**新冠疫情下互联网医疗的资源配置及对分级诊疗的影响**

蒋世华 郑经天 齐淏仑 熊绍琳

天津大学 天津市 300072

****[基金项目]****基金：国家自然科学基金（202110056274）

**摘要：**“互联网 + 医疗”是我国改善就医环境、促进医疗资源合理分配、创新医疗服务模式和理念的重要一环，突破了时间和空间对医疗资源分布的限制，为患者提供更为便捷、高效的医疗服务。本文在针对线下医疗存在的城乡医疗资源服务差距大、病患就医扎堆和各科室医生分布不均等问题，利用K-Prototypes技术对“好大夫在线”平台获取的合计2952447条2019年、2020年医生及订单数据进行医生画像聚类研究，分析“互联网 + 医疗”服务的现有医疗资源的资源配置问题和对分级诊疗制度实施的影响。

**关键词：**“互联网 + 医疗”；医生画像；医疗资源配置；分级诊疗；K-Prototypes

**Resource allocation of Internet medical treatment and its impact on graded diagnosis and treatment under COVID-19**

Jiang Shihua Zheng Jingtian Qi Haolun Xiong Shaolin

School of Tianjin, Tianjin, 300072

**Abstract**:"Internet + medical" is an important part of improving the medical environment, promoting the rational allocation of medical resources and innovating the mode and concept of medical service in China. It breaks through the limitation of time and space on the distribution of medical resources and provides patients with more convenient and efficient medical services. In view of the large gap in urban and rural medical resource services, the clustering of patients' medical treatment, and the unequal distribution of doctors in various departments, this paper used K-prototypes technology to carry out a clustering study of doctor portraits for the total 2,952,447 doctors and order data acquired from The "Good Doctor Online" platform in 2019 and 2020. This paper analyzes the allocation of existing medical resources of "Internet + medical service" and its impact on the implementation of hierarchical medical system.

**Key words**: "Internet and medical care"; user portrait; allocation of medical resources; graded diagnosis and treatment; K-Prototypes

2015年7月，国务院发布了《关于积极推进互联网 + 行动的指导意见》，要求各地方发展在线医疗卫生新模式，要求推动医疗信息跨地区、跨机构的互联互通，保障在线预约诊疗、诊断信息查询、医疗费用交付、药品供应的便捷可行，促进智慧养老服务的推广，充分开发互联网、大数据等信息技术在分级诊疗的潜力，这是从国家层面第一次提出将互联网与医疗卫生行业结合，标志着传统医学诊疗路径的格局被打破。随后国家卫生健康委员会和国家中医药管理局在2018 年制定了《互联网医院管理办法（ 试行） 》《互联网诊疗管理办法（ 试行） 》《远程医疗服务管理规范（ 试行） 》三项政策，对“互联网 + 医疗”机构的门槛准入、申请流程、人员和仪器配备、机构监管等作出了详尽的说明，“互联网 + 医疗”进一步规范。2019 年底新冠肺炎席卷世界多国，线下医疗的艰难性和风险性推动了“互联网 + 医疗”继续发展，线上就医注册用户数量突飞猛进，“互联网 + 医疗”逐渐在我国医疗方式中占据了一席之地。不过“互联网 + 医疗”现阶段仍处于快速发展和扩张的早期，在许多方面存在不足，因此需要理清“互联网 + 医疗”的相关主体和运行过程，多方位全面查找影响因素并进行干预，才能保证“互联网 + 医疗”政策的深入推进。

根据现有的理论和分析[[[1]](#endnote-0)]，“互联网 + 医疗”突破了当前医疗资源分布的时间和空间局限，为患者提供更为便捷、高效的医疗服务。互联网医疗是“互联网 + 医疗”与传统医院和企业有机结合后产生的，依托各地方医院，运用互联网技术提供便捷合适的医疗服务。但根据近期的发展现状显示，“互联网 + 医疗”不仅有可能阻碍医药卫生体制改革政策中分级诊疗制度的实施，而且在信息安全、诊治一体化、服务监管等方面都存在一定问题。本研究针对的是线下医疗体系存在的“金字塔”结构导致的城乡医疗资源服务差距大、病患就医过于集中和各地方医生分布不均导致的医疗资源的服务效率降低的现象，通过构建医师画像和现有数据对比分析“互联网 + 医疗”服务对现有医疗资源配置的问题，判断“互联网 + 医疗”对分级诊疗制度实施的影响。

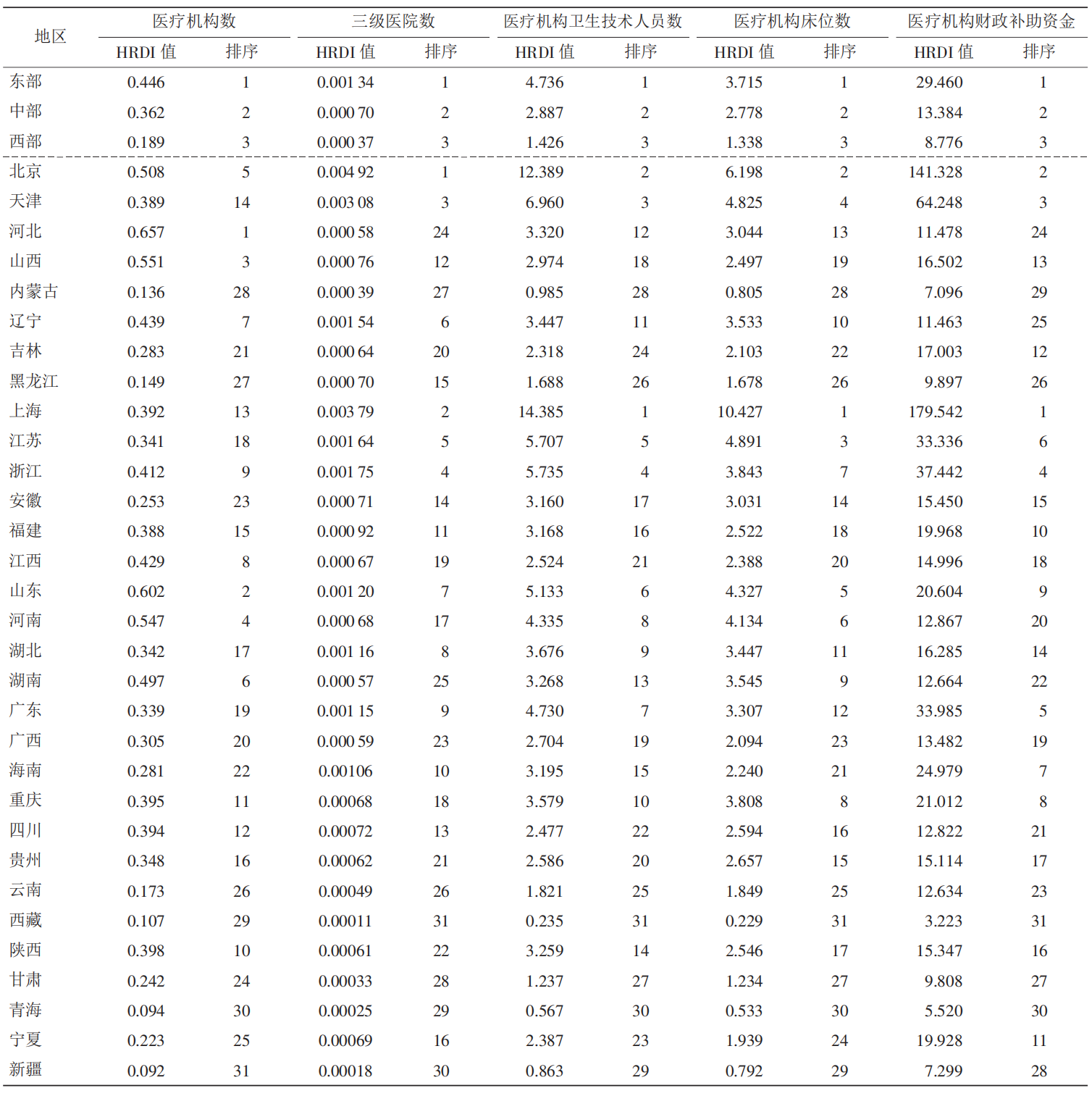
1. **我国医疗资源现状和“互联网 + 医疗”**

**1.1我国国医疗资源现状**

**1.1.1 我国医疗资源配置**

在现代医疗体系构建和人口快速增长的今天，如何合理配置并高效利用现有医疗资源是卫生事业可持续发展的基础。2020 年 6 月 2 日，习近平总书记在专家学者座谈会上强调，“人民安全是国家安全的基石，人类健康是社会文明进步的基础。要加大卫生健康领域科技投入，推动将健康融入所有政策，发展卫生健康事业始终处于基础性地位”。在2016 年国务院发布了《“十三五”卫生与健康规划》，规划中指出中国医疗卫生资源总量不足，结构不合理，制约了卫生与健康事业的改革发展。胡洪曙和武锶芪（ 2019) 研究发现，在中国七类基本公共服务中，医疗卫生模块的供需失调最严重。2019年有学者[[[2]](#endnote-1)]从统计年鉴中获取我国共计31个省级行政区（不包括澳门、香港特别行政区和台湾地区，下同）的医疗服务数量和效率、医疗资源数量、土地面积、人口数量等数据，数据具体包括：各地医疗机构数、三级医院数、医疗机构卫生技术人员数、医疗机构床位数、医疗机构财政补助资金5 项医疗资源指标，医疗机构门急诊人次、医疗机构出院人数、医疗机构住院病人手术人次、医院病床使用率、医院平均住院日、医院医师日均担负诊疗人次、医院医师日均担负住院床日 7 项医疗服务利用指标，以及各地方常住人口数量、土地面积的数据。研究的医疗机构对象包括医院、社区卫生服务中心（站）、街道卫生院、乡镇卫生院、村卫生室、门诊部、诊所（卫生所、医务室、护理站）、专科疾病防治院（所、站）、妇幼保健院（所、站）等。最终发现我国各地的各类医疗资源卫生资源密度指数（Health Resources Density Index，HRDI）值均呈现从东部向西部依次递减的趋势。全国31个省级行政区的卫生资源密度指数分布如表1。

**表**12016**年我国医疗资源密度指数**



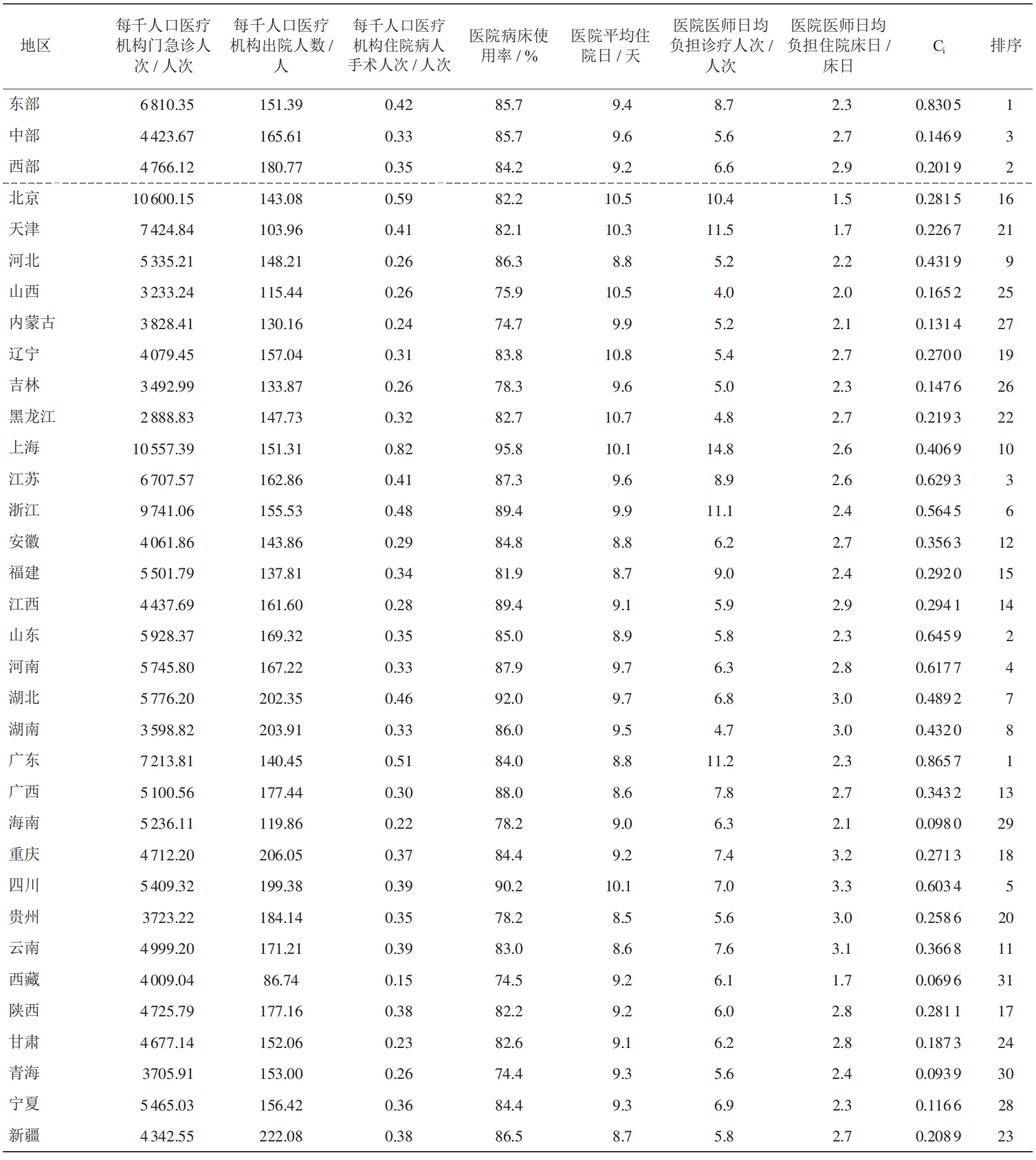
在综合各地方人口数量、土地面积两方面因素后，我国各类别的医疗资源的卫生资源密度指数值均呈现从东部向西部依次递减的趋势。

其中卫生资源密度指数（Health Resources Density Index，HRDI ）值等于每千人口卫生资源和每平方千米卫生资源的几何平均数，反映了卫生资源人口分布和地理分布的综合水平[[[3]](#endnote-2)]。

**1.1.2 我国医疗资源利用率**

在对现有医疗资源的利用率方面，我国31个省级行政区中，只有4个地区的医疗资源密度超过全国平均水平，同时有相对较高的医疗服务利用水平；而医疗资源密度高、服务利用水平低的省级行政区达到了10个。在部分地区，各类医疗资源配置水平和服务利用水平均较低，充分说明现阶段我国现有医疗体系中存在具体且明显的医疗资源配置失衡和浪费问题。全国各地医疗资源利用水平详细如表2。

**表**22016 **年我国医疗资源利用情况**



其中Ci为各地区医疗服务指标值与最优值的相对接近程度[[[4]](#endnote-3)]，Ci越大表明当地医疗服务水平越接近最优水平，从而得到各地区医疗服务利用数量和效率。

**1.2 分级诊疗和互联网医疗**

**1.2.1 分级诊疗的实施**

分级诊疗制度作为实现医疗从全科到专业化的制度，要求按照患者疾病的轻重缓急及治疗的难易程度进行分级，最终让不同级别的医疗机构承担不同级别疾病治疗的医疗制度。[国务院办公厅](https://baike.baidu.com/item/%E5%9B%BD%E5%8A%A1%E9%99%A2%E5%8A%9E%E5%85%AC%E5%8E%85/468516" \t "https://baike.baidu.com/item/%E5%88%86%E7%BA%A7%E8%AF%8A%E7%96%97/_blank)在2015年9月8日发布的《[关于推进分级诊疗制度建设的指导意见](https://baike.baidu.com/item/%E5%85%B3%E4%BA%8E%E6%8E%A8%E8%BF%9B%E5%88%86%E7%BA%A7%E8%AF%8A%E7%96%97%E5%88%B6%E5%BA%A6%E5%BB%BA%E8%AE%BE%E7%9A%84%E6%8C%87%E5%AF%BC%E6%84%8F%E8%A7%81/18611492" \t "https://baike.baidu.com/item/%E5%88%86%E7%BA%A7%E8%AF%8A%E7%96%97/_blank)》首次提出以强化基层医疗资源为重点，通过完善分级诊疗服务体系、建立健全分级诊疗保障机制、组织实施等方面构建分级诊疗制度的意见。

在 2020年初新冠肺炎疫情的爆发后，疫情不仅对全世界人民的生命健康安全造成巨大威胁，也暴露出我国长期存在的医疗资源配置不均衡问题。北京清华长庚医院董家鸿院士在参与疫情增援和调研后表示“医疗资源不是最大问题，资源配置不合理才是核心问题”面对国内不均匀的医疗分布和部分地区存在的医疗资源剩余。国务院办公厅在2020年7月发布的《国务院办公厅关于印发深化医药卫生体制改革2020年下半年重点工作任务的通知》中明确指出，要推进分级诊疗和医药卫生信息化建设，提升各城乡社区医疗服务能力，强化基层卫生防疫能力，促进全国医药卫生管理和服务模式。

**1.2.2 “互联网 + 医疗”和分级诊疗**

在 2020 年底，我国“互联网 + 医疗”的服务应用已达千余个，主要分为以医院主导和以企业为主导的两种“互联网 + 医疗”服务，其中医院主导的“互联网 + 医疗”服务具有运营权，就医服务由医院提供，医生的医疗行为受医院管理，问诊费用由医院统一结算后分配，就医流程相对规范且质量更接近线下就医的服务。在我国全力开展新冠肺炎疫情防控阻击战中，医院主导的“互联网 + 医疗”服务快速发挥了互联网特有的跨地域、零接触的优势，为患者提供了日常诊疗与预防诊疗的功能。各大医疫情防控间间纷纷开启“互联网 + 医疗”的实时预防诊疗平台，在降低院内感染几率、阻止疫情蔓延方面发挥重要作用。但医院主要的“互联网 + 医疗”服务用户群体相对较窄，诊疗服务尚未得到广泛应用，目前主要提供仍然然是预约挂号、智能分诊、科室医生查询、健康咨询、医院导航等功能；且处方单最终会流转交至企业，整体存在医药分离的局面。企业主导的“互联网 + 医疗”服务不受线下医院的束缚，有很高的自主权，收益主要来源于问诊费分成及处方药返点，服务水平取决于所聘请的医生的数量和质量，但提供的功能相对医院主导的“互联网 + 医疗”服务更少，所提供的服务并不能满足患者的医疗需求。

然而，根据“互联网 + 医疗”的现状来看，其在一定程度上可能阻碍了我国医改政策中分级诊疗制度的实施，而且在线上线下一体化、服务监管、信息安全等方面存在诸多问题。

分级诊疗制度作为实现医疗从全科到专业化的制度，要求按照患者疾病的轻重缓急及治疗的难易程度进行分级，最终让不同级别的医疗机构承担不同级别疾病治疗的医疗制度。在现有的两种“互联网 + 医疗”模式中，以企业为主导的“互联网 + 医疗”服务以盈利为目的难以兼顾医疗服务的公益性，同时行业内部竞争激烈，医疗服务同质化严重，通常无法解决患者实质问题，在形成完整医疗服务商业闭环的过程存在诸多问题，使以企业为主导的“互联网 + 医疗”服务在出现后的7年以来始终无法摆脱盲目扩张、恶性竞争的发展阶段；以医院主导的“互联网 + 医疗”服务具有更为良好的发展环境和空间，根据相关资料显示，国家对省级医院全面开展“互联网 + 医疗”服务后，各医院在互联网平台上提供的服务不断增加。但是患者通过各医院的线上平台进行预约挂号、在线问诊等更为方便的医疗活动，跨越了基层医疗机构的首诊机制，忽视双向转诊的医疗政策，使得“大医院吃不了，基层医院吃不饱”的现象愈发严重，进一步突出了我国现存的优质医疗资源过度集聚和部分地区医疗资源配置效率有待提高的问题。

**1.2.3 “互联网 + 医疗”的资源配置和对分级诊疗制度的影响**

针对2020年我国现有的“大医院吃不了，基层医院吃不饱”愈发严重的现象，其即受到来自于2020年爆发持续至今的新冠疫情对现有医疗系统的压力的影响，也受到来自“互联网 + 医疗”服务的影响，无法明确“互联网 + 医疗”的医疗资源配置是否激化了我国现存的优质医疗资源过度集聚和部分地区医疗资源配置效率有待提高的问题。

**1.3 文献综述**

Wijayanto等通过K-Prototypes对多维度的贫困数据进行了聚类分析[[[5]](#endnote-4)]，Liu等将聚类技术用于构建商业大数据的用户画像中[[[6]](#endnote-5)]，Wu等将用户画像应用于会计数据之中[[[7]](#endnote-6)]，张瑞利等分析了互联网+医疗服务的三种模式并进行对比，好大夫平台属于实体医院与互联网医疗平台的服务融合模式(H+I模式)[[[8]](#endnote-7)]，张泽洪等使用无序多项Logistic回归方法对好大夫平台进行分析，分析影响服务供给的主体驱动因素[[[9]](#endnote-8)]，武慧娟等使用K-Means聚类方法对数字阅读用户进行了聚类分析并形成用户画像[[[10]](#endnote-9)]，唐晓波等通过医生的特征对医生构建医生画像[[[11]](#endnote-10)]。

付硕雄等提出了可通过提升收入或福利等多种途径，对线上医务人员进行激励从而吸引更多人才涌入互联网医疗行业中[[[12]](#endnote-11)]，同样地，张泽洪等在政策建议中也提出了注重对在线医生的价值主张与诉求，满足不同在线医生的个性化诉求，例如在线医生的收入和未来的职业发展等各个维度9。因此对于线上医生的个人价值与相关需求方面值得被重视。

本文运用K-Prototypes技术聚类分析，通过构建医生的医生画像对比线上医疗资源的配置与线下医疗资源的配置，从医生的地区、职称、科室、擅长、性别、医院级别和工龄七方面对“互联网 + 医疗”服务中的医生进行数据挖掘，构建不同的医生画像，通过不同医生画像在“互联网 + 医疗”服务中的比例和在“互联网 + 医疗”服务中所获得的收入进行分析并且和线下医疗资源配置和利用率进行对比，判断“互联网 + 医疗”的资源配置失衡程度和对分级诊疗制度实施的影响。

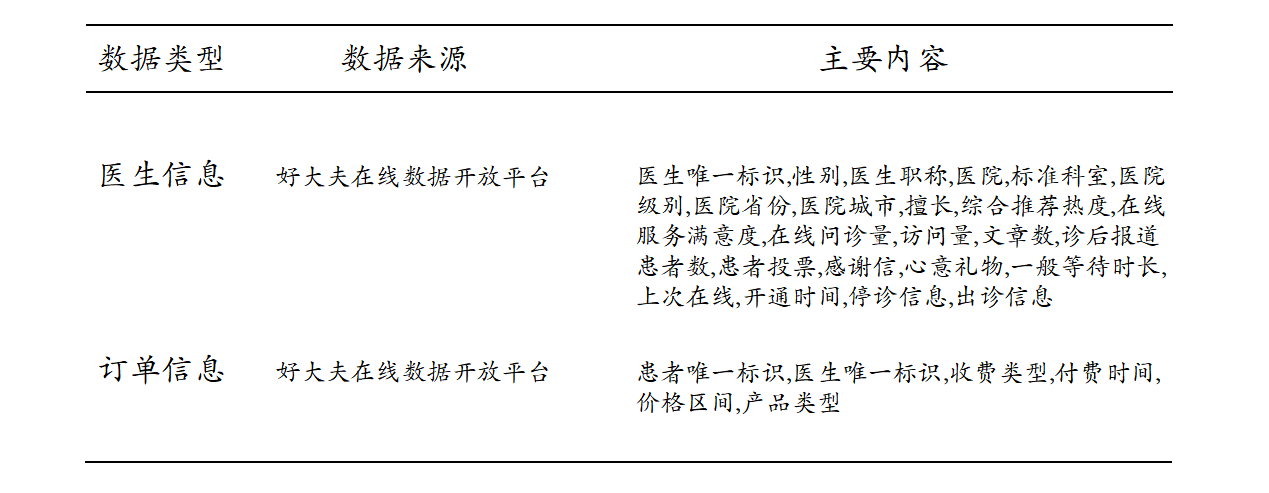
**2 .基于K-Prototypes聚类的医生画像构建**

**2.1数据收集**

本文选定国内大型互联网医疗平台“好大夫在线”进行研究，在“好大夫在线”平台官网申请到了2019、2020年的医生信息和订单成交信息，其中2019年的医生数据和订单数据分别为207169条和1420859条，2020年的医生数据和订单数据分别为234037条和1090382条，考虑到数据量比较庞大，本文将分别对两年的数据进行聚类分析，并对2020年的数据的聚类结果和用户画像进行详细的分析阐述。

医生数据集由医生唯一标识、性别、医生职称、医院、标准科室、医院级别、医院省份、医院城市、擅长、综合推荐热度、在线服务满意度、在线问诊量、访问量、文章数、诊后报道患者数、患者投票、感谢信、心意礼物、一般等待时长、上次在线、开通时间、停诊信息和出诊信息组成，主要包括医生人口统计信息、专长信息、科研信息、患者评价这几方面的信息.订单数据集由患者唯一标识，医生唯一标识，收费类型，付费时间，价格区间，产品类型构成，包括了基本的医患信息和订单成交情况。

**表**3 **数据来源、主要内容、数据类型**



**2.2数据预处理**

**2.2.1 删除缺失值、重复值**

在数据预处理过程中，首先对缺失值、重复值进行了处理，剔除了部分重要信息缺失或完全重复地记录。在数据中存在部分医生所属医院显示未收录，由于该特征是医生特征的重要标识，剔除了所属医院为“未收录医院”的数据。

**2.2.2 变量说明**

此外，为了后续研究和分析的方便，同时使医生的特征更加突出，统一了医院级别、医院地域、医生职称、标准科室的划分，医院被划分为一级医院、二级医院、三级医院；医院省份被划分为的西部、中部、东部；医生职称被划分为正高、副高、中级、初级医生；标准科室按照“好大夫在线”平台官网所划分科室类型，划分为内科、外科、妇产科、儿科、五官科、急诊、医技、中医等更高层次的类别特征11。

**2.2.3 数据融合**

根据订单数据集中的信息，通过医生信息与订单信息的共同特征——医生唯一标识，将两个数据合二为一，最终得到2019年的有效数据65409条，2020年的有效数据64029条。

**2.2.4 序号编码及分词**

然后对类别特征（性别、医生职称、标准科室、医院级别、医院省份）进行序号编码，并利用哈尔滨工业大学的LTP分词对医生的擅长特征进行分词，以实现后续医生画像的构建11。

**2.2.5 文本特征**

擅长特征属于文本特征，对分词后的擅长特征进行向量化，然后通过SVD降维最终筛选包含总信息53%的200个维度作为数值特征。

**2.2.6 标准化**

Kmodes库中的K-Prototypes默认使用欧式距离计算数值特征，同时由于本文所选用的大数据集，为提高算法的计算效率，对数值特征（文章数、工龄）进行标准化。

**2.3 数据处理**

**2.3.1特征选择**

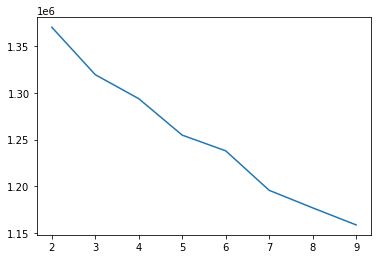
根据研究需要，首先从已有数据集中选择了能够代表医生个人特征的人口统计信息、专长信息、科研信息，并根据医生开通时间在线问诊账户的时间构造了表示医生在线服务时长的“工龄”特征，最终得到性别、医生职称、医院级别、医院省份、工龄、标准科室、文章数7个特征用于聚类分析。

**2.3.2 K-Prototypes聚类**

由于数据集呈现类别变量与数值变量混合的特点，而常用K-Means算法如果对混合型变量进行聚类，需要将各个类别变量转换为哑变量，使其成为数值型变量，然后通过计算欧几里得距离得出聚类结果。本文选取的聚类特征中类别变量比较多，如果使用K-Means算法，计算过程会比较繁琐。

因此，本文使用了处理混合属性聚类的典型算法：K-Prototypes算法。K-Prototypes算法继承了Kmean算法和Kmode算法的思想，并且提出了混合属性簇的原型。混合属性中存在数值属性和类别属性，数值属性原型为属性中所有属性取值的均值，类别属性原型为类别属性中选取属性值取值频率最高的属性。K-Prototypes算法无需创建哑变量，将分别为类别变量计算汉明距离、为数值型变量计算欧几里得距离然后得出聚类结果，算法运行结束后将使用成本函数（所有点到各自簇质心的距离之和）评价聚类效果。本文使用K-Modes库中的K-Prototypes方法来进行聚类。

手肘法是一种启发式方法，用于确定数据集中的聚类中心数目[[[13]](#endnote-12)]，该方法绘制了一个关于聚类中心数目和成本函数的图，并选择曲线的拐点作为最终选择的聚类中心数目。为了找到K-Prototype聚类算法的最佳聚类数，在数据集上运行手肘法，利用手肘法计算k值范围为2到9的成本函数值。然后，使用Matplotlib对不同的聚类中心数目分别绘制对应的成本函数，通过寻找成本函数下降的“拐点”确立最优的聚类中心数，“拐点”即为成本函数曲线下降幅度最快的点。



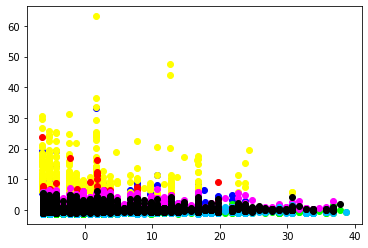
**图1 手肘图**

根据图1确定聚类中心数目K=7时，聚类效果最佳。

**2.4 模型优化**

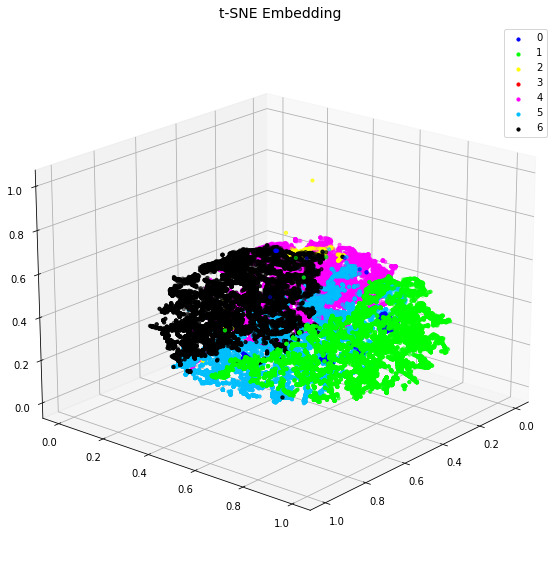
**2.4.1 聚类可视化**

通过使用降维对聚类结果进行可视化，通过降维到二维或者三维空间从而观察聚类的结果是否可行[[[14]](#endnote-13)]。同时降维的方法有线性与非线性的，由于聚类属于非监督学习，于是先对聚类后的结果进行线性降维，如果原始数据在二维或者三维空间可用线性结构或其组合准确描述或近似描述，则视为聚类结果可用线性降维显示，不然则通过非线性的降维方法来进行降维。

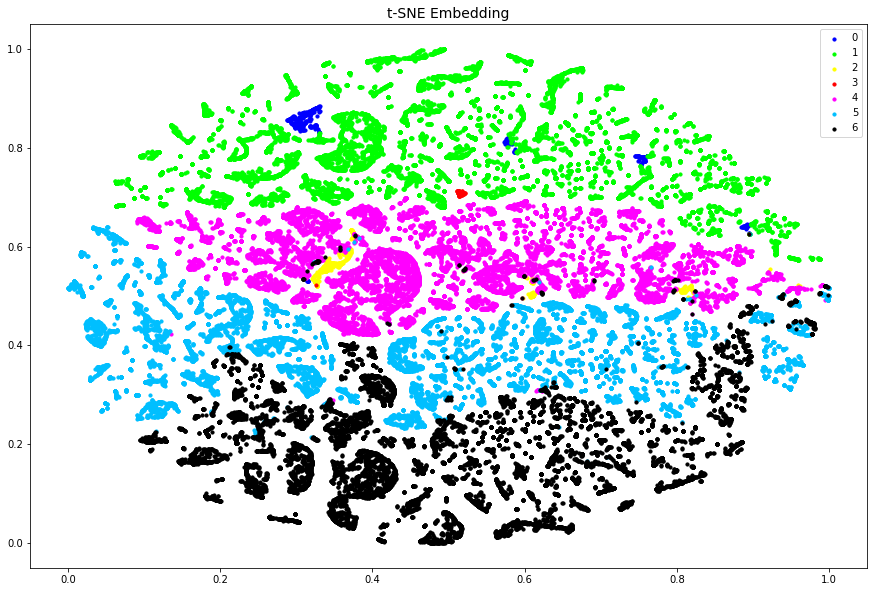


**图2 PCA降维(n\_clusters=7)**

首先使用了PCA降维进行可视化，由图2看出线性降维对于聚类结果的显示效果并不是很好，于是再使用非线性降维方法——tsne降维。

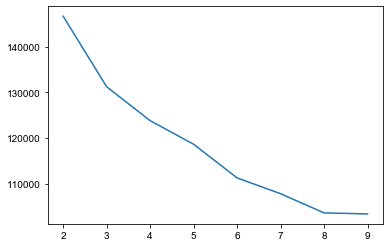


**图3 t-sne3D降维(n\_clusters=7)**

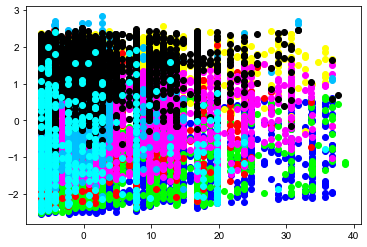


**图4 t-sne2D降维(n\_clusters=7)**

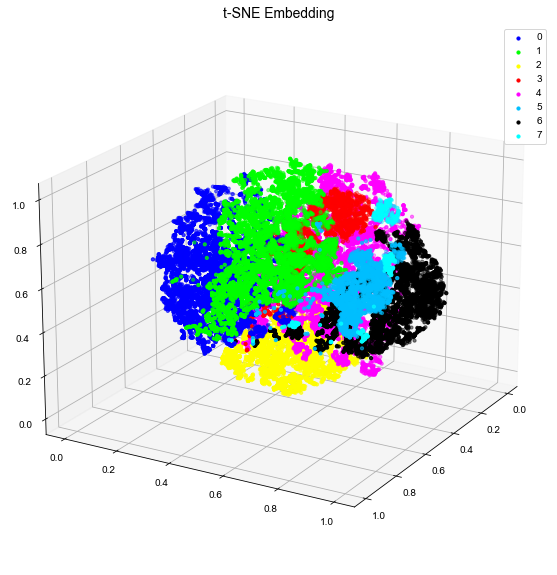
分别对t-sne进行了2D和3D的可视化，由图4可看出t-sne对该聚类结果降维效果好，且该聚类效果好，由图3可看出第2类存在异常值，于是删除异常值后重新聚类[[[15]](#endnote-14)]。



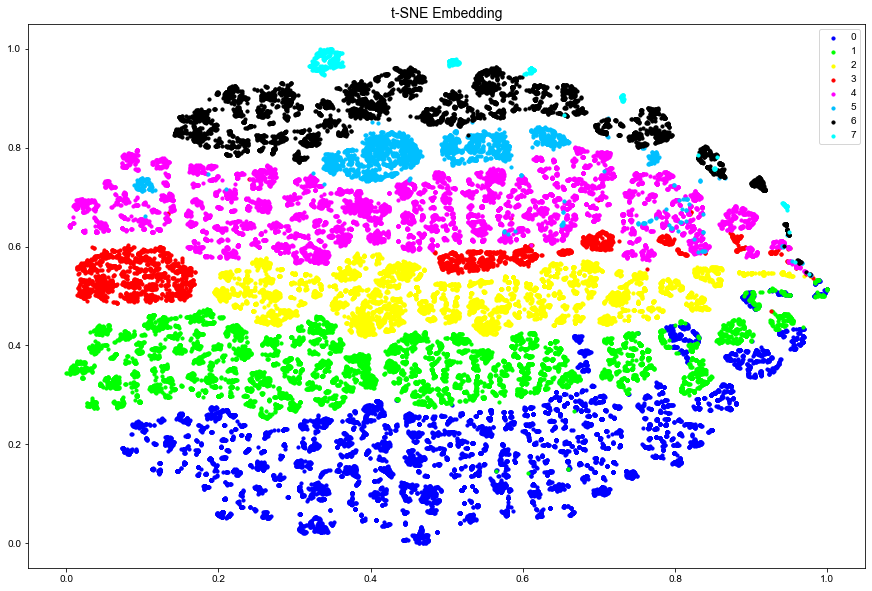
**图5 手肘图(无异常值)**



**图6 PCA降维(无异常值)**



**图7 t-sne3D降维(无异常值)**



**图8 t-sne2D降维(无异常值)**

由图8可看出聚类数目为8时聚类效果好，最终选择聚类数目为8并观察每类的特征。

**2.4.2 改进的K-Prototypes算法**

Chen提出K-Prototypes算法的初始中心选取的改进方法[[[16]](#endnote-15)]，K-Prototypes算法中对于数值变量是使用K-Means来计算，K-Means初始点的选取是通过均值与方差选取的点，改进的K-Prototypes则采取分组平均法来取得初始点，本文通过采取Chen16提到的分组平均法改进对数值变量的初始点选取，比较了两种方法的成本，结果如表4。

**表4 K-Prototypes与改进的K-Prototypes的成本对比**

|  |  |  |
| --- | --- | --- |
| init设置  算法 | K-Prototypes | 改进的K-Prototypes |
| Huang | 103606.42 | 104837.62 |
| Cao | 104558.18 | 105248.52 |

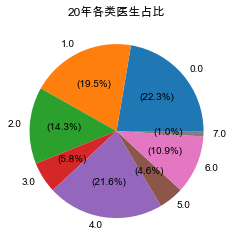
由表4对两种算法的比较看出，分组平均法的聚类结果没有原来的聚类结果好，于是最后选取了Kmodes库中的K-Prototypes方法进行聚类。

**2.5 医生画像构建**

为使聚类结果更为直观，对聚类得到的结果进行进一步的可视化，对聚类过程中的数值变量进行图表的可视化，对类别属性进行词云的绘制形成医生画像的构建。本文使用Python中的wordcloud库，根据文本中词语出现的频率等参数绘制词云，构建医生画像。

**3.互联网医疗平台医生诊疗收入影响因素分析**

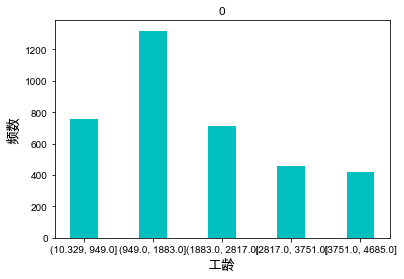
**3.1聚类分析**



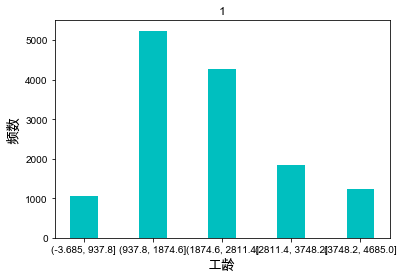
**图9 各类医生占比图**

**3.1.1 图表分析**

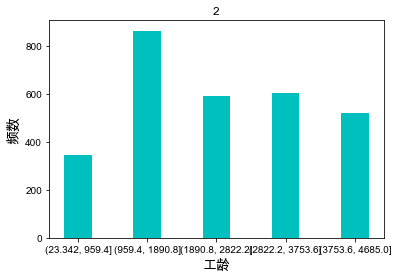
通过图表对每类医生的工龄进行可视化分析，第0类、第1类、第2类、第3类、第4类、第5类、第6类、第7类的图表分别如图10、图11、图12、图13、图14、图15、图16、图17。



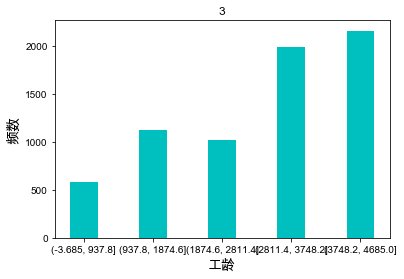
**图10 工龄（第0类）**

****

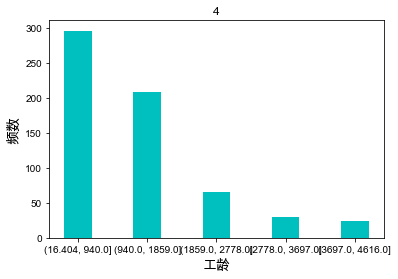
**图11 工龄（第1类）**



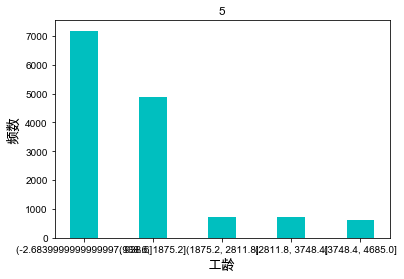
**图12 工龄（第2类）**

****

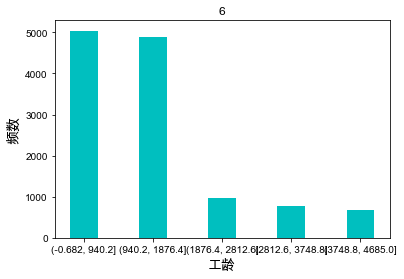
**图13 工龄（第3类）**



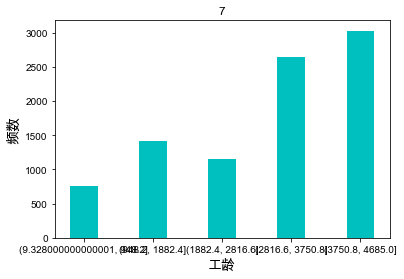
**图14 工龄（第4类）**

****

**图15 工龄（第5类）**



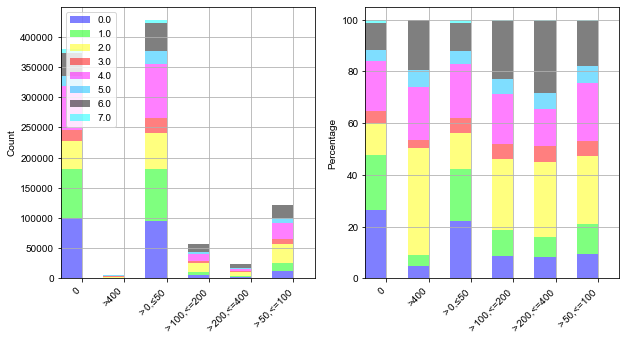
**图16 工龄（第6类）**

****

**图17 工龄（第7类）**

**3.1.2 价格区间**

通过图表和医生画像对每类医生各种特征的分析，并观察每类医生在不同价格区间上的分布情况，每类医生在不同价格区间的数量及占比如图18。



**图18 价格区间**

**表5 不同价格区间对应医生人数**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 价格区间 | 0 | ＞0,≤50 | ＞50,<=100 | ＞100,<=200 | ＞200,<=400 | >400 |
| 医生人数 | 382454 | 430233 | 121815 | 56747 | 24000 | 6259 |

第0类：在“0”、“>0,<=50”区间中大于在20%，在“>50,<=100”、“>100,<=200”、“>200,<=400”、“>400”区间的占比5%左右。第0类医生在低价格、中等价格的订单数一般，在高价格的订单数较少。总体的订单数偏少且从低价格订单数到高价格订单数逐渐减少。

第1类：在“0”、“>0,<=50”区间占比在20%左右，其余区间占比都在5%～10%之间。第3类医生在低价格的订单数较多，在中等、高价格的订单数较少。总体的订单数一般且分布差距较大，偏向低价格。

第2类：在“>400”区间中占比最大，接近55%，且在“>50,<=100”、“>100,<=200”、“>200,<=400”三个区间中占比都在20%左右。第2类医生在高价格的订单数较多，中等价格的订单数分布较为平均，低价格的订单数较少。总体的订单数较多。

第3类：在所有区间中占比在5%左右。第3类医生总体的订单数偏少且分布均匀。

第4类：在“>200,<=400”区间中占比在15%左右，在“>50,<=100”、“>100,<=200”、“0”、“>0,<=50”、“>400”区间中占比在20%左右。第4类医生在整体订单数分布较为平均，一般且分布较分散。

第5类：和第3类相似，在在所有区间中占比在5%左右。第5类医生在低价格、中等价格的订单数和第3类差不多，在高价格的订单数多于第3类。总体的订单数一般且分布均匀分散。

第6类：在“0”、“>0,<=50”区间中占比在10%左右，在“>50,<=100”、“>100,<=200”、“>400”区间的占比在20%左右，在“>200,<=400”区间占比25%。第6类医生在低价格、中等价格的订单数一般，在高价格的订单数较多。总体的订单数一般且从低价格订单数到高价格订单数逐渐增加。

第7类，在“0”、“>0,<=50”区间中占比少于5%，其他区间几乎没有。第7类医生在低价格订单数很少，总体订单数少。

**3.1.3 医生画像**

通过医生画像对每类医生的医生职称、医院级别、医院省份、标准科室、性别、擅长进行可视化分析，第0类、第1类、第2类、第3类、第4类、第5类、第6类、第7类的图表分别如图19、图20、图21、图22、图23、图24、图25、图26。



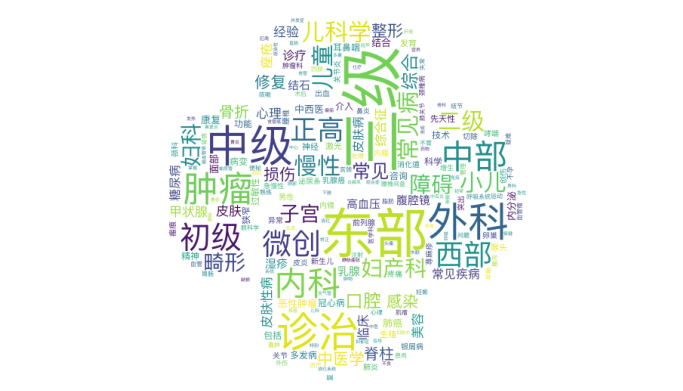
**图19 医生画像（第0类）**

第0类医生集中在三级医院，地区偏东部、中部其次，职称偏正高级别，且外科较多、内科其次，男性较多，擅长肿瘤、微创手术、诊治较多。

****

**图20 医生画像（第1类）**

第1类医生集中在三级医院，地区偏东部、中西部其次，职称偏正高级别，且外科较多、内科其次，男性较多，擅长肿瘤、微创手术、慢性方面较多。

****

**图21 医生画像（第2类）**

第2类医生集中在三级医院，地区偏东部、中部其次，职称偏正高级别、副高其次，且外科较多、内科、儿科学其次，男性较多，擅长畸形、肿瘤、微创手术方面较多。

****

**图22 医生画像（第3类）**

第3类医生集中在三级医院，地区偏东部、中部其次，职称偏副高级别、正高其次，且外科较多、内科其次，男性较多，擅长妇科、慢性、感染、肿瘤、修复、损伤等多方面且分布较为分散。

****

**图23 医生画像（第4类）**

第4类医生集中在三级医院，地区偏东部，职称偏副高级别、中级和正高其次，且外科、儿科学、内科较多、妇产科学其次，男性比女性较多，擅长常见病、儿童、肿瘤方面较多。

****

**图24 医生画像（第5类）**

第5类医生集中在三级医院，地区偏东部，职称偏副高级别、中级和正高其次，且外科、儿科学、内科较多、妇产科学其次，男性比女性较多，擅长常见病、儿童、肿瘤方面较多。

****

**图25 医生画像（第6类）**

第6类医生集中在三级医院，地区偏东部、中部和西部紧随其后，职称偏中级级别、副高和正高其次，且外科较多、内科、皮肤性病科、儿科学、妇产科学校多，男性比女性较多，擅长常见病、中医学、高血压、皮肤病、糖尿病等很多方面且分布较为分散。

****

**图26 医生画像（第7类）**

第7类医生集中在三级医院，地区偏东部、中部其次，职称偏正高、副高其次，且外科较多、内科其次，男性较多，擅长常见病、慢性、感染、肿瘤等多方面且分布较为分散。

**4.结论及建议**

**4.1有关“互联网 + 医疗”服务对线下医疗资源配置的结论**

根据“互联网 + 医疗”服务中收集的医生数据构建的八类医生画像，在医院级别方面，“互联网 + 医疗”的八类医生几乎都集中在三级医院，二级和一级医院的医生在线上平台的占比极小，根据2020年我国卫生健康事业发展统计公报的数据，全国三级医院2996个（其中：三级甲等医院1580个），二级医院10404个，一级医院12252个，未定级医院9742个，去除未定级医院的数据，三级医院在我国医疗卫生机构的占比仅为11.68%，在不同级别的医院中，三级医院医师日均担负诊疗人次为6.3人，二级医院为5.8人，一级医院为4.5人，并无显著差异，“互联网 + 医疗”平台在对不同医院级别的医疗资源配置中极度不合理。

在地区方面，“互联网 + 医疗”的八类医生主要分布在东部地区、中部地区较少、西部地区最少，根据表1所提供的数据，我国在各类医疗资源的卫生资源密度指数值均呈现从东部向西部依次递减的趋势，其中东部地区的医疗机构卫生技术人员数的卫生资源密度指数值为4.736，中部地区为2.887，西部地区为1.426，东部地区三级医院的卫生资源密度指数值为0.00134，中部地区为0.00070，西部地区为0.00037，“互联网 + 医疗”平台在对不同地区的医疗资源配置中相对合理。

在医生职称方面，“互联网 + 医疗”的六类医生主要分布在副高级和正高级，中级较少，据2020年我国卫生健康事业发展统计公报的数据，在技术职务（聘）结构中，高级（主任及副主任级）占8.8%、中级占19.8%、初级占61.7%、待聘占9.7%，其中医生职称与能力之间存在正相关，但在2021年人社部发布的《关于深化卫生专业技术人员职称制度改革的指导意见》中指出医生职称和工作量挂钩，可以预见职称越高的医生在实际医疗环境下承受着更大的医疗压力，同时文献[[[17]](#endnote-16)]指出不同职称、医院等级之间的差异并不存在明显的患者满意度差异，“互联网 + 医疗”在不同职称的医生的医疗资源配置中相对不合理。

在医生科室方面，“互联网 + 医疗”的六类医生主要集中在外科，内科、儿科学、妇产科学依次递减，据2020年我国卫生健康事业发展统计公报的数据，各科门诊量中内科占比28%、外科占比25%、儿科占比8%、妇科占比15%。由于不同科室的患者之间存在在隐私和医疗需求方面的差异，部分科室在线上线下存在较大的医疗服务质量差异，缺乏不同科室的医师数量比例和相关研究，无法得出“互联网 + 医疗”在不同科室的医生的医疗资源配置的合理程度。但在医生擅长方面，“互联网 + 医疗”的八类医生包含了慢性、微创、感染、肿瘤、修复、损伤、常见病、儿童、肿瘤、中医学、高血压、皮肤病、糖尿病等多方面，基本覆盖了患者实际需求的各个层面，所以“互联网 + 医疗”在不同专长和科室的医生的医疗资源配置上相对合理。

在医生性别方面，“互联网 + 医疗”的八类医生主要为男性，女性较少，根据联合国妇女署2020年数据显示，我国女性医生占到了医生总数的约50%，“互联网 + 医疗”服务平台上的男女比例相对失衡，“互联网 + 医疗”在不同性别的医生的医疗资源配置上相对不合理。

在“互联网 + 医疗”服务中所获得的收入中，第2类医生群体拥有着最高的收入，其他类别的医生群体收入相对平均，由于第0、1、4类医生的数量是其他类别医生的两倍左右，实际上第0、1、4类医生在“互联网 + 医疗”服务中所获得的收入是最低的。由于不同职称、医院等级之间的差异并不存在明显的患者满意度差异，在2020年中国医院薪酬报告中也显示出不同科室之间不存在明显的收入差距，“互联网 + 医疗”在不同类别的医生的收入配置上相对不合理。

总体而言，“互联网 + 医疗”的资源配置失衡程度较高，没有对现存的优质医疗资源过度集聚和部分地区医疗资源配置效率有待提高的问题进行改进，过于专注于引入三级医院、高职称的医师，并且将“互联网 + 医疗”的收入倾斜向这两个群体，很可能激化了线下医疗资源配置的问题。

**4.2有关“互联网 + 医疗”服务对分级诊疗影响的结论**

面对分级诊疗制度的实施中优质医疗资源过度集聚的问题，“互联网 + 医疗”在不同医院级别的医师的配置上表现得让人失望，几乎全部为三级医院的医师，在一定程度上阻碍了我国医改政策中分级诊疗制度的实施；面对分级诊疗制度的实施中不同地区医疗资源配置不均匀的问题，“互联网 + 医疗”在不同地区的医师的配置中没有对实际情况做出改进，但在一定程度上缓解了部分地区医疗资源配置效率较低的问题，部分促进了我国医改政策中分级诊疗制度的实施。

总体而言，“互联网 + 医疗”没有发挥它在互联网平台上的跨地域优势，没有合理运用信息技术和先进理论配置现有医疗资源，也没有在分级诊疗制度的实施中应发挥的作用，并在一定程度上阻碍了我国医改政策中分级诊疗制度的实施，需要对“互联网 + 医疗”结构进行一定程度的改革和优化。

**4.3建议**

1. 促进“互联网 + 医疗”医疗线上线下一体化，深入地方基层医疗服务。吸收一、二级医院的医师进入“互联网 + 医疗”平台，凭借广泛的医院基数、互联网技术的支持和医疗信息支撑，促进优质医疗资源向各地方流动，提高基层医疗机构的医疗服务能力和诊疗水平，提升基层的承接能力，逐步构建“基层首诊、双向转诊、急慢分治、上下联动”的分级诊疗新格局。此外，面对我国现有的慢性病患者“寻医难、看病难”问题，“互联网 + 医疗”可以通过线上医疗服务，在社区卫生信息平台中加入慢性病的线上诊差和电子处方系统，让医生可以在线上对患者进行复诊，从而缓解各地方医院对慢性病患者的就诊压力。
2. 促进商业保险加入“互联网 + 医疗”服务。自2013年以来，我国在医疗服务过程中频繁出现医疗事故甚至医闹、医暴等恶性事件，其原因主要是线下医院中医护人员不可避免地会在接诊、检查、护理等过程出现未尽责任或措施不当、治疗态度消极、误诊漏诊等不良行为，与患者及其家属直接的接触和不当的沟通进一步激化了双方的矛盾，给医师、患者和医院都带来了巨大的精神、健康和经济损失。“互联网 + 医疗”服务作为一种发生在网络平台的医疗服务，在避免了医闹、医暴等恶性事件的同时，应当保证患者遭遇医疗事故的权益，引入商业保险则可以在患者遭遇医疗事故时给予患者及时且适当的经济补偿，也可以降低医疗事故对“互联网 + 医疗”服务的负面影响，保证“互联网 + 医疗”服务在群众中的信誉和自身的良性发展。此外，面对患者所关注的信息泄露问题，商业保险也可以作为有信誉的第三方保障患者的信息安全，降低医院或企业在维护信息安全和患者维权时的风险，一步增强“互联网 + 医疗”服务的保障水平。
3. 引入先进的医疗理念和专业的医疗流程。在实际生活中，面对不同医院和地区所存在的医疗资源差异，患者已经形成了对不同医院和地区的医疗服务的主观偏见，而理论研究v已经证明其中部分偏见是非理性的、不客观的，“互联网 + 医疗”作为由医师直接提供和决定医疗服务水平的平台，不存在因医院级别和地区导致的医疗服务水平差异，构建在线下医院的医疗理念并不完全适合“互联网 + 医疗”，应当引入更为先进的医疗理念进行管理和发展。此外，由于人工智能的发展和普及，并且医疗流程中存在简单重复的劳动过程，原本由医师负责的部分医疗流程可以被人工智能取代，降低医师在“互联网 + 医疗”服务中的医疗压力，提高现有医疗资源的利用率。
4. 优化现有的“互联网 + 医疗”平台的人员结构和收入分配。现阶段“互联网 + 医疗”中平台过于看重医院级别、医师职称和医师地区，导致平台将资源和注意力集中在现有医疗资源的一小部分上，让部分医师承担了大部分的互联网医疗压力，平台应该主动承担改变患者的观念的责任，自主优化人员结构，吸收更多的女性医师、中部和西部地区的医师和一、二级医院的医师，吸收现有医疗系统中更多的劳动力，消化部分地区多余的医疗资源，提高全国医疗资源的利用率，并且优化薪酬机制，不再对部分医师提供更高的收入，平均“互联网 + 医疗”平台医师的薪酬水平，刺激更多的医师在平台上发挥自己的能力，进一步增强“互联网 + 医疗”服务的质量和水平。

**参考文献：**

1. [] 周忠良.“互联网+医疗”的现状、问题与发展路径[J].人民论坛, 2021(22):88-91. [↑](#endnote-ref-0)
2. [] 雷鹏,冯志昕,丁荆妮,段睿,余红,刘奇川.中国医疗资源配置与服务利用现状评价[J].卫生经济研究, 2019, 36(05):50-55. [↑](#endnote-ref-1)
3. [] 李丽清,周绪,赵玉兰,卢祖洵.我国东中西部地区基层医疗资源配置与经济发展耦合协调关系研究[J].中国全科医学, 2021, 24(22):2777-2784. [↑](#endnote-ref-2)
4. [] 周明华,谭红,何思长,邵茂,冯潇.基于差别指数和集聚度的四川省基层医疗卫生资源公平性分析[J].现代预防医学, 2021, 48(03):469-472+490. [↑](#endnote-ref-3)
5. [] A. Wijayanto, Y. K. Suprapto and D. P. Wulandari, "Clustering on Multidimensional Poverty Data using PAM and K-prototypes Algorithm : Case Study: Jambi Province 2017," 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2019, 210-215. [↑](#endnote-ref-4)
6. [] Q. Liu, "Business User Portrait Modeling and Clustering Analysis Under The Background of Big Data," 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2021, 666-669. [↑](#endnote-ref-5)
7. [] Y. Wu et al., "User Portraits and Investment Planning Based on Accounting Data," 2020 IEEE International Conference on Services Computing (SCC), 2020, 404-411. [↑](#endnote-ref-6)
8. [] 张瑞利,王刚.“互联网”医疗服务供给：模式比较及优化路径[J].卫生经济研究, 2022, 39(03):32-37. [↑](#endnote-ref-7)
9. [] 张泽洪,熊晶晶.互联网医疗服务供给的主体驱动因素[J].河海大学学报(哲学社会科学版), 2022, 24(01):38-46+110. [↑](#endnote-ref-8)
10. [] 武慧娟,赵天慧,孙鸿飞,史天骄.基于支付意愿的数字阅读用户画像聚类研究[J/OL]. 情报科学:1-11[2022-05-25]. [↑](#endnote-ref-9)
11. [] 唐晓波,高和璇.基于特征分析和标签提取的医生画像构建研究[J].情报科学, 2020, 38(05):3-10. [↑](#endnote-ref-10)
12. [] 付硕雄,陈皓阳,赵海静,莫雯茜,时亚楠,张新平.我国互联网+医疗服务的优化研究——基于复杂适应系统理论[J].中国卫生事业管理, 2022, 39(03):166-169+203. [↑](#endnote-ref-11)
13. [] R. Gupta, H. Kumar, T. Jain, A. Shrotriya and A. Sinha, "Demographic Customer Segmentation of banking users Based on k-prototype methodology," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022, 578-584. [↑](#endnote-ref-12)
14. [] 王开军.基于降维的聚类可视化技术[J].福建师范大学学报(自然科学版), 2011, 27(04):50-54+60. [↑](#endnote-ref-13)
15. [] H. Cho and S. M. Yoon, "Issues in Visualizing Intercultural Dialogue Using Word2Vec and t-SNE," 2017 International Conference on Culture and Computing (Culture and Computing), 2017, 149-150. [↑](#endnote-ref-14)
16. [] X. Chen, "An Improved Clustering Algorithm for Mixed Attributes Data Based on K-prototypes Algorithm," 2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), 2015, 396-399. [↑](#endnote-ref-15)
17. []马弘新,黄秀清.互联网医疗的医院等级歧视研究[J].卫生经济研究, 2017(07):40-44. [↑](#endnote-ref-16)