

On the Significance of the Method of Multi-Source Heterogeneous Data Fusion in TCM

Hanqing Zhao, Zhiguo Wang*

Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing
Email: zhhqtt@163.com, `biaozhun432@126.com

Received: Aug. 18th, 2018; accepted: Aug. 26th, 2018; published: Sep. 3rd, 2018

Abstract

Multi-source isomerism is one of the basic features of large data. It is a hot issue in recent years to study traditional Chinese medicine diagnosis and treatment methods based on data. Building a generalization model is one of the methods to solve multisource heterogeneous data fusion and shares and extends the scope of traditional Chinese medicine data. However, the complexity of the large data of traditional Chinese medicine is high. Many problems, such as rich semantics, uneven distribution and poor objectivity, have greatly restricted the research and application of big data in Chinese medicine. In this paper, the importance of multi-source heterogeneous data fusion method under the background of Internet+ large data is discussed, and the importance of multi-source heterogeneous data fusion method based on the combination of disease and syndrome is discussed. It is the original cause of the important component of the large data of traditional Chinese medicine in the future, and the further study of the multi-source isomerism of traditional Chinese medicine. The method of large data fusion provides a theoretical reference.

Keywords

TCM Informatization, Diagnosis and Treatment Mode, Combination of Disease and Syndrome, Big Data, Multi-Source Heterogeneous Fusion

论中医药多源异构大数据融合方法研究的意义

赵汉青, 王志国*

中国中医科学院中医临床基础医学研究所, 北京
Email: zhhqtt@163.com, `biaozhun432@126.com

收稿日期: 2018年8月18日; 录用日期: 2018年8月26日; 发布日期: 2018年9月3日

*通讯作者。

摘要

多源异构是大数据的基本特征之一, 以数据为导向研究中医诊疗方法是近年来的热点问题, 构建泛化模型是解决多源异构数据融合共享、扩展中医数据研究范围的方法之一, 然而由于中医药多源异构大数据具有复杂度高、语义丰富、分布不均、客观化程度差等诸多问题, 极大地制约了中医药大数据的研究和应用。本文以多源异构大数据模型为主要切入点, 讨论了互联网+大数据背景下中医多源异构数据融合方法的重要性, 阐述了基于病证结合的中医药多源异构数据是未来中医药大数据重要构成部分的原因, 为进一步研究中医药多源异构大数据融合方法提供了理论参考。

关键词

中医信息化, 诊疗模式, 病证结合, 大数据, 多源异构融合

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

多源异构是大数据的基本特征之一[1], 由于数据来源不同、数据结构不同、数据标准不同, 多源异构数据在类型、性质、形式、内容和时空关系上均存在极大的差异[2] [3] [4] [5], 容易导致数据信息共享和互联互通障碍。近年来, 在医学[6]、工业、农业、气象、海洋、地质等多个领域均对其相关的多源异构大数据的融合分析问题进行了研究, 而在中医领域针对此问题的研究略显不足。

2. 中医数理化研究已陷入瓶颈

自 20 世纪 80 年代以来, 国内开始对中医辨证论治信息化等数理问题进行研究, 并从模糊数学、自动向量机、人工神经网络、概率论等数理角度解析中医理论, 开始尝试对中医要素进行量化分析, 产生了阴阳五行数学、中医专家系统及各类中医辩证思想的数学模型[7]。

进入 90 年代后, 由于信息技术的发展, 国内外运用包括聚类分析、关联分析、贝叶斯网络、回归和判别分析、文本挖掘、因子分析和主成分分析、隐结构模型、马尔科夫模型、主题模型、信息熵等多种数据挖掘技术对中医辨证论治规律的发现和应用开展了广泛的研究, 发现或验证了许多理论规律, 例如挖掘整理名老中医经验等, 至今仍为研究热点[8] [9] [10]。

迈入 21 世纪时, 正逢互联网技术快速发展, 卫生信息化工作迅速开展, 大型医院均装备有医院信息系统, 中医临床诊疗数据基本上实现了信息化存储, 同时国内学者建设了多个大型中医药数据库、中医临床科研共享系统等中医药网络数据平台, 集合了中医电子病历、中医药文献或其他专题数据, 使中医药数据资源得到极大丰富, 并在此基础之上探索中医证、治、法、药或其他中医理论规律, 提出了多种中医诊断模型, 设计了多种中医辩证人工智能算法[11], 在中医信息化、数字化、智能化研究特别是中医临床辅助支持研究方面取得了许多有重要意义的成果。

然而, 由于近几年云计算、大数据技术的高速发展以及精准医学概念的提出, 大规模的数据显然不适合采用人工方法处理, 传统数理分析和数据挖掘方法在中医诊疗模型的研究中表现出的局限性越来越大, 已不能适应当前大数据背景下中医诊疗规律的研究工作。而现代信息技术的进步, 尤其是人工智能

领域的发展, 提出了许多新的理论和新的研究方法, 将其应用在当前大数据背景下的中医多源异构数据分析中, 具备研究的可行性。

3. 未来中医信息主要由病证结合诊疗数据构成

近年来, 病证结合的诊疗模式研究再次成为业内讨论和关注的热点[12], 现代中医已不太可能逃避学习现代医学知识。有学者[13]认为, 随着医学对疾病认识的不断深入, 现代医学辨病不断细化, 病证结合中的辨病也应相应细化, 让病证结合成为中医临床的诊疗模式, 有助于清楚认识疾病的本源和特点, 利于中医处方规律的研究。

中医学内容博大精深, 病证结合的临床诊疗模式需要医生积累大量的中西医知识, 而中医临床实际中面临的庞大疾病谱却又与西医不同, 如何让中医在临床处方过程中得到正确的辅助诊疗信息是当前中医药信息化研究需要解决的重要问题。而在未来大数据收集处理过程中, 医院信息系统中存储的诊疗数据占据了相当一部分, 这部分数据不可避免会混入相当的现代医学信息, 这就造成未来中医药大数据将会包含大量病证结合诊疗数据。

4. 数据融合方法是中医药大数据应用研究的基础

2016年12月6日, 中国国务院发表《中国的中医药》白皮书, 白皮书指出, 中医药发展上升为国家战略, 中医药事业进入新的历史发展时期。随着云计算、大数据、物联网等现代信息科学技术的不断进步, 如何将中医药与信息技术相结合[14]、总结发现中医诊疗规律的信息学原理和应用方法是目前尚未完全解决的理论难题[15]。《中医药信息化发展“十三五”规划》指出, 中医药信息化将进入实现“融入、整合、跨越”的关键时期, 以国家级、省级数据中心为核心的中医药数据平台将不断收集并提供中医药大数据资源, 如何将大规模病证结合中医药多源异构数据进行有效融合[16][17], 是未来进行中医药大数据深层分析应用研究的关键基础问题之一。

5. 中医药多源异构大数据融合方法具有重要的应用价值

笔者认为, 以多源异构中医药数据融合及其支持下的中医处方的数据融合方法为研究对象, 采用定量研究方法, 设计一种泛化的中医处方大数据模型, 使语义丰富的中医数据可以在一定标准下得到存储和识别, 揭示大数据和“互联网+”背景下中医病证结合诊疗规律的新特性, 是当前急需进行的研究课题。

目前中医处方规律的研究多集中在一类病或一类证上, 某些特定的数据可能只适合使用某种特定的方法进行分析, 而大数据具有未分类、高维度的特性, 发现并描述科学合理的中医诊疗规律是当前行业内需要解决的重要科学问题。该问题的成功解决可以为中医药理论和临床的信息学提供一种新的方法, 能为中医“互联网+”服务和中医临床辅助医疗提供理论支撑, 具有重要的理论意义和工程应用价值。

基金项目

本研究得到国家科技部基础性工作专项(2008FY230500)资助。

参考文献

- [1] Jagadish, H.V., Gehrke, J., Labrinidis, A., et al. (2014) Bid Data and Its Technical Challenges. *Communications of the ACM*, 57, 86-94. <https://doi.org/10.1145/2611567>
- [2] Safari, S., Shabani, F. and Simon, D. (2014) Multirate Multisensory Data Fusion for Linear Systems Using Kalman Filters and a Neural Network. *Aerospace Science and Technology*, 39, 465-471. <https://doi.org/10.1016/j.ast.2014.06.005>
- [3] Lin, G.P., Liang, J.Y. and Qian, Y.H. (2015) An Information Fusion Approach by Combining Multigranulation Rough

- Sets and Evidence Theory. *Information Sciences*, **314**, 184-199. <https://doi.org/10.1016/j.ins.2015.03.051>
- [4] Bostrom, H., Andler, S.F., Brohede, M., *et al.* (2007) On the Definition of Information Fusion as a Field of Research. School of Humanities and Informatics, University of Skovde,
- [5] Naumann, F., Bilke, A., Bleiholder, J., *et al.* (2006) Data Fusion in Three Steps: Resolving Inconsistencies at Schema-, Tuple-, and Value-Level. *IEEE Data Engineering Bulletin*, **29**, 21-31.
- [6] Suk, H.-I., Lee, S.-W. and Shen, D.G. (2014) Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis. *Neuroimage*, **101**, 569-582. <https://doi.org/10.1016/j.neuroimage.2014.06.077>
- [7] 蔡庆生, 卢贤若, 黄克诚. 中医数学化初探[J]. 自然杂志, 1982, 5(11): 813-816, 880.
- [8] 杨丽. 基于知识推理的中医临床诊疗决策支持方法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2014.
- [9] 朱志鹏, 杜建强, 刘英锋, 等. 基于 LDA 主题模型的中医药方剂相似度计算[J]. 计算机应用研究, 2017, 34(6): 1-5.
- [10] 洪燕珠, 周昌乐, 张志枫, 等. 基于随机森林法的慢性疲劳证候要素特征症状的选择[J]. 中医杂志, 2010, 51(7): 634-638.
- [11] 陈楚湘. 中医诊疗标准建立及应用的智能方法研究[D]: [博士学位论文]. 郑州: 解放军信息工程大学, 2011.
- [12] 王阶, 熊兴江, 张兰凤. 病证结合模式及临床运用探索[J]. 中国中西医结合杂志, 2012, 32(3): 297-299.
- [13] 陈可冀, 宋军. 病证结合的临床研究是中西医结合研究的重要模式[J]. 世界科学技术——中医药现代化, 2006, 8(2): 1-5.
- [14] 崔蒙, 李海燕, 杨硕, 等. 中医药信息学理论科学领域研究进展[J]. 中国中医药图书情报杂志, 2015, 39(5): 1-6.
- [15] 于琦, 崔蒙, 李海燕, 等. 从大数据角度探讨中医药信息学特征[J]. 中国数字医学, 2014, 9(4): 32-34.
- [16] Yager, R.R. (2004) A Framework for Multi-Source Data Fusion. *Information Sciences*, **163**, 175-200. <https://doi.org/10.1016/j.ins.2003.03.018>
- [17] 化柏林, 李广建. 大数据环境下多源信息融合的理论与应用探讨[J]. 图书情报工作, 2015, 59(16): 5-10.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2166-6067, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: tcm@hanspub.org