台和本文设计平台展开同期测试,具体测试结果如图2所示。

表1 测试用例

测试项目	测试内容	预期结果	实际结果
权限日志性能测试	使用注册成功用户名,查看目标病例	获取目标病例信息,其他功能不可见	符合预期设定
	使用注册成功用户名,申请转诊	确定目标病例后,按照需求,拟发送转诊信息	符合预期设定
电子病历测试	用户可操作电力病历相关功能,录入原始病例信息,并按要求进行存储	在平台数据库中可查询到目标病例	病患信息与原始信息一致
转诊测试	申请转诊,完成转诊审批	填写转诊信息,发起转诊申请	接收到转诊申请,并同意
	转诊申请失败后,填写驳回理由	驳回理由保存成功	信息保存成功,刷新后存在
	转诊闭环处理	可获取到重新上传的电子病历	信息保存成功,刷新后存在
医院资源查询	输入目标设备或检测项目名称,获取相关资源信息	输入目标词汇后,可得到相应的信息	获取到目标项目信息,并及时反馈



(a) 静止状态



(b) 信息交互状态

图2 平台数据承载量测试结果

通过上述测试结果可以发现,当平台处于静止状态时,本文平台与其他两平台的数据承载量较为一致,但总体偏高。当测试状态为信息交互状态,三种平台的数据承载量差异逐渐明显。文中平台的数据承载量与其在静

止状态较为一致,其他两种平台的数据承载量明显下降,说明此两种信息化平台的运行稳定性较差,无法对大量的数据进行处理与分析。

3.2.2 平台信息推荐能力测试

此次研究中,将平台需要推荐的信息数据设定为图像与数据两部分。将表2中数据存储的平台中,并按照预设的查询要求,输出目标数据,以此判定平台信息处理能力。具体测试数据见下表。

表2 诊疗信息化平台测试数据

测试组序号	信息数据总数/条	数据信息/条	图像信息/条
1	4 178	1 524	2 654
2	6 795	3 154	3 641
3	4 788	3 146	1 642
4	7 655	1 204	6 451
5	8 570	6 415	2 155
总计	31 986	15 443	16 543

选取具有特征的诊疗图像作为测试对象,并将其划分为3类,使用文中平台与其他两种平台调取此部分图像,分析推荐结果,确定不同平台的信息调取效果。

按照上述设定,获取测试结果,具体如表3所示。



(a) 文中平台



(b) 文献[1]平台



(c) 文献[2]平台

图4 图像调取结果

对以上测试结果进行分析可以看出,在进行图像调取的过程中,文中平台可获取到目标图像且不会出现异常。其他两种平台虽然也可以调度到相应的图像,但此部分图像存在异常,由此证实了此两种平台的信息分析处理能力不佳。在进行上述测试的过程中,对信息数据的并发时间进行统计,所得结果如表4所示。

表4 信息数据并发耗时统计/ms

事务名称	文中平台耗时	文献[1]平台耗时	文献[2]平台耗时
电子病历数据录入	1.21	2.61	1.64
电子病历数据查询	1.45	2.35	1.70
电子病历图像导入	2.15	4.68	4.85
电子病历图像查询	2.61	5.10	5.00

当电子病历数据并发量一致时,文中平台可在最短时间内完成电子病历的整理与查询工作。与文中系统相比,其他两种平台的数据并发处理时间相对较短,更进一步证实了本文平台具有较高的信息化处理能力,将其应

表3 信息化平台数据推荐结果

测试组序号	文中平台查询结果/条	文献[1]平台查询结果/条	文献[2]平台查询结果/条
1	4 178	4 100	4 178
2	6 795	6 154	6 795
3	4 788	4 031	4 788
4	7 651	7 044	7 640
5	8 568	8 147	8 145

对上表中数据进行分析可以看出,文中平台应用后可对99%的数据进行推荐,且数据较为完整。另外两种平台查询结果存在大量缺失,且数据完整度不高。此测试结果证实了基础平台与人工智能平台的数据处理能力较差。根据测试要求,对其进行细化得到下述医疗图像推荐结果,如图3。图像调取结果如图4所示。



图3 图像推荐目标

图3 图像推荐目标

用在在线上诊疗中,用户可通过此平台与医生沟通,并及时从数据库提取相关病历信息。

4 结束语

针对当前医院线上诊疗信息化平台在应用过程中存在的问题,研究了一种以大数据和人工智能算法作为媒介的诊疗信息化推荐算法。此算法在研发过程中应用了多种互联网数据处理技术,并对原有的平台框架进行了优化,以期提高此平台的应用效果。在后续的研究过程中,还需不断对此平台的硬件部分进行完善与优化,保证平台软件与硬件具有较高的适配性。

参考文献:

- [1] 徐莉,顾逸君,邵琰婷,等.单病种质量管理信息化平台建设与应用[J].中国卫生质量管理,2022,29(7):15-18.
- [2] 黄冰琳,方令平,郭锦华,等.基于中心化管理的药物临床试验药房信息化平台设计与实现[J].中国现代应用药学,2022,39(21):

2781-2786.

[3] 刘倩倩,张素兰,田丽,等.跨院帮扶肿瘤专科护理信息化平台的构建及应用[J].中华护理杂志,2021,56(6):805-810.

[4] 魏盈盈,刘东,尹雄章,等.药品集中带量采购工作标准化流程与信息化平台建设实践[J].中国药房,2022(9):1136-1140.

[5] 孙秋,张铁林,聂双莲,等.我院医疗质量管理信息化平台建设的实践探讨[J].中国医院管理,2022,42(8):67-69.

[6] 宋勤,包芳芳,金妙景,等.以麻醉药房为载体的预防用抗菌药物信息化管理平台的构建和应用[J].中华医院感染学杂志,2021,31(24):3822-3826.

[7] 姜倩,李秦川,郭茜茜.基于德尔菲法构建“互联网+双向+游戏+科普”模式的药学科普交互平台[J].中国医院药学杂志,2021,41(19):2015-2020.

[8] 阳国兴,彭文英,钱邦富,等.医院运营信息化平台建设体系的构建在门诊管理中的应用[J].中国卫生事业管理,2021,38(7):503-506.

[9] 王慧,姜若,赵虎,等.基于信息化平台的公立医院党委会议事规则运行实践[J].中华医院管理杂志,2021,37(9):781-783.

[10] 孙晓.综合性医院数字化档案信息自动分类方法研究[J].自动化技术与应用,2023,42(7):74-77.

[11] 吴培,董晨东,茅宁莹.“互联网+”背景下处方外流推进障碍与对策分析—基于利益相关者视角[J].中国药房,2021,32(12):1409-1415.

[12] 高艺萌.医院财务信息智能化管理平台系统设计[J].自动化技术与应用,2023,42(2):162-166.

作者简介:李娇(1988—),女,硕士,工程师,研究方向:医院信息化管理。

(上接第98页)

可以看出,在此次研究的方法下,CNN-LSTM的运行时间明显地减少,虽然运行时间相较于CNN有略微增加,但是在高准确度的保障下,研究认为这部分增加的时间是可以接受的。

3 结束语

近年来,电气化铁路在我国得到了全面的发展。作为其主要组成部分,牵引供电系统的故障识别对于电气化铁路安全具有重要作用。目前牵引供电系统故障识别还少有采用深度学习方法的研究出现。为此,研究针对复线自耦变压器的供电故障,提出了一种结合了CNN和LSTM的故障识别模型。实验结果中,CNN-LSTM故障识别的平均准确率平均为0.95。相较于CNN和LSTM模型,分别提高了0.18和0.14,验证了模型的准确性。此外,在30组数据的测试中,CNN-LSTM识别准确率发生明显偏差的情况明显少于其他两种模型,验证了模型的稳定性。最后,在数据量为100时,相较于LSTM,CNN-LSTM的运行时间减少了46.83%,验证了研究方法的有效性。虽然此次研究取得了一定的成果,但是由于设备条件有限,研究并未进行大样本的实验模型性能验证。因此此次研究仅作参考,进一步提升此次研究的可靠性将成为今后的主要课题。

参考文献:

[1] 程鹏,刘文泉,陈冲,等.面向电气化铁路牵引供电的光伏发电分相电流控制策略[J].电力系统自动化,2022,46(19):145-153.

[2] 施力仁.基于光纤传感器的电气设备状态检测方法[J].微型电脑应用,2022,38(3):148-150.

[3] Zhou H,Duan F,Liu Z,et al.Study on electric spark discharge between pantograph and catenary in electrified

railway[J].IET electrical systems in transportation,2022,12(2):128-142.

[4] 翁植,文成林.基于深度特征聚类和RNN的电网故障诊断[J].控制工程,2022,29(1):175-181.

[5] 郭瑜,童丽娜,倪旭明.基于改进卷积神经网络的电力通信网故障诊断研究[J].计算机测量与控制,2022,30(2):24-30.

[6] Chu W,Song Y,Duan F,et al.Development of steady arm damper for electrified railway overhead contact line with double pantographs based on numerical and experimental analysis[J].IET Electrical Systems in Transportation,2021,11(3):269-277.

[7] 赵斌,贾智,王东,等.电气化铁路轨道电路钢轨互阻抗计算研究[J].铁道学报,2021,43(8):54-61.

[8] 李群湛,王帅,易东,等.电气化铁路贯通同相供电AT牵引网故障辨识与自愈技术研究[J].铁道学报,2022,44(7):46-54.

[9] 郭爱军,鞠晨.基于深度信念网络的模拟电路故障诊断[J].自动化技术与应用,2023,42(2):76-80.

[10] 郑乐辉,孙君杰,牛润,等.改进CNN和Bi-LSTM的集成化装备故障检测研究[J].计算机测量与控制,2022,30(11):52-58.

[11] Chen H,Zhang X.Path Planning for Intelligent Vehicle Collision Avoidance of Dynamic Pedestrian Using Attention-LSTM,MSFM,and MPC at Unsignalized Crosswalk[J].IEEE Transactions on Industrial Electronics,2022,69(4):4285-4295.

[12] He J,Li C,Wang P,et al.Soliton Molecule Dynamics Evolution Prediction Based on LSTM Neural Networks[J].IEEE Photonics Technology Letters,2022,34(3):193-196.

作者简介:翟政中(1990—),男,工程师,研究方向:铁道电气化。

通信作者:刘华云(1984—),男,助理工程师,研究方向:电气工程。