

# 基于 PCA-LSTM 神经网络的辽宁个人卫生支出占比变化趋势预测\*

辽宁中医药大学经济管理学院(110847) 梁旭 赵悦 马月丹 王锐 屈嘉怡

**【摘要】** 目的 预测辽宁个人卫生支出占卫生总费用比重的变化趋势,为辽宁未来卫生健康、医疗保障相关政策制定提供参考依据。方法 利用主成分分析法将影响个人卫生支出占比的多个因素进行降维处理,并将主成分得分输入长短时记忆神经网络(long short term memory network, LSTM)神经网络模型进行变化趋势预测。结果 前 3 个主成分特征值均大于 1 且累积方差贡献率达 84.592%,主成分分析(principal components analysis, PCA)-LSTM 模型的预测结果显示未来 9 年辽宁个人卫生支出占比会呈缓慢下降趋势。2025 年、2030 年的预测值均略高于规划值。结论 辽宁个人卫生支出占比缓步下降,实现规划目标任务有一定难度,应合理控制个人卫生支出上涨,加大政府卫生支出力度,完善多元筹资机制。

**【关键词】** 卫生总费用 个人卫生支出占比 主成分分析 长短时记忆神经网络

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.06.024

卫生总费用是全面反映一个国家或地区在一定时期内全社会用于医疗卫生服务所消耗的资金总额<sup>[1]</sup>。从筹资角度看,卫生总费用由政府、社会和个人卫生支出三部分组成<sup>[2]</sup>。其中个人卫生支出是指由个人或家庭直接支付的医疗卫生服务费用。个人卫生支出占卫生总费用的比重是衡量一个国家或地区卫生筹资水平与居民就医负担的重要指标<sup>[3]</sup>,也是《“健康中国 2030”规划纲要》《“十四五”国民健康规划》《“十四五”全民医疗保障规划》等规划文件中的主要约束性指标之一。

党和国家高度重视减轻居民就医负担问题,《“健康中国 2030”规划纲要》明确了“到 2025 年个人卫生支出占卫生总费用比重降至 27%,到 2030 年降至 25%左右”的目标。《辽宁省“十四五”卫生与健康发展规划》中提出该指标规划目标为 27%左右,与全国规划水平保持一致。但从近十年的数据看,2010 年、2015 年、2020 年辽宁个人卫生支出占卫生总费用比重分别为 39.4%、36.0%、30.8%,均高于全国相应年份 35.3%、29.2%、27.7%。从省际比较看,2019 年除河北、湖北外辽宁个人卫生支出占比高于其他省份。辽宁个人卫生支出占比虽然与全国同样呈下降趋势但依然偏高,辽宁在 2025、2030 年完成该指标的规划任务具有一定的难度。如何顺利实现该规划目标是辽宁卫生健康事业发展面临的艰巨任务之一,而探索该指标的影响因素并对其进行科学预测不仅对于辽宁具有重要的现实意义,对于面临同样问题的其他省份也具有一定的借鉴价值。本文通过主成分分析方法将影响个人卫生支出占比的诸多因素进行降维处理,并建立长短时记忆神经网络(long short term memory network,

LSTM)神经网络模型预测其变化趋势,为辽宁未来卫生健康、医疗保障相关政策制定提供参考依据。

## 资料与方法

### 1. 研究数据

#### (1) 指标选取

通过文献回顾、小组讨论与专家咨询,按照数据可获得性、强相关性、可比性原则,建立省级个人卫生支出占比影响因素指标体系,如表 1 所示指标体系由内部相关因素、居民健康状况、医疗卫生资源、社会经济因素 4 个一级指标和 16 个二级指标组成。

表 1 省级个人卫生支出占比影响因素指标体系

一级指标	二级指标
内部相关因素	1. 政府卫生支出占财政支出比重(%)
	2. 人均社会卫生支出(元)
	3. 卫生总费用占 GDP 比重(%)
	4. 人均卫生总费用(元)
居民健康状况	5. 65 岁以上人口占比(%)
	6. 人口出生率(‰)
	7. 人口死亡率(‰)
	8. 传染病发病率(1/10 万)
医疗卫生资源	9. 每千常住人口执业(助理)医师数(人)
	10. 每千人口医疗卫生机构床位数(张)
	11. 医疗卫生机构数(个)
	12. 卫生技术人员数(万人)
社会经济因素	13. 人均 GDP(元)
	14. 城镇化率(%)
	15. 居民消费价格指数(%)
	16. 失业率(%)

#### (2) 数据来源

数据来自于《中国统计年鉴》《中国卫生健康统计年鉴》和《辽宁省统计年鉴》,数据公开可查,真实可靠。

\* 基金项目:国家卫生健康委卫生发展研究中心委托科研项目(2022000000430)

(3) 缺失值处理

对于部分指标数据获取不全的情况,采取以下两种方法进行缺失值填补:①对于有明显变化趋势倾向的指标数据缺失,根据前两年增幅对缺失年份数据进行推算并填补缺失值;②对于无明显变化趋势情况且缺失值小于全部数据 10%的指标数据缺失,采用均值法进行缺失值填补。

2. 研究方法

(1) 主成分分析

主成分分析法 (principal components analysis, PCA) 是一种降维方法,即通过正交变换将多个相关性强的变量转化为若干个线性不相关的变量。为剔除原有指标中冗余的成分首先利用 SPSS 对原始数据进行标准化处理,再计算样本相关系数矩阵及其特征值和对应的特征向量,进而得出方差贡献率及累计方差贡献率,最后在确保主成分的累计贡献率大于 80%的前提下,确定主成分的数量。

(2) LSTM 模型

LSTM 模型是一种改进的循环神经网络,可以用来解决循环神经网络无法处理的长距离依赖问题<sup>[4]</sup>。LSTM 模型在网络状态内添加了记忆细胞状态,并通过增加遗忘门、输入门、输出门使得某一时刻的输出受当前时刻的输入和上一时刻的输出共同影响<sup>[5]</sup>。LSTM 通过遗忘门结构选择性地控制信息的流动,它决定了之前时刻的历史信息有多少保留到当前时刻,然后将当前的记忆和长期的记忆组合在一起,形成新的细胞状态,通过遗忘的信息可以减少时间维度的数据冗余。LSTM 模型擅长处理具有时间先后顺序的数据,因此适用于时间序列预测<sup>[6]</sup>。

结 果

1. 主成分分析结果

对 1995—2021 年 16 个指标的数据进行统计分析,KMO 及巴特利特检验结果显示 KMO 值为 0.972, KMO 值>0.5, 巴特利特球形度检验  $P<0.001$ 。通过 SPSS 26.0 进行主成分分析得到主成分的特征值、方差贡献率和累积贡献率(见表 2),结果表明前 3 个主成分特征值均大于 1 且累积方差贡献率达 84.592%,大于 80%,因此,原始的 16 个指标转换为 3 个主成分。

主成分载荷数值的大小代表各个变量与主成分之间的关联度。如表 3 所示,由主成分载荷矩阵可知,第一主成分中政府卫生支出占财政支出比重、人均社会卫生支出、卫生总费用占 GDP 比重、人均卫生费用、地区 65 岁以上人口占比、人口死亡率、每千常住人口执业(助理)医师数、每千人口医疗卫生机构床位数、医疗卫生机构数、卫生技术人员数、人均 GDP、城镇化率的载荷系数较大,说明第一主成分主要反映卫生、人

口、经济等方面综合因素的作用。第二主成分中人口出生率、失业率的载荷系数较大,第三主成分中传染病发病率、居民消费价格指数的载荷系数较大,说明第二、三主成分是和居民相关的因素在起作用。

表 2 主成分的特征值、方差贡献率和累积贡献率表

成分	特征值	方差贡献率 (%)	方差累积贡献率 (%)
1	10.204	63.778	63.778
2	1.913	11.955	75.733
3	1.417	8.859	84.592
4	0.787	4.919	89.511
5	0.633	3.955	93.466
6	0.470	2.936	96.403
7	0.249	1.557	97.960
8	0.124	0.773	98.733
9	0.093	0.581	99.314
10	0.053	0.330	99.644
11	0.035	0.221	99.864
12	0.014	0.086	99.951
13	0.005	0.034	99.985
14	0.001	0.008	99.993
15	0.001	0.005	99.998
16	0.000	0.002	100.000

表 3 主成分载荷矩阵

指标序号	成分		
	1	2	3
1	0.911	0.260	-0.107
2	0.993	0.009	-0.006
3	0.931	-0.168	-0.096
4	0.991	-0.004	0.052
5	0.959	-0.124	0.019
6	-0.430	0.712	-0.052
7	0.628	0.150	0.077
8	-0.286	0.360	0.804
9	0.844	0.116	-0.162
10	0.945	0.180	-0.121
11	0.848	0.088	0.299
12	0.942	0.158	-0.127
13	0.970	-0.007	0.117
14	0.885	-0.198	0.206
15	-0.115	0.492	-0.705
16	-0.108	-0.885	-0.196

由主成分载荷数值与对应特征值算术平方根相除得出主成分系数,进而形成个人卫生支出占比影响因素主成分的计算式。将标准化后的原始数据带入计算式中即可得到主成分得分,如表 4 所示。

表 4 主成分得分

年份	$F_1$	$F_2$	$F_3$	年份	$F_1$	$F_2$	$F_3$	年份	$F_1$	$F_2$	$F_3$
1995	-0.99	2.37	-2.80	2004	-0.82	-1.39	-0.50	2013	0.40	0.25	0.53
1996	-0.93	1.06	-2.06	2005	-0.70	-1.15	0.16	2014	0.45	0.92	1.18
1997	-0.96	0.67	-0.40	2006	-0.62	-0.85	0.45	2015	0.79	0.12	0.58
1998	-0.85	0.23	0.26	2007	-0.66	-0.11	0.04	2016	1.04	0.13	0.00
1999	-0.90	0.50	1.56	2008	-0.44	-0.02	-0.05	2017	1.49	0.64	0.22
2000	-0.87	1.08	0.65	2009	-0.10	-0.27	0.95	2018	1.48	-0.02	-0.27
2001	-0.82	-0.02	1.13	2010	0.01	0.67	0.38	2019	1.62	-0.29	-0.48
2002	-0.96	-1.98	-0.14	2011	0.02	0.39	0.72	2020	2.03	-0.96	-1.48
2003	-0.76	-2.46	-1.33	2012	0.22	0.57	1.03	2021	1.83	-0.09	-0.33

2.LSTM 模型预测

将表 4 主成分得分数据集作为神经网络模型的输入数据,利用 Python3.10.6 建立 LSTM 模型对 2025 年和 2030 年辽宁个人卫生支出占比进行预测。根据前期测试结果,选取 3 个主成分每 3 年的得分数据作为网络输入,第 4 年的主成分得分作为网络输出,以此类推,得出 24 组样本供神经网络训练和测试使用。将 1~23 组样本作为训练集,通过训练挖掘出数据之间存在的规律并构建模型,第 24 组作为测试集,对训练好的模型进行误差检验。

首先设定神经网络训练次数为 310 次,训练结果显示 3 个主成分的最小均方误差分别为 0.0012、0.0035、0.0074,说明训练模型效果较好。然后将 1995—2021 年辽宁个人卫生支出占比数据集输入模型,与 1995—2021 年 3 个主成分得分一起作为创建向

量机的数据基础,探测 3 个主成分与辽宁个人卫生支出占比之间的关系,并以此作为预测逻辑基础。

在将数据进行归一化、标准化处理后以每 3 年为单位创建数组来预测下一年的数值,直至预测出 2030 年主成分得分。为验证预测的准确度,构建 ARIMA (0,1,0)模型进行预测,并将 2019—2021 年辽宁个人卫生占比两种方法的预测结果与真实值进行对比(见表 5),结果显示 PCA-LSTM 的预测误差显著小于 ARIMA 模型,且与真实值的误差逐年缩减,趋近于零,模型预测效果比较理想。

在预测出未来 9 年的主成分得分后,系统根据之前探测到的 3 个主成分与辽宁个人卫生支出占比之间的关系对 2022—2030 年辽宁个人卫生支出占比进行预测。2022—2030 年主成分得分与辽宁个人卫生支出占比预测结果如表 6 所示。

表 5 2019—2021 年辽宁个人卫生支出占比预测结果

年份	真实值 (%)	PCA-LSTM 预测值 (%)	PCA-LSTM 绝对误差 (%)	PCA-LSTM 相对误差 (%)	ARIMA 预测值 (%)	ARIMA 绝对误差 (%)	ARIMA 相对误差 (%)
2019	32.51	31.54	-0.97	-2.98	33.71	1.20	3.69
2020	30.78	30.42	-0.36	-1.17	33.69	2.91	9.45
2021	29.98	29.95	-0.03	-0.10	33.68	3.70	12.34

表 6 2022—2030 年主成分得分与辽宁个人卫生支出占比预测值

年份	第一主成分	第二主成分	第三主成分	预测值 (%)	降幅 (%)
2022	1.932590	0.377170	0.277986	28.95	
2023	2.046024	0.393925	0.304330	28.35	0.60
2024	2.104203	0.458671	0.317017	27.78	0.57
2025	2.228265	0.453328	0.288808	27.13	0.65
2026	2.295485	0.453776	0.253971	26.74	0.39
2027	2.348425	0.453582	0.218202	26.36	0.38
2028	2.404357	0.451054	0.181810	26.06	0.30
2029	2.438371	0.449221	0.155565	25.85	0.21
2030	2.465753	0.447538	0.133096	25.68	0.17

根据 PCA-LSTM 神经网络模型的预测结果显示未来 9 年辽宁个人卫生支出占比呈缓慢下降趋势。2025 年、2030 年的预测值均略高于规划目标值。

## 讨 论

1. 辽宁个人卫生支出占比缓步下降, 实现规划目标承压

根据预测结果, 2022—2030 年辽宁个人卫生支出占比呈现持续下降趋势, 但下降幅度逐步减小, 2025 年指标预测值略高于当年规划值, 2030 年的指标预测值与国家规划目标值相差 0.68 个百分点, 预测结果提示辽宁未来完成该指标的规划任务还十分艰巨(见表 6)。从历史数据上看, 这与辽宁该指标值长期较高有很大关系, 2021 年辽宁个人卫生支出占比为 30.0%, 相较全国同期平均水平 27.7% 尚有 2.3 个百分点的差距, 完成与国家一致 2025 年规划目标显然有较大的压力。辽宁个人卫生支出占比偏高是多方因素叠加的结果, 因此对于如何有效促进该指标的稳步下降, 需要统筹各方因素, 多措并举。

2. 合理控制个人卫生支出上涨, 减轻居民疾病经济负担

卫生总费用反映了全社会用于医疗卫生的资金消耗情况, 受健康状况、人口总量及结构影响, 同时医疗资源配置、医疗保障、现代医疗技术、物价水平等社会经济因素也与其高度密切相关<sup>[7]</sup>。随着社会经济快速发展与先进医疗技术的广泛应用, 居民卫生健康服务需求不断释放。更需要关注的是, 辽宁 2020 年 65 岁及以上人口占比 15.9%, 除上海外高于其他省份, 可见其人口老龄化程度很高, 而人口老龄化必然会驱动医疗卫生支出的长期增长<sup>[8]</sup>。因此应根据地方人口结构特点与卫生健康需求优化医疗卫生服务体系合理控制个人卫生支出的上涨, 以预防为主, 完善疾病预防控制体系, 针对重点人群特别是老年人群进行健康干预, 提升健康素养水平。以基层为重点, 促进优质资源下沉, 提升基层医疗卫生服务能力, 加强健康管理, 促进居民合理就医。

3. 强化政府主体责任, 加大医疗卫生投入力度

个人卫生支出占比受卫生总费用内部结构的影响, 与政府卫生支出、社会卫生支出占比存在此长彼消的关系。2010 年、2015 年、2020 年辽宁政府卫生支出占比分别为 22.8%、20.7%、23.4%, 远低于全国平均水平。数据表明辽宁政府卫生支出相对不足, 这直接影响卫生筹资的结构, 致使社会与个人的筹资压力上升<sup>[9]</sup>。因此应进一步强化政府在医疗卫生投入的主体责任, 健全在卫生健康领域的投入机制, 在政府财政允许的情况下适度提升筹资水平。

4. 完善多元筹资机制, 推动各筹资渠道协同发展

增加社会卫生筹资力度可以减轻政府卫生支出大幅上升的压力, 利于减少个人卫生支出占比。应拓宽社会筹资渠道, 深化医疗保障制度改革, 持续提高基本医疗保险的筹资能力, 在保证医疗基金可持续性的基础上, 提升居民医疗保障待遇水平, 完善城乡居民大病保险制度、医疗救助制度, 避免低收入群体的灾难性卫生支出和因病致贫。健全老年人等重点人群的医疗保障制度。鼓励商业健康保险企业根据居民个人需求开发个性化、多样化、多层次的医疗保险产品, 特别是支持和推广普惠健康保险, 切实减轻居民的医药消费开支过高的疾病负担。

5. 科学预测是编制规划指标的前期必要工作, 具有重要的参考价值

个人卫生支出占比是各类卫生健康规划的重要约束性指标, 必须科学谨慎制定。个人卫生支出与卫生总费用的影响因素众多, 且各因素之间存在一定的相关性, 单变量的时间序列分析如 ARIMA 模型可以充分利用挖掘历史数据本身的信息和变化特点, 建模方便, 但对于受多种因素指标预测往往误差较大。PCA-LSTM 模型通过主成分分析将众多因素进行降维处理, 并构建神经网络预测模型, 将历年的各主成分得分输入模型进行预测, 模型预测在充分反映数据特点的同时, 又考虑到各因素的影响, 可以较准确地判断变化趋势, 提高预测的精度。该方法对于卫生健康领域受多因素影响的规划指标编制提供了重要的分析工具, 具有一定的推广应用价值。在规划此类指标前, 可利用该方法进行预测, 为卫生健康领域的科学规划提供重要参考。

## 参 考 文 献

- [1] 孟庆跃. 卫生经济学[M]. 北京: 中国协和医科大学出版社, 2017: 66.
- [2] 苗淦, 石福艳, 梁益嘉, 等. 基于优势组合模型的“十三五”期间全国卫生总费用预测研究[J]. 中国卫生统计, 2021, 38(2): 243-245.
- [3] 万泉, 李涛, 柴培培, 等. 我国个人卫生支出占卫生总费用比重监测评价与预测研究[J]. 中国卫生经济, 2021, 40(5): 33-36.
- [4] 翟梦梦, 王旭春, 任浩, 等. 基于 Python 的 LSTM 模型对流感预测的研究[J]. 中国卫生统计, 2022, 39(2): 162-166.
- [5] 宋丽娜. 基于情感分析和 PCA-LSTM 模型的股票价格预测[J]. 中国管理信息化, 2021, 24(21): 159-161.
- [6] 张可, 崔乐. 基于 PCA-LSTM 模型的多元时间序列分类算法研究[J]. 统计与决策, 2020(15): 44-49.
- [7] 李秋莎, 尹文强, 宋燕楠, 等. 城镇化、老龄化及政府卫生支出对我国卫生总费用的影响研究[J]. 中国卫生经济, 2019, 38(9): 42-46.
- [8] 刘丽丽, 周恭伟. 我国卫生总费用现状分析及影响因素研究[J]. 卫生软科学, 2022, 36(8): 45-49.
- [9] 张毓辉, 万泉, 王秀峰, 等. “十三五”时期中国卫生总费用监测预警研究[J]. 卫生经济研究, 2017(1): 8-13.

(责任编辑: 郭海强)