

doi:10.3969/j.issn.1673-9833.2025.03.010

基于证据理论的在线健康社区医生 回答群决策方法研究

邹筱¹, 刘垣春¹, 周欢¹, 袁义²

(1. 湖南工业大学 经济与管理学院, 湖南 株洲 412007; 2. 湖南工业大学 计算机学院, 湖南 株洲 412007)

摘要: 在复杂群决策环境下, 针对在线健康社区中体量庞大且良莠不齐的医生回答择优问题, 基于证据理论, 结合文本分析法与复杂网络分析法, 提出了一种推荐最优医生回答的分析框架。首先, 通过 python 获取在线健康社区中医生的回答数据, 并对数据进行预处理; 其次, 运用 TextRank 主题模型分析医生回答文本中表达的主题, 将提取主题后的医生回答作为证据, 运用证据理论方法对证据间的信任度和冲突度进行测度, 并以此得到证据的初步得分; 再次, 基于回答的冲突因子构建医生关联网, 进而根据网络结构特征确定专家权重; 最后, 结合专家权重重新调整证据初步得分, 得到方案的最终得分, 从而选出最优方案。比较分析结果表明, 经典证据理论决策结果与专家决策结果的一致率为 71.4%, 而本方法决策结果的一致率达 85.7%, 准确率提高了 14.3%。

关键词: 在线健康社区; 证据理论; 回答推荐; 群体决策

中图分类号: TP391.1

文献标志码: A

文章编号: 1673-9833(2025)03-0073-09

引文格式: 邹筱, 刘垣春, 周欢, 等. 基于证据理论的在线健康社区医生回答群决策方法研究 [J]. 湖南工业大学学报, 2025, 39(3): 73-81.

Research on the Decision-Making Method of Online Health Community Doctor Response Group Based on Evidence Theory

ZOU Xiao¹, LIU Yuanchun¹, ZHOU Huan¹, YUAN Yi²

(1. College of Economics and Management, Hunan University of Technology, Zhuzhou Hunan 412007, China;

2. College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: In a complex group decision-making environment, an analysis framework has been proposed based on evidence theory, combined with text analysis and complex network analysis, for a recommendation of the optimal doctor response for the problem of selecting the best answer among doctors with large and uneven quality in online health communities. Firstly, the response data of medical students in online health communities can be obtained through Python and pre-processed data. Secondly, the TextRank topic model is used for an analysis of the topics expressed in the doctor response text, with the extracted topics used as evidence to measure the trust and conflict between the evidence by adopting evidence theory methods, thus obtaining a preliminary score for the evidence based on this. Thirdly, a doctor association network is constructed based on the conflict factors of the responses, with expert weights determined

收稿日期: 2024-04-17

基金项目: 国家自然科学基金资助青年项目 (71801090); 湖南省自然科学基金资助面上项目 (2023JJ30220); 湖南省教育厅科研基金资助重点项目 (23A0440); 湖南省自然科学基金资助项目 (2023JJ50203)

作者简介: 邹筱, 女, 湖南工业大学教授, 硕士生导师, 主要研究方向为社会计算, 物流与供应链管理,
E-mail: 7805463@qq.com

通信作者: 周欢, 女, 湖南工业大学教授, 硕士生导师, 主要研究方向为社会计算, 网络舆情,
E-mail: huanzhou@hut.edu.cn

according to the network structure characteristics. Finally, by adjusting the preliminary score of the evidence based on expert weights, the final score of the plan can be obtained, with the optimal plan selected accordingly. The consistency rate between the decision results of classical evidence theory and expert decision results is 71.4%. The consistency rate of the decision results of this method reaches 85.7%, and the accuracy has been improved by 14.3%.

Keywords: online health community; evidence theory; response recommendation; group decision making

1 研究背景

随着互联网与数字技术的快速发展,互联网迅速融入各行各业,在促进金融、银行、传统制造业、医疗和其他不同行业的发展与数字化转型等方面发挥着重要作用。在医疗健康领域,“互联网+医疗”模式不断深入发展。而在线健康平台作为“互联网+医疗”模式的产物,也得到了迅速发展。据第53次《中国互联网络发展状况统计报告》显示,截至2023年12月,我国网民规模达10.92亿,其中互联网医疗用户规模达4.14亿,占网民整体的37.9%,用户规模和使用率都呈现出上升的趋势^[1]。

用户规模不断扩大,吸引了越来越多的医生专家在在线健康平台上注册账号,为用户答疑解惑。社区中逐渐积累了大量的健康问诊数据,相较于以往用户发布问题后需耐心等待医生回答的问诊方式,用户更倾向于先浏览社区中已有类似疾病问题下的医生回答,以期更快速地解决自身健康问题^[2]。值得重视的是,医生针对患者提问所给出的回答方案的优劣,可以直接或间接地影响患者后续的就诊意向和治疗效果。因此,如何对海量的医生回答方案进行分析处理,识别出最优治疗方案,以协助在线健康社区(online health community, OHC)不断优化服务、满足用户医疗健康需求、促进“平台-医生-用户”在线健康咨询生态系统协同发展成为了学者们关注的焦点。

近年来,随着在线健康社区不断发展成熟,在线健康社区中的医生回答方案文本数据也急剧膨胀,针对同一患者提问,不同的医生由于个人知识结构和能力水平的差异,有时会出现见解不同甚至意见截然相反的情况。在线健康社区中医生回答的信息质量良莠不齐^[3],而普通用户缺乏专业医疗知识、信息素养不高,所以用户对平台中健康信息的辨识能力较差,易产生健康信息焦虑^[2]。证据理论(evidential theory, ET)在融合具有不确定性甚至冲突性的群决策问题的可信度上具有不错的效果,其核心思想是将问题和证据细化成小的问题和证据,根据证据间的相似程度,然后通过证据理论的相关规则,计算出证据可信度,并对方案进行优劣排序。借助证据理论,能

够对医生决策群体给出的不确定甚至有冲突的决策方案进行处理分析,已成为该领域重要的研究方法。

本文基于证据理论,综合运用文本分析法和复杂网络分析法,为解决医生回答群决策的择优问题提出一种新的解决方法。并且为了验证方法的可行性,选用具有代表性的“寻医问药”在线健康社区作为实验对象,获取“寻医问药”经典问答板块中的真实医生回答文本数据展开实验。

2 相关研究

2.1 在线健康社区

在线健康社区是一个可以提供医疗健康知识共享、进行专家咨询、沟通交流的网上虚拟社区^[4]。近年来,关于在线健康社区的研究集中于社区用户行为和与信息内容分析与治理两个方面。

在社区用户行为研究方面,学者对用户行为的自身特点和其影响因素方面展开了深入研究。用户行为具体包括医生贡献行为^[5]、知识共享行为^[6]、参与行为^[7]、信息搜寻行为^[8]、信息使用行为^[9]及隐私披露行为^[10]等,通过分析用户的行为自身特点,从而提高平台和管理人员对用户的认识。进一步探析影响用户行为的因素有哪些,辨析其中的关键因素,这将有利于有重点、有针对性地优化社区服务水平和质量。

在社区信息内容分析与治理方面,不少学者运用文本分析法、情感分析法以及复杂网络等方法,利用在线健康社区中的在线患者评论^[11]、医患问答交互数据^[12]、电子病历数据^[13]等网络健康数据展开实验,研究成果对于治理网络虚假健康信息^[14],改善在线健康社区中的信息质量,提高在线健康社区的知识管理水平等具有积极的促进作用。

2.2 群体决策

群体决策(group decision making, GDM)是指针对某一较为复杂的决策问题,决策群体根据某种科学的评价标准,给出自己的决策意见,然后再依据某种规则对决策群体所给出的全部决策意见进行融合处理,最终形成群体共识意见,或对备选方案进行优劣排序以选出最优方案^[15]。群体决策的优势在于可以充分发挥集体智慧,凸显各个专家的优势,

取长补短, 有效弥补单个决策者由于自身专业知识的不足、专业能力的欠缺或环境资源信息的缺失等多方面因素对结果造成的影响, 使决策结果更加科学化、合理化。

郭永辉等^[16]指出, 学者们关于群体决策问题的研究主要集中于评价指标体系的构建、决策者观点表达的形式、权重的确定以及群体决策的综合集成4个方面。而近年来, 通信技术的飞速发展导致群体决策者规模越来越庞大, 决策问题和决策环境都变得越来越复杂化, 这使得获得大规模群体共识意见的难度也大幅提高, 群体决策的综合集成问题, 即共识问题成为解决大规模群体决策问题的关键要素之一^[17]。共识度量是一种基于距离函数的使用, 用来描述决策者之间相似度的函数^[13]。本文利用证据理论测度决策者之间的成对距离, 以表示医生间意见的相似度与冲突度, 距离越近, 专家意见相似度越高。

2.3 证据理论

证据理论最早由 A. P. Dempster^[18] 提出, 后来由 G. Shafer^[19] 对其进一步完善和推广, 故又被称为 D-S 理论。证据理论在解决具有不确定性问题方面具有良好的效果^[20], 是决策领域中被广泛应用的一种理论。证据理论的核心思想就是将复杂的证据和问题细化分解为简单的证据和问题, 再通过证据理论的计算规则得到各条证据的可信度, 按照得分的高低

对决策方案进行排序。因其能较好地处理决策问题的未知性和不确定性而被广泛应用于专家系统^[21]、模式识别^[22]、可靠性分析^[23]和信息融合^[24]等领域。为了得到更合理的决策结果, 现有学者主要从改进融合规则^[25]以及对证据主体进行预处理^[26]两个方面着手。

综上所述, 首先, 已有大量学者在在线健康社区研究领域取得了一定的进展, 奠定了良好基础, 但对在线健康社区的医患问答领域, 尤其是针对于医生回答的“对中择优”问题的研究仍需进一步深入研究。其次, 随着群决策环境越来越复杂化, 通过主观的方式对方案进行择优难度急剧增加。所以, 如何更加客观合理地解决群决策问题成为了学者们研究的重点。而在医疗诊断问题中, 降低不确定性以提高准确性是最关键且最艰巨的问题之一。证据理论方法因其能够有效地处理知识的分类以及灵活地描述不确定性问题, 能有效地对医生回答进行知识发现。

3 基于证据理论的医生回答群决策方法

本文的研究框架如图1所示, 由5大模块构成, 分别为数据获取及预处理、证据集合构建、证据得分计算、医生权重计算以及最优方案选择。

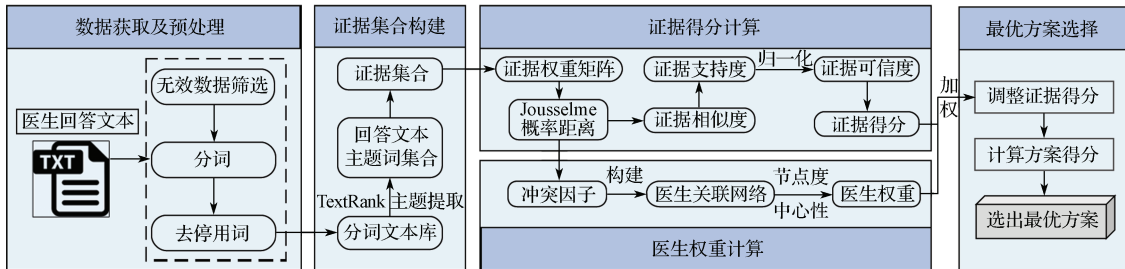


图1 基于证据理论的医生回答群决策方法研究框架

Fig. 1 Research framework of doctor response group decision-making method based on evidence theory

3.1 基于回答主题的证据集合构建

3.1.1 医生回答主题提取

采用基于图模型的方法——TextRank 主题提取, 对医生回答文本进行主题提取。TextRank 算法是一种基于图模型的排序算法。利用 TextRank 算法将文本数据进行分割, 得到若干关键词, 把得到的关键词作为图模型的节点, 再计算出节点之间的相似度, 并将其作为图模型的边, 最后利用节点间相似度对各关键词进行排序, 即与其他节点相似度越高的节点对应的得分越高。

假定医生回答文本是 T , 将 T 分割处理之后得到集合 $T=[S_1, S_2, \dots, S_m]$, 从而可以得到医生回答文本 T

的无向加权图 $G=(V, E, W)$, 其中, V 为所有节点的集合, E 为节点间所有边的集合, W 为每条边所对应权重的集合。图 G 节点间的概率转移矩阵 $SM_{n \times n}$ 为一个对称相似度矩阵:

$$SM_{n \times n} = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{nn} \end{bmatrix}. \quad (1)$$

因为节点 V_i 对 V_j 的贡献度和节点 V_j 对 V_i 的贡献度是一样的, 所以 $SM_{n \times n}$ 对角线上的元素都是 1。

根据无向加权图 G 和图 G 节点间的概率转移矩阵 $SM_{n \times n}$, 通过迭代的方式计算出每个节点的权重, 权重计算如式 (2) 所示:

$$WS(V_i) = (1-d) + d \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} ws(V_j) \quad (2)$$

式中： $WS(V_i)$ 为节点 V_i 的权重值； d 为阻尼系数，通常取值 0.85； V_i 、 V_j 为关键词； w_{ji} 为连接 V_i 和 V_j 的边的权重； $In(V_i)$ 为指向 V_i 的节点集合； $Out(V_i)$ 为 V_i 指向的节点集合。

权重越高的节点的重要性越高，把每条医生回答中重要性较高的节点形成医生回答文本主题集合。

3.1.2 证据集合构建

在本研究中，基于回答主题构建证据集合，将主题提取后的医生回答文本数据作为证据进行下一步实验，一条医生回答文本数据对应一条证据，所有医生的回答即构成群决策问题的证据集合。

3.2 证据得分计算

3.2.1 证据可信度计算

令 $\Theta = \{H_1, H_2, \dots, H_N\}$ 为相互独立且包含所有可能结果的集合，被称为识别框架。基本概率分配 (BPA) 是一个函数 $m: 2^\Theta \rightarrow [0, 1]$ ，也被称为 mass 函数，满足：

$$m(\emptyset) = 0 \text{ 并且 } \sum_{A \in \Theta} m(A) = 1。$$

式中： \emptyset 为空集； A 为 Θ 的子集，且 2^Θ 是 Θ 的幂集，包含 Θ 所有子集，即

$$2^\Theta = \{\emptyset, \{H_1\}, \{H_2\}, \dots, \{H_N\}, \{H_1 \cup H_2\}, \dots, \{H_1 \cup H_N\}, \dots, \Theta\}。$$

基本概率分配 $m(A)$ 测定的是精确分配给 A 的信度，表示证据直接支持 A 为真的强度。所有命题的基本概率之和为 1，空集不产生任何信度，分配 Θ 的概率，即 $m(\Theta)$ ，被称为未知度。任意子集 $A \subseteq \Theta$ ，若 $m(A) > 0$ ，则 A 被称为 m 的焦点，所有焦点的集合称为核，所有相关的焦点元素统称为证据主体。

利用证据之间的 Josselme 距离来衡量证据之间的冲突度，证据之间的 Josselme 距离值越大，表示证据之间的冲突度越高。具体计算公式如下：

$$D(d_a, d_b) = \sqrt{\frac{1}{2} (\|d_a\|^2 + \|d_b\|^2 - 2\langle d_a, d_b \rangle)} \quad (3)$$

(a, b = 1, 2, \dots, n)。

式中： $D(d_a, d_b) \in [0, 1]$ ； $\|d_a\|^2 = \langle d_a, d_a \rangle$ ；

$$\|d_b\|^2 = \langle d_b, d_b \rangle；$$

$$\langle d_a, d_b \rangle = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} d_a(e_i) d_b(e_j) \frac{|e_i \cap e_j|}{|e_i \cup e_j|}, e_i, e_j \in R(\theta) \quad (4)$$

则由 d_a 、 d_b 之间的 Josselme 距离所构成的证据距离矩阵为

$$D_j = \begin{bmatrix} 0 & D(d_1, d_2) & \dots & D(d_1, d_{n-1}) & D(d_1, d_n) \\ D(d_2, d_1) & 0 & \dots & D(d_2, d_{n-1}) & D(d_2, d_n) \\ \vdots & \vdots & 0 & \vdots & \vdots \\ D(d_{n-1}, d_1) & D(d_{n-1}, d_2) & \dots & 0 & D(d_{n-1}, d_n) \\ D(d_n, d_1) & D(d_n, d_2) & \dots & D(d_n, d_{n-1}) & 0 \end{bmatrix} \quad (5)$$

由于 $D(d_a, d_b)$ 与 $D(d_b, d_a)$ 都是计算证据 d_a 与 d_b 距离之间的 Josselme 距离，所以 D_j 为一个对称矩阵。由此可知，当有 n 条证据进行融合处理时，计算所有证据间的距离需要进行 $n(n-1)/2$ 次。

设 d_a 、 d_b 为识别框架 Θ 上的信念结构， $D(d_a, d_b)$ 为 d_a 和 d_b 之间的 Josselme 距离，则 d_a 和 d_b 之间的相似度为

$$Sim(d_a, d_b) = 1 - D(d_a, d_b) \quad (6)$$

设有 n 个定义在识别框架 Θ 上的信念结构，且设 $Sim(d_a, d_b)$ 为 d_a 和 d_b 之间的相似度，则区间证据 d_a 的支持度为

$$Sup(d_a) = \sum_{b=1, b \neq a}^n Sim(d_a, d_b), a = 1, 2, \dots, n \quad (7)$$

设有 n 个定义在识别框架 Θ 上的信念结构，且 $Sup(d_a)$ 为区间证据 d_a 的支持度，则证据 d_a 的可信度 Crd_a 为

$$Crd_a = \frac{Sup(d_a)}{\sum_{a=1}^n Sup(d_a)}, a = 1, 2, \dots, n \quad (8)$$

3.2.2 得分计算

在本研究中，将主题提取后的医生回答文本作为证据集合，构建证据权重矩阵，利用证据理论实现对主题提取后的医生回答文本间 Josselme 距离进行测度，计算得出每条证据的可信度，并将其作为初步的证据得分。

3.3 医生权重计算

3.3.1 关系网络构建

复杂网络是将复杂系统描述为网络，系统的组成要素描述为节点，要素间的联系描述为边，整个系统是由节点和边组成的网络。图是复杂网络常用的一种表示方法。假设图 $G=(V, E)$ 描述的是一个无向无权的复杂网络的拓扑结构， V 是图 G 的节点集， E 是图 G 的边集，其中， $|V|=n$ 、 $|E|=m$ 分别表示图 G 有 n 个节点、 m 条边。本文将每条证据所对应的医生作为网络节点，以证据间的证据冲突度作为边，构建医生关联网。

3.3.2 网络节点重要性计算

在复杂网络中，节点的度中心性 (degree centrality) 是在网络分析中刻画节点重要性的一个重

要指标, 节点度表示一个节点的邻接节点的数目, 节点 v_i 的度中心性计算公式如下:

$$C_d(v_i) = \sum_{j=1}^n e_{i,j} \quad (9)$$

式中 $e_{i,j}$ 为顶点 v_i 与其它顶点的连接边。

根据医生关联网络的节点度中心性的大小确定专家权重, 节点度中心性越大将被分配更高的专家权重, 反之分配较低的专家权重。其主要思想是: 通常采用少数服从多数的方式进行决策, 采纳与主体意见相同或接近的意见, 拒绝与主体意见相悖的意见。

3.4 回答支持程度计算

专家的意见越统一, 其结果的可靠性往往越强。首先, 基于证据理论计算出证据的可信度和冲突因子, 得到证据初步得分; 然后, 根据构建的医生关联网络确定专家权重, 从而进一步调整初步得分, 进而得出最终方案得分。

4 实验分析

为验证所提方法在处理具有不确定性医生群决策问题的有效性和科学性, 选取在线健康社区中的真实数据展开实验予以验证。

4.1 数据来源及预处理

4.1.1 数据来源

本研究运用 python 从寻医问药 (<https://www.xywy.com/>) 平台爬取实验所需数据。截至 2022 年底, 寻医问药平台主要为在线健康社区中的用户提供寻医、问药、查资讯等服务, 功能全面、成功就诊的患者有上百万, 是全国有代表性的在线健康平台之一, 故选用该平台获取实验数据。

已有研究表明, 若针对在线健康问答社区开展研究, 将研究数据限制在一个特定疾病领域, 甚至一个特定的类别, 可以避免来自其他类别的干扰。所以通过分析寻医问药平台的网页布局, 获取医生基于相同疾病与主题下的患者提问数据, 并收集各个问题下 30 名医生回答的正文、是否采纳情况、已帮助用户数、点赞量、关联问题等相关数据。

4.1.2 数据预处理结果

由于获取的数据为不易被后续处理的非结构化文本数据, 故对所获取的医生回答集进行数据预处理, 以获取更易处理的结构化数据。首先, 对获取的数据进行规范化, 转化为表 1 所示的结构化形式 (为保护隐私, 本文将不展示医生姓名); 然后, 为了保护患者隐私, 数据需要被匿名化, 但因为患者的姓名已被平台进行保护, 所以无需此操作; 最后,

去除患者提问数据和医生介绍数据等次要信息, 并利用 jieba 中文分词工具对医生回答数据进行处理, 包括清洗数据、文本分词和去停用词等。

表 1 爬取数据样例

Table 1 Crawling data samples

编号	正文	是否采纳	关联问题
医生 01	脑血管硬化病人需要控制血压血脂与血糖, …… , 清除斑块, 有时还需降脂溶栓治疗。	是	1
医生 02	治疗脑血管硬化, …… , 治疗时摒弃了西医的副作用和依赖性。	否	1
⋮	⋮	⋮	⋮
医生 30	脑血管硬化患者可以采取物理加中医的治疗方法, …… , 治疗时摒弃了西医的副作用和依赖性。	否	21

获取寻医问药平台的医生回答数据后, 预处理后的部分医生回答数据如表 2 所示。

表 2 预处理后医生回答数据示例

Table 2 Samples of a doctor response data after preprocessing

编号	正文
医生 01	['脑血管','硬化','病人','控制','血压','血脂','血糖','检测','发现','早防','早治','心态','平和','生活','规律','低盐','低脂','清淡','饮食','多喝水','少','烟酒','吃','蔬菜水果','减肥','运动','锻炼','治疗','情况','采用','动脉','内膜','剥离','手术','清除','斑块','降脂','溶栓','治疗']
医生 02	['治疗','脑血管','硬化','患者','物理','加','中医','治疗','方法','中医','理论','现代化','治疗','仪器','作用','治疗','方法','患者','恢复','治疗','摒弃','西医','副作用','依赖性']
⋮	⋮
医生 30	['脑血管','硬化','患者','物理','加','中医','治疗','方法','中医','理论','现代化','治疗','仪器','作用','治疗','方法','患者','恢复','治疗','摒弃','西医','副作用','依赖性']

4.2 实验步骤与结果分析

4.2.1 证据集合

通过 TextRank 主题模型对预处理后的医生回答数据进行主题提取。利用 jieba 工具计算词项的 TextRank 值, 截取其中排名前 10 的词项和权重 (保留 4 位小数), 如表 3 所示。然后进一步对获取的主题词进行归一化处理, 处理结果如表 4 所示, 并将其作为识别框架, 将经过主题提取后的医生回答作为证据进行下一步实验。

利用 python 工具, 基于分词文本库, 采用 TextRank 绘制的医生回答文本关键词的词云图 (见图 2), 图中显示靠前 50 个关键词。根据词云图初步可知, 在脑血管硬化的治疗过程中, 首先, 病人需

要高度重视对于血糖、血压以及血脂的检测与监控；其次，病人也要积极配合医生的治疗，采用药物治疗，听取医生建议；最后，病人需要养成健康的生活习惯，例如适当锻炼、多喝水、清淡饮食、规律作息以及保持平和的心态等。

表 3 医生回答 TextRank 主题词提取示例

Table 3 TextRank subject words extraction example for doctor responses

编号	医生回答主题词
医生 1	{'治疗': 1.0, '血压': 0.8506, '血脂': 0.7525, '控制': 0.7340, '情况': 0.6806, '采用': 0.6725, '清除': 0.6651, '手术': 0.6629, '运动': 0.6620, '动脉': 0.6577}
医生 2	{'治疗': 1.0, '方法': 0.4283, '患者': 0.3851, '作用': 0.2549, '仪器': 0.2511, '恢复': 0.2375, '现代化': 0.2248, '摒弃': 0.2219, '物理': 0.2118, '中医理论': 0.1950}
⋮	⋮
医生 30	{'治疗': 1.0, '方法': 0.4938, '患者': 0.4331, '作用': 0.2909, '仪器': 0.2868, '恢复': 0.2708, '现代化': 0.2544, '摒弃': 0.2523, '物理': 0.2319, '中医理论': 0.2204, '副作用': 0.1820, '脑血管': 0.1635, '硬化': 0.1631, '依赖性': 0.1439}

表 4 医生回答 TextRank 主题词归一化示例

Table 4 TextRank topic normalization example for doctor responses

编号	医生回答主题词归一化
医生 1	{'治疗': 0.1363, '血压': 0.1159, '血脂': 0.1026, '控制': 0.1000, '情况': 0.0927, '采用': 0.0917, '清除': 0.0906, '手术': 0.0903, '运动': 0.0902, '动脉': 0.0896}
医生 2	{'治疗': 0.2932, '方法': 0.1256, '患者': 0.1129, '作用': 0.0747, '仪器': 0.0736, '恢复': 0.0696, '现代化': 0.0659, '摒弃': 0.0651, '物理': 0.0621, '中医理论': 0.0572}
⋮	⋮
医生 30	{'治疗': 0.2280, '方法': 0.11261, '患者': 0.0987, '作用': 0.0663, '仪器': 0.0654, '恢复': 0.06173, '现代化': 0.0580, '摒弃': 0.0575, '物理': 0.0529, '中医理论': 0.0502, '副作用': 0.0415, '脑血管': 0.0373, '硬化': 0.0372, '依赖性': 0.0328}



图 2 医生回答关键词词云图

Fig. 2 Keyword word cloud for doctor responses

4.2.2 证据得分

根据式 (3) (4) (6)~(8) 计算得出证据的

可信度，结果如图 3 所示。由图可知，最小的证据得分为医生 28，这表明该医生给出的回答与其他医生的观点冲突程度最高，故该条证据被赋予的得分最小；反之，与其他医生观点相似度更高的医生 21 的证据被赋予更高得分。

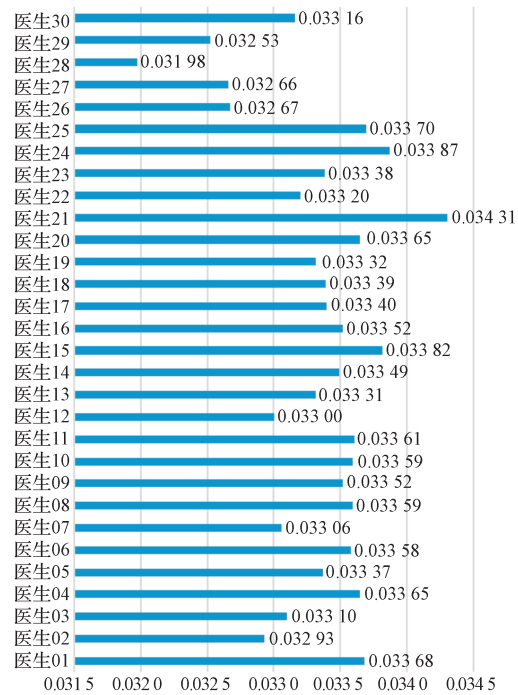


图 3 基于证据理论的证据得分

Fig. 3 Evidence score based on evidence theory

4.2.3 医生权重

根据证据理论计算得证据之间的 Joussemle 距离，以此作为证据间冲突因子构成证据冲突因子矩阵，如表 5 所示。数值越大意味着证据之间的 Joussemle 距离越远，表示此医生的意见与其他医生的意见相差较大，即冲突程度越大，反之，冲突程度越小。

表 5 证据冲突因子矩阵

Table 5 Evidence conflict factor matrix

	医生 1	医生 2	⋮	医生 29	医生 30
医生 1	0	0.290 0	⋮	0.318 4	0.287 7
医生 2	0.290 0	0	⋮	0.351 2	0.019 3
⋮	⋮	⋮	0	⋮	⋮
医生 29	0.318 4	0.351 2	⋮	0	0.344 3
医生 30	0.287 7	0.019 3	⋮	0.344 3	0

根据表 5 的结果，以医生作为网络节点、以证据冲突因子的值作为网络的边来构建医生关联网络，且为了使所构建的医生决策网络中既包含所有专家，又要避免回答冲突程度过高的医生之间建立连边，本文通过多次调试后，最终将网络的阈值设置为 0.33。构建出医生关联网络，如图 4 所示，其中，颜色越深、节点越大表示该节点的节点度中心性越大，其重要性

越高;反之越低。得出医生关联网络后,基于医生关联网络的节点度中心性赋予专家权重值。

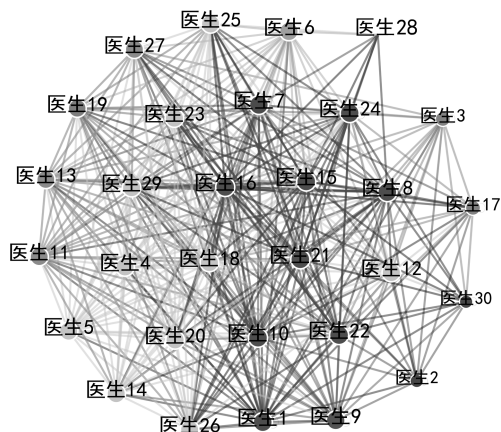


图4 医生关联网络

Fig. 4 Doctor association network

4.2.4 回答支持程度

确定证据得分与医生权重后,通过将两者加权的方式对证据的初步得分作进一步调整,得出最终的方案得分,得分最高的方案即为群决策问题中的最优方案。调整专家权重后医生回答得分的结果如图5所示。

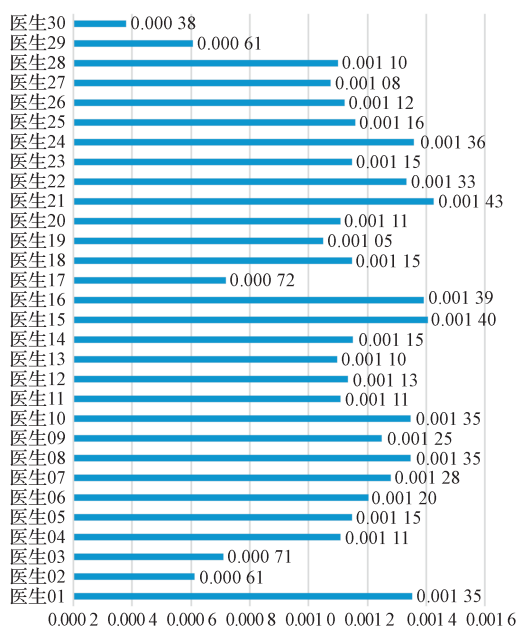


图5 基于证据理论和医生关联网络的证据得分

Fig. 5 Evidence score based on evidence theory and doctor association network

4.2.5 最优回答

通过实验得出的回答支持信度最高为医生21的回答,最终得分约为0.00143,对应的最优回答内容为“脑动脉硬化是全身动脉硬化的一部分,西医治疗本病多采用降低血脂、扩张血管、改善精神症状类药物,提倡低脂、低盐、多碘食物。一般治疗:应注意劳逸结合,生活有规律、避免情绪激动和进行适度的

体育锻炼。”关于心脑血管硬化等疾病,医生强调患者需要注意血压血脂、药物治疗、规律饮食和作息以及适量运动等注意事项,实验结果与前文中根据所有医生回答所得出的词云图中重点词具有一致性,即实验得出的最优医生回答具有代表性和可靠性,与实际情况相符。

4.3 比较分析

为了验证本文方法的普适性和有效性,爬取了脑动脉硬化、肝硬化以及高血压等不同疾病相同或相似问题下的医生回答开展实验,并将本文方法得到结果与经典证据理论方法^[9]进行对比分析。评判医生回答是否有效往往需要用户认可或权威专家认可,其中用户认可度可以通过用户点赞或转发等互动行为来判断,但鉴于在线健康社区中这部分数据存在缺失,因此通过对中南大学湘雅医院与株洲市中心医院相应科室的权威专家进行访谈予以判断,最后将本文决策结果与专家决策结果进行一致性检验,决策结果准确率,并对比不同方法的结果准确率,实验结果如表6所示。

表6 改进算法组合结果比较

Table 6 Comparison of improved algorithm combination results

疾病	医院	专家	决策结果		
			专家	本方法	经典证据理论
脑动脉硬化	中南大学湘雅医院	1			
		2	21号	21号	21号
	株洲市中心医院	3			
		4			
肝硬化	株洲市中心医院	1			
		2	3号	3号	3号
	中南大学湘雅医院	3			
		4			
高血压	中南大学湘雅医院	1			
		2	15号	15号	7号
	株洲市中心医院	3			
		4			
尿毒症	株洲市中心医院	1	7号		
		2			
	株洲市中心医院	3	8号	8号	8号
		4			
胃炎	株洲市中心医院	1			
		2	14号	14号	14号
	中南大学湘雅医院	3			
		4	14号和7号		
中风	中南大学湘雅医院	1			
		2	1号	1号	1号
	株洲市中心医院	3			
		4			
		5	15号		

由表6中数据可知,经典证据理论决策结果与专家决策结果的一致率为71.4%,本方法决策结果的一致率达85.7%,准确率提高了14.3%,产生这种结果的原因可能是当专家意见出现较大差异时,经典证据理论方法融合结果有时会出现违背常理的情况,即Dempster组合规则在冲突证据面前存在明显缺陷。而本文方法基于医生回答文本间的客观联系计算证据冲突度和可信度,再据此构建医生联系网络的网络结构特征调整专家权重,这样以最原始的文本数据为实验对象开展实验,极大程度克服了经典证据理论用专家对方案打分作为实验数据可能带来的人为因素的影响,实验结果更为客观。

5 结语

本文提出了一种新的获取最优决策方案分析框架,基于证据理论,将文本分析法与复杂网络分析方法相结合,研究在线健康社区中对医生回答的群决策择优问题。实验结果符合少数服从多数的原则,即最优医生回答是更加全面、可靠且有指导性的医生回答。一方面,将复杂网络分析方法应用于医生回答优越性分析,拓展了复杂网络分析方法的应用范围;另一方面,基于证据理论获取证据得分,然后基于医生关联网络的网络特征获取专家权重,对证据得分调整后得到最终的证据得分,为患者筛选出最优的医生回答,不仅提高了患者决策效率和质量,更好地满足用户的医疗健康需求,有助于在线健康社区的平台管理人员有效管理社区信息,也为大规模的群体决策问题提供了新的解决思路。未来,将考虑与智能化方法结合,实现医生回答优越性的自动判断,并为用户的就诊决策提供高效的辅助。

参考文献:

- [1] 中国互联网络信息中心. 第53次《中国互联网络发展状况统计报告》: 互联网发展研究 [EB/OL]. [2023-11-07]. <https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html>.
China Internet Network Information Center. The 53rd Statistical Report on Internet Development in China: Research on Internet Development [EB/OL]. [2023-11-07]. <https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html>.
- [2] 林萍, 吕健超. 基于Stacking集成学习的在线健康社区问答信息采纳识别研究 [J]. 情报科学, 2023, 41(2): 135-142.
LIN Ping, LÜ Jianchao. Information Adoption Recognition in Online Question and Answer Health Communities Based on Stacking Ensemble Strategy [J]. Information Science, 2023, 41(2): 135-142.
- [3] 成全, 张燕刚. 面向用户需求主题的在线问答社区信息多层级分类研究 [J]. 情报学报, 2022, 41(8): 860-871.
CHENG Quan, ZHANG Yangang. Information Multi-Hierarchical Classification of Online Q & a Community Oriented to Users' Needs Topics [J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(8): 860-871.
- [4] 赵栋祥. 国内在线健康社区研究现状综述 [J]. 图书情报工作, 2018, 62(9): 134-142.
ZHAO Dongxiang. Review on Domestic Research Status of Online Health Community [J]. Library and Information Service, 2018, 62(9): 134-142.
- [5] 王盼盼, 吴志艳, 罗继锋. 有偿奖励对医生在线健康社区中贡献行为的影响 [J]. 系统管理学报, 2022, 31(2): 343-352.
WANG Panpan, WU Zhiyan, LUO Jifeng. Effect of Monetary Incentive on Physicians' Contribution Behavior in Online Healthcare Community [J]. Journal of Systems & Management, 2022, 31(2): 343-352.
- [6] 张星, 吴忧, 夏火松, 等. 基于S-O-R模型的在线健康社区知识共享行为影响因素研究 [J]. 现代情报, 2018, 38(8): 18-26.
ZHANG Xing, WU You, XIA Huosong, et al. A Study of Influencing Factors of Knowledge Sharing Behavior in Online Health Communities Based on S-O-R Model [J]. Journal of Modern Information, 2018, 38(8): 18-26.
- [7] 周涛, 王盈颖, 邓胜利. 基于社会资本理论的在线健康社区用户参与行为研究 [J]. 信息资源管理学报, 2020, 10(2): 59-67, 129.
ZHOU Tao, WANG Yingying, DENG Shengli. Research on Online Health Community Users' Participation Based on Social Capital Theory [J]. Journal of Information Resources Management, 2020, 10(2): 59-67, 129.
- [8] 朱云琴. 双路径视角下在线健康社区信息搜寻行为影响因素研究 [D]. 昆明: 昆明理工大学, 2021.
ZHU Yunqin. Research on Influencing Factors of Users' Health Information Seeking Behavior in Online Health Community Based on the Perspective of Dual Paths [D]. Kunming: Kunming University of Science and Technology, 2021.
- [9] 刘婵君, 贾晓萌, 林泳淇. 国外老年用户网络健康信息使用行为研究: 理论基础、影响因素与未来启示 [J]. 情报杂志, 2021, 40(6): 178-186.
LIU Chanjun, JIA Xiaomeng, LIN Yongqi. A Research on Internet Health Information Use Behavior of Elderly Users Abroad: Theoretical Basis, Influencing Factors

- and Future Enlightenment[J]. *Journal of Intelligence*, 2021, 40(6): 178–186.
- [10] 王瑜超. 在线医疗社区用户健康隐私信息披露意愿的影响因素研究[J]. *信息资源管理学报*, 2018, 8(1): 93–103, 113.
WANG Yuchao. Research on the Influencing Factors of Users' Health Information Disclosure Intention in Online Medical Community[J]. *Journal of Information Resources Management*, 2018, 8(1): 93–103, 113.
- [11] 叶 艳, 吴 鹏, 周 知, 等. 基于LDA-BiLSTM模型的在线医疗服务质量识别研究[J]. *情报理论与实践*, 2022, 45(8): 178–183, 168.
YE Yan, WU Peng, ZHOU Zhi, et al. Research on Online Medical Service Quality Identification Based on LDA-BiLSTM Model[J]. *Information Studies (Theory & Application)*, 2022, 45(8): 178–183, 168.
- [12] LIN C Y, WU Y H, CHEN A L P. Selecting the Most Helpful Answers in Online Health Question Answering Communities[J]. *Journal of Intelligent Information Systems*, 2021, 57(2): 271–293.
- [13] 杨鑫禹, 牟冬梅, 彭 浩, 等. 基于数据特征的电子病历数据驱动临床决策模型研究[J]. *情报理论与实践*, 2022, 45(5): 181–188.
YANG Xinyu, MOU Dongmei, PENG Hao, et al. Research on Clinical Decision Model Driven by Electronic Medical Records Based on Data Characteristics[J]. *Information Studies (Theory & Application)*, 2022, 45(5): 181–188.
- [14] 邓胜利, 顾一飞. 网络虚假健康信息研究综述: 认知、行为与治理[J]. *图书馆杂志*, 2022, 41(5): 14–22.
DENG Shengli, GU Yifei. A Review of Online Health Misinformation: Recognition, Action and Governance[J]. *Library Journal*, 2022, 41(5): 14–22.
- [15] COOK W D, KRESS M. Ordinal Ranking with Intensity of Preference[J]. *Management Science*, 1985, 31(1): 26–32.
- [16] 郭永辉, 尚战伟, 邹俊国, 等. 群决策关键问题研究综述[J]. *统计与决策*, 2016, 32(24): 63–67.
GUO Yonghui, SHANG Zhanwei, ZOU Junguo, et al. Summary of Research on Key Problems of Group Decision Making[J]. *Statistics & Decision*, 2016, 32(24): 63–67.
- [17] 钟晓芳, 兰红娟, 江文奇. 共识驱动的区间直觉模糊型多准则群体决策信息融合模型[J]. *系统工程与电子技术*, 2020, 42(7): 1558–1566.
ZHONG Xiaofang, LAN Hongjuan, JIANG Wenqi. Consensus Driven Information Fusion Model of Interval-Valued Intuitionistic Fuzzy Multi-Criteria Group Decision Making[J]. *Systems Engineering and Electronics*, 2020, 42(7): 1558–1566.
- [18] DEMPSTER A P. Upper and Lower Probabilities Induced by a Multivalued Mapping[J]. *The Annals of Mathematical Statistics*, 1967, 38(2): 325–339.
- [19] SHAFER G. A Mathematical Theory of Evidence[M]. Princeton: Princeton University Press, 1976: 35–37.
- [20] 熊 山. 基于证据理论的不完整信息融合研究[D]. 重庆: 西南大学, 2022.
XIONG Shan. Research on Incomplete Information Fusion Based on Evidence Theory[D]. Chongqing: Southwest University, 2022.
- [21] 王金波, 吴伟志. 基于证据理论的广义多尺度覆盖决策系统的最优尺度组合[J]. *模式识别与人工智能*, 2022, 35(4): 291–305.
WANG Jinbo, WU Weizhi. Evidence-Theory-Based Optimal Scale Combinations in Generalized Multi-Scale Covering Decision Systems[J]. *Pattern Recognition and Artificial Intelligence*, 2022, 35(4): 291–305.
- [22] 余付平, 黄益恒, 沈 堤, 等. D-S证据理论在空中目标识别中的应用现状与展望[J]. *电光与控制*, 2024, 3(4): 75–86.
YU Fuping, HUANG Yiheng, SHEN Di, et al. Application Status and Prospect of D-S Evidence Theory in Aerial Target Identification[J]. *Electronics Optics & Control*, 2024, 3(4): 75–86.
- [23] YU Y, LI Y, XIAO F, et al. Application of D-S Evidence Theory in Marine Environmental Security Posture Assessment[J]. *E3S Web of Conferences*, 2023, 370: 01003.
- [24] 刘 康, 何明浩, 韩 俊, 等. 基于本征向量和Jousselme距离的高冲突证据融合方法[J]. *系统工程与电子技术*, 2022, 44(7): 2175–2180.
LIU Kang, HE Minghao, HAN Jun, et al. A High Conflict Evidence Fusion Method Based on Eigenvector and Jousselme Distance[J]. *Systems Engineering and Electronics*, 2022, 44(7): 2175–2180.
- [25] LEFÈVRE E, ELOUEDI Z. How to Preserve the Conflict as an Alarm in the Combination of Belief Functions?[J]. *Decision Support Systems*, 2013, 56: 326–333.
- [26] YUAN K J, XIAO F Y, FEI L G, et al. Conflict Management Based on Belief Function Entropy in Sensor Fusion[J]. *SpringerPlus*, 2016, 5(1): 638.

(责任编辑: 申 剑)