

大数据驱动的老年人护理需求预测模型开发^{*}

杨娟 宁珂雪 蒲华秀

龚海燕

(四川广播电视大学 成都 610000)

(四川大学锦城学院 成都 610000)

〔摘要〕 采取大数据驱动的机器学习技术开发老年人护理需求预测模型, 阐述研究目标与内容、关键技术与路线, 分析该技术产品化后的应用前景, 指出该技术应用有助于预测未来老年人护理需求态势和智能生成个性化护理方案。

〔关键词〕 大数据; 机器学习; 老年人; 护理需求; 预测模型

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.04.007

Development of Big Data – driven Prediction Model for the Elderly Nursing Needs YANG Juan, NING Kexue, PU Huaxiu, Sichuan Radio and TV University, Chengdu 610000, China; GONG Haiyan, Jincheng College of Sichuan University, Chengdu 610000, China

〔Abstract〕 A prediction model for the elderly nursing needs is developed by adopting big data – driven machine learning technology. The paper expounds on the study goals and contents, key technologies and routes, analyzes the application prospect of the technology after its productization, points out that the application of this technology is conducive to predicting the future trend of the elderly nursing needs and intelligently generating personalized nursing plans.

〔Keywords〕 big data; machine learning; the elderly; nursing needs; prediction model

1 引言

人口老龄化现已成为我国极为严峻的社会问题。截至 2018 年我国 60 岁及以上人口已达 2.49 亿, 失能老人超过 4 000 万。四川省是人口大省, 60 岁及以上老年人数超过 2 000 万, 失能老年人超过 300 万; 另一方面, 作为劳务输出大省, 四川空

巢老人较多, 对于专业护理服务呈现庞大而刚性的需求。传统基于问卷调查收集老年人护理需求的方式存在信息滞后、信息散乱、护理项目不合理、服务价格随意性大、护理效果无法跟踪等诸多问题, 亟需利用新兴技术全面分析老年人健康状态、优化护理方案、预测护理需求。本文通过人工智能与机器学习等技术分析四川省基层医疗机构信息系统中 2014 – 2019 年老年人海量健康数据, 构建护理需求预测模型。

〔收稿日期〕 2020 – 03 – 11

〔作者简介〕 杨娟, 工程师, 讲师, 发表论文 6 篇; 通讯作者: 宁珂雪, 硕士, 讲师。

〔基金项目〕 四川广播电视大学“基于微门户平台的招生管理系统”(项目编号: KTZXJY2018001Q)。

2 研究目标与内容

2.1 目标

通过分析四川省基层医疗机构信息管理系统中

海量公共卫生服务数据,全面掌握全省 55 岁以上人群健康状态与护理需求时空分布,利用机器学习和人工智能技术揭示健康状态的发展态势,构建四川省老年人护理需求预测模型,为整体描述全省未来老年人护理需求数量与类别分布、个性化护理服务定制、引导社会要素流入护理供给侧和构建老年护理市场化机制提供参考。

2.2 内容

2.2.1 刻画 2014 - 2019 年 55 岁以上人群“三间分布” 根据全省 21 个地市州 55 岁以上人群的生活自理能力、认知功能、抑郁量化以及随访体检数据对其健康状态指标进行描述性分析,刻画老年人健康状态的时空分布。

2.2.2 分析健康指标变化,构建具有护理需求的人群画像 研究 2014 - 2019 年之间 55 岁以上人群公共卫生健康数据,基于机器学习算法设计开发全省老年人健康指标演进模型,根据模型系统构建未来需要护理的老年人群画像。

2.2.3 预测未来护理需求分布 结合老龄化进程、主要健康指标变化以及基于专家知识的护理方法,利用人工智能和机器学习技术构建护理需求模型,从整体水平上预测未来 1 年四川省老年护理市场总需求和个性化需求,有效促进老年人护理服务市场化发展水平。

3 关键技术

3.1 概述

首先进行数据清洗和多元数据融合,其次描述健康状态指标分布特征,再次分析四川省健康状态主要指标的变化趋势,最后通过机器学习预测未来

1 年全省护理需求态势。所涉及的关键技术包括:多源数据融合、健康指标分布统计学研究、基于机器学习构建预测模型、效果评价指标。

3.2 多源数据融合反馈技术

护理需求要素的日常变化以及由各种疾病引起的病源性护理需求等与养老护理运营及其效率密切相关,且随着经济发展,行业管理精细化水平逐步提升,对按需护理服务精度的要求也越来越高,需要多源数据的监测和汇总。多源数据融合,见图 1。现阶段护理需求要素收集和评价手段都是单一化的,如何联合这些单一化的手段得到一项综合性、具有市场反馈机制的多源数据融合采集技术,是提供按需护理的重要手段。

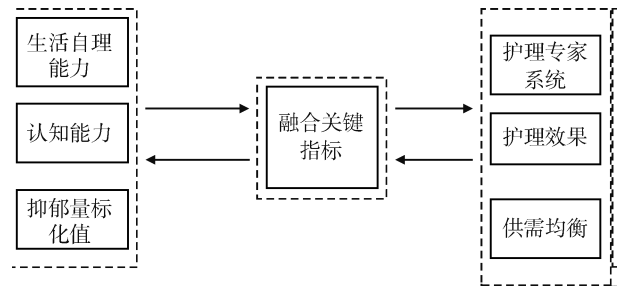


图 1 多源数据融合

3.3 健康指标分布统计学研究

在实际中由于健康指标的复杂性,研究对象往往表现为多指标数据。多指标数据的结构更加复杂,严格上应用三维表来表示,在平面上可以将其转换为一个二级二维表的形式,见表 1。研究总体共有 N 个,每个样本的特征用 p 个指标表示 ($X_1, X_2, \dots, X_j, \dots, X_p$),时间长度为 T ,则 $X_{ij}(t, N)$ 表示第 i 个样本第 j 个指标在 t 时间。

表 1 健康指标分布统计学研究

时间	1			...	t			...	T		
样本	$X_1 \dots$	$X_j \dots$	X_p	...	$X_1 \dots$	$X_j \dots$	X_p	...	$X_1 \dots$	$X_j \dots$	X_p
1	$X_{11}(1) \dots$	$X_{1j}(1) \dots$	$X_{1p}(1)$	...	$X_{11}(t) \dots$	$X_{1j}(t) \dots$	$X_{1p}(t)$	...	$X_{11}(T) \dots$	$X_{1j}(T) \dots$	$X_{1p}(T)$
2	$X_{21}(1) \dots$	$X_{2j}(1) \dots$	$X_{2p}(1)$	...	$X_{21}(t) \dots$	$X_{2j}(t) \dots$	$X_{2p}(t)$	...	$X_{21}(T) \dots$	$X_{2j}(T) \dots$	$X_{2p}(T)$
...	...	...	...	...	...	...	...	...	...	...	...
i	$X_{i1}(1) \dots$	$X_{ij}(1) \dots$	$X_{ip}(1)$	...	$X_{i1}(t) \dots$	$X_{ij}(t) \dots$	$X_{ip}(t)$	...	$X_{i1}(T) \dots$	$X_{ij}(T) \dots$	$X_{ip}(T)$
...	...	...	...	...	...	...	...	...	...	...	...
N	$X_{N1}(1) \dots$	$X_{Nj}(1) \dots$	$X_{Np}(1)$	...	$X_{N1}(t) \dots$	$X_{Nj}(t) \dots$	$X_{Np}(t)$	...	$X_{N1}(T) \dots$	$X_{Nj}(T) \dots$	$X_{Np}(T)$

3.4 基于机器学习构建预测模型

3.4.1 概述 随着机器学习广泛应用, 受益于各种算法, 研究如何高效地将训练好的机器学习模型部署到生产环境。从数据采集开始, 经历数据分析、变形、验证、拆分、训练、模型创建、模型验证、大规模训练、模型发布、提供服务、监控和日志, 再到构建完整机器学习预测模型, 见图 2。根据健康指标特征 (长期趋势、季节性和自相关性), 采用机器学习算法预测未来健康指标的演进趋势。拟采用的机器学习方法主要有长短期记忆网络、误差反传算法神经网络和最小二乘支持向量机, 对其预测效果进行评价。

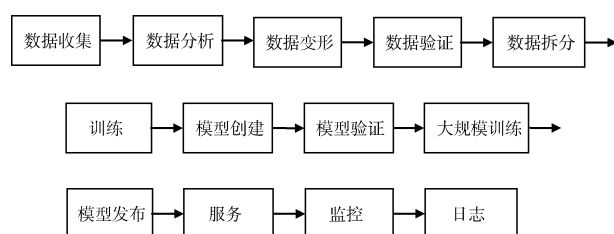


图2 机器学习预测模型

3.4.2 长短期记忆网络 (Long - Short Term Memory, LSTM) 循环神经网络 (Recurrent Neural Network, RNN) 的一种。神经网络是一种节点定向连接成环的人工神经网络 (Artificial Neural Networks, ANNs)。循环神经网络内部状态可以展示动态时序行为, 与其他网络的不同之处在于其隐含层是能够跨越时间点的自连接层, 隐含层的输出不仅进入输出端, 还进入下一个时间步骤的隐含层, 所以能够持续保留信息, 根据之前状态推出后面的状态^[1]。RNN 每个时间状态的网络拓扑结构相同, 在任意 t 时间下包含输入层、隐含层、输出层。RNN 隐含层的输出一分为二, 一份传给输出层, 一份与下一时刻输入层的输出一起作为隐含层的输入。

3.4.3 误差反传算法 (Back - Propagation, BP) 神经网络 ANNs 作为智能信息处理的工具之一, 是模仿人的大脑神经系统信息处理功能的智能化系统, 具有并行处理、容错性、自学习功能, 已在模拟识别、自动化控制、知识处理及运输与通行等领域取得很好的效果。BP 神经网络是 ANNs 中最

具代表性和应用最广泛的一种。通常由具有多个节点的输入层、隐含层和多个或一个输出节点的输出层组成^[2], 见图 3。

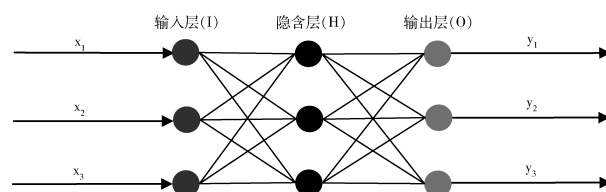


图3 BP神经网络原理

3.4.4 最小二乘支持向量机 (Least Squares Support Vector Machine, LSSVM) 是一种简单的支持向量机 (Support Vector Machine, SVM)。在支持向量机的基础上, 通过最小二乘法, 利用误差平方和选择超平面, 构造平方损失函数最小二乘支持向量机, 同时将 SVM 的不等式约束条件转换为线性等式, 将二次规划问题转化为线性求解, 其求解速度比 SVM 更快、效率更高^[3]。四川省老年人护理需求特征的 LSSVM 预测模型基于该省基层医疗机构信息化系统 (以下简称系统) 采集的数据建立, 同时还要考虑到系统所采集老年人群的变化, 因此选取的输入因子包括: 前 i ($i = 1, 2, 3, 4, 5$) 年的 5 年数据。根据四川省 5 年 (2014 - 2019 年) 的资料作预测研究, 得出仿真结果, 再将试预报值与实测值进行对比分析。LSSVM 建模的关键在于参数 γ 和核函数参数 δ 的确定。研究表明由于径向基核函数 (Radial Basis Function, RBF) 对数值限值条件相对较少, 模型复杂度降低, 从而大大提高训练速度。采用 RBF 作为模型的核函数, 其算法实现过程如下: (1) 确定参数 γ 和 δ 的取值范围。(2) 在参数最大取值范围内构建参数对 $(\gamma i, \delta i)$ 的二维网格平面。(3) 输入每个网格节点参数对 $(\gamma i, \delta i)$ 到 LSSVM 中, 采用学习样本进行学习, 并输出学习误差。取最小误差对应的节点值 $(\gamma i, \delta i)$ 为最优参数对。(4) 以 $(\gamma i, \delta i)$ E min 为中心, 构建新的二维网格平面, 选取数值相近的参数值进行学习, 可获得更高精度的学习结果。

3.4.5 预测评价指标 依次为: 平均绝对误差 (Mean Absolute Error, MAE), 平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE), 均方根误

差 (Root Mean Square Error, RMSE) 和 R2, 其计算公式为:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - y_i|$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(|Y_i - y_i| / |Y_i| \right)$$

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^n (Y_i - y_i)^2 \right)^{\frac{1}{2}}$$

$$R^2 = 1 - \frac{\sum (Y_i - y_i)^2}{\sum (Y_i - Y_{mean})^2}$$

其中 Y_i 为 t 年度护理需求的老年人群, y_i 为机器学习算法预测的 t 年达到护理需求的老年人群。 Y_{mean} 为需要护理的老年人的年均值线。

4 技术路线 (图4)

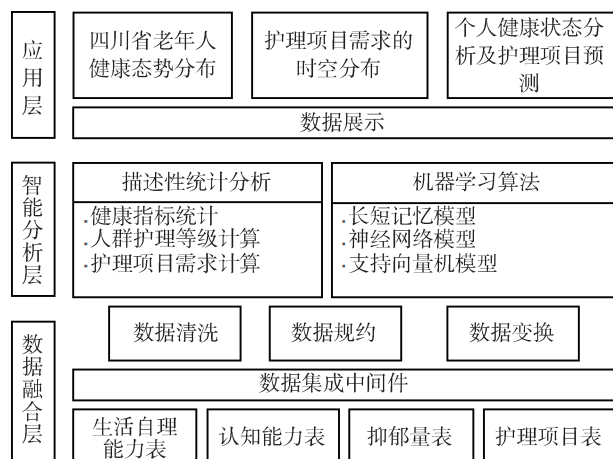


图4 技术路线

5 结语

本研究计划在全省范围内系统、全面分析 55 岁以上人群的健康状态, 从数据层面揭示整体健康状态主要指标的变化轨迹, 为政府相关部门制定养老助老政策、构建护理服务市场化机制提供参考依据。尝试将大数据、机器学习等新兴技术应用到老年人健康状态与护理需求预测, 弥补传统问卷调查、样本统计、专家知识等研究方式的不足, 进一步提升护理需求评估全面性、需求信息发布集中性、护理行为规范性。该技术经产品化后, 如果在全省范围内实施, 既是对四川省基层医疗机构信息化管理系统功能的有力扩展, 也能从全省整体水平上预测未来全省老年人护理需求态势和智能生成个性化护理方案, 帮助护理需求侧与供给侧在一定约束条件下交易成本更低、护理效果更好。根据未来发展态势, 一方面能够指导老年人加强自身锻炼, 避免或减少未来高费用的护理项目, 控制晚年生活成本, 提高生活质量; 另一方面引导社会资源根据护理项目的发展态势提前布局投入, 为未来护理需求储备技术和人力。

参考文献

- 1 杨茂, 朱亮. 基于 FA-PCA-LSTM 的光伏发电短期功率预测 [J]. 昆明理工大学学报 (自然科学版), 2019, 44 (1): 61-68.
- 2 张莹. 我国典型城市空气污染特征及其健康影响和预报研究 [D]. 兰州: 兰州大学, 2016.
- 3 刘其琛, 穆炜炜. 最小二乘支持向量机的核函数及参数选择算法研究 [J]. 电脑知识与技术, 2015, 11 (19): 160-162.

《医学信息学杂志》开通微信公众号

《医学信息学杂志》微信公众号现已开通, 作者可通过该平台查阅稿件状态; 读者可浏览当期最新内容、过刊等; 同时提供国内外最新医学信息研究动态、发展前沿等, 搭建编者、作者、读者之间沟通、交流的平台。可在微信添加中找到公众号, 输入“医学信息学杂志”进行确认, 也可扫描右侧二维码添加, 敬请关注!



《医学信息学杂志》编辑部