



# 中医药与大数据

石鑫

山东省济南市长清区山东中医药大学

目前, 医疗产业已经产生海量数据, 其数据类型包括中医古籍专著、名老中医医案、病历、文献期刊、临床经验总结等各种中医诊疗相关数据, 且数据类型和数量仍在增加。然而, 这些医疗相关数据在过去并未得到有效的利用, 主要原因归结于两个方面: 数据收集整理困难, 缺乏数据安全共享平台; 标准化难题即中医固有的经验性、不确定性及模糊性, 难以转化为线性逻辑, 本研究就这两方面的问题及现状进行综述。

大数据相关技术紧紧围绕数据展开, 数据的采集、整理、传输、存储、安全、分析、呈现和应用等等都属大数据的范畴。中医药在抗击新冠疫情的过程中发挥了重要作用, 面临中医药发展的重要契机, 中医药如何借助数字化平台、大数据和人工智能更好的服务于行业的发展和人民健康是目前亟待解决的难题。

## 一、数据收集整理难题

人工智能是交叉学科的产物, 机器学习、自然语言处理等是其主要的研究方向。人工智能和大数据关系密不可分, 大数据是人工智能“思考”和“决策”的基础, 人工智能是大数据进行价值化操作的必备条件。林树元<sup>[1]</sup>基于人工智能结合文献计量学

可视化分析, 总结了现今中医智能化研究中的数据困境导致深度学习效果不佳的难题。深度学习的模型训练需要以大数据为基础, 目前, 中医古籍专著、医院病历、名老中医经验等数据分布散乱, 数据涉及病人的隐私而变得敏感, 因此增加了收集的难度, 且缺乏权威的、安全的数据管理和共享平台, 来源复杂的资料的收集、整合和数据管理将是一项繁琐艰巨的任务。

## 二、标准化难题

中医药大数据主要来源于长期的临床实践和临床记录, 这些数据都是以自然语言的形式存在, 只有将其转换成计算机语言才能被进一步的使用, 其转化的核心是自然语言处理和医学文本识别。然而, 机器的决策是在理解文本语义的基础上进行的, 由于中医学的病例文本大多属于非结构化数据(大数据结构化是指数据标准化处理、指标标准化处理、去重去敏、安全处理等, 只有在数据结构化之后, 数据才能用于挖掘和分析), 中医学有较多非标准化的表达方式, 人工标注成本极高, 导致中医学命名体识别的难度极大<sup>[2]</sup>。其次, 与循证医学不同, 中医辨证论治的治疗原则具有复杂性, 同种疾病会有不同证型和症状, 同种症状和证型可能出现在同一种疾病中, 不同疾病也没有清晰的边界, 这些因素会导致疾病算法模型生成难度增加并使诊断结果不稳定。

DOI: 10.14218/MRP.2021.017

通讯作者: 石鑫 Email: 17862971602@163.com



### 三、近几年的国内外发展

2015年发布《北京中医药文化传播发展报告》认为，将大数据对患者的信息收集、整理、分析以及在中医“治未病”方面的长期信息跟踪和预测用于中医药学术研究和临床经验总结，将成为中医药学术在现代可持续性发展中的正确道路。近年来由于受中医药临床数据的信息采集技术和设备的限制，很多研究机构对中医大数据的研究和建设停留在对古籍文献整理和基础资料汇集的层次，未形成真正符合现代大数据分析意义上的中医药大数据系统，与大数据信息采集连接的中医智能分析系统的形成差之甚远。

随着近几年深度学习的迅猛发展，基于这些数据难题，在 Watson 辅助系统的支持下，国际商业机器公司 (International Business Machines Corporation) 宣布已经拥有 1 亿份患者病历，3 千万份影像数据以及 2 亿份保险记录，数据总量超过 60 万 TB，覆盖人数约 3 亿，然而这些是在花费了 40 亿美元的基础之上获得的。谷歌公司也在英国国家医疗服务系统 (National Health Service, NHS) 的帮助之下获得了 160 万患者的健康数据，但由于医疗数据涉及到患者的隐私，因此引起来了很多争议，数据收集也变的敏感而困难。潘玉颖<sup>[3]</sup>提出文本挖掘、文本搜索及 Python 技术是解决数据难题必备

的关键技术，并应提出矩阵分析方法构建“药物组合 = f(症状组合)”的函数式，建立中医数字化模型，利用人工智能构建函数 f，证素赋值法构建药物组合和症状组合，模糊数学方法赋值证药信息，从而构建适应中医非线性对应的数理模型，成功实现了在 tensorflow 系统环境下输入症状可产生处方的效果。

### 四、讨论

中医学讲求天人合一的整体原则，研究对象始终是人这个整体，而非人的病，因此中医药构建的是关系本体而非实体本体，面对这样一个复杂的巨系统，我们要处理其中诸元素的相关性；其次我们需要将研究的对象置于自然、社会、心理的不同环境之中，从整体出发去考虑问题，因此发展出非线性的辩证体系。最后，将生命科学与人文科学有机地结合起来也是中医药治疗疾病的重要方法。

大数据具有全体性、混杂性和相关性，同时，大数据的思维也是非线性的。医学领域的大数据提供了研究对象在社会中各个维度的数据，所以在理想的情况下，一个从社会到身体再到心理的全方位的分析无疑是生命科学与人文科学的结合。

中医药大数据成果的产出不仅需要互联网技术和人工智能方面的人才，在中医数字化模型的构建中，中医学、统计学、数学等相关专业知识也同样重要，因此同时具备这些专业素养的综合性人才的培养是中医药大数据获得成果的重要一步。同时，如何安全管理这些数据，防止个人隐私数据泄露又能合理利用，也是需要严肃对待的问题。

### 参考文献

- [1] 林树元, 瞿溢谦, 刘畅, 李煜, 曹灵勇. 中医药人工智能发展述评及技术融合趋势探讨. 中华中医药杂志 2020; 35(11):53854-53859.
- [2] 柴华, 路海明, 刘清晨. 中医自然语言处理研究方法综述. 医学信息学杂志 2015;36(10):58-63.
- [3] 潘玉颖, 崔伟锋, 范军铭. 中医药大数据应用核心问题探究. 中医学报 2020;35(05):928-930.