

doi:10.3772/j.issn.1000-0135.2015.007.008

应急参考咨询团队构建模式研究¹⁾

叶光辉¹ 李纲¹ 武川²

(1. 武汉大学信息资源研究中心, 武汉 430072; 2. 武汉大学信息管理学院, 武汉 430072)

摘要 基于既有多源专家特征信息融合研究,从单一资源的不同分面出发,统计了每一分面排名的参考指标,据此生成专家分面排名。鉴于分面排名之间可能存在的冲突,结合 D-S 理论和熵运算,设计了专家排名融合的算法,算法核心围绕 mass 值的计算和合成。获取专家综合排名,依托群体决策需求,应用加权平均法生成 Top N 位专家的内容特征,通过二次遴选完成应急参考咨询团队的构建。以多主题领域的应急专家遴选过程为例,通过评测排名值的重复率和一致率,证实本文方法相较单一方法效率更高。

关键词 应急专家 参考咨询 分面排名 D-S 理论 信息熵

Research on Construction Mode of Emergency Experts Team for Virtual Reference

Ye Guanghui¹, Li Gang¹ and Wu Chuan²

(1. Center for the Studies of Information Resources of Wuhan University, Wuhan 430072;
2. Information Management School of Wuhan University, Wuhan 430072)

Abstract Based on previous research on information fusion for multi-source expert features, this paper has firstly designed reference indicators of different facets, by which expert faceted rankings are obtained. Secondly, taking the conflict given rise to by varied rankings into account, the authors have applied D-S theory and information entropy to expert faceted ranking fusion algorithm, which focuses on the calculation and synthesis of mass values. Thirdly, weighted average method has been used to integrate content features of top N experts required in emergency decision, so emergency experts team for virtual reference can be constructed via selection for two times. Finally, making the selection process of emergency expert from different areas as examples, outcome has demonstrated that new method in this paper is more efficient than single method by the evaluation of ranking values in repetition and consistency.

Keywords emergency expert, virtual reference, faceted ranking, D-S theory, information entropy

1 引言

长期以来,突发事件对公众生命、财产安全以及社会秩序造成了很大危害。凭借较深的理论积淀和

丰富的工作经验,应急参考咨询在突发事件事前、事中及事后处置中发挥着重要作用^[1],集中体现在群体决策过程中。群体决策关联方方面的知识,需要跨主题甚至跨领域的专家参与进来,因此为保证其科学性,应急专家的遴选过程要严格规范,参考咨

收稿日期:2015年1月15日

作者简介:叶光辉,男,1986年生,博士研究生,主要研究方向:信息检索与文本挖掘,竞争情报,E-mail:3879-4081@163.com。李纲,男,1966年生,博士生导师,教授,武汉大学信息资源研究中心副主任,主要研究方向:情报学,竞争情报。武川,男,1989年生,博士研究生,主要研究方向:实体检索。

1) 本文系国家自然科学基金重大项目“智慧城市应急决策情报体系建设研究”(项目编号:13&ZD173)、中央高校基本科研业务费项目“跨学科专家科研团队发现研究”(项目编号:2014104010202)研究成果之一。

询团队的知识结构也要系统合理。

应急参考咨询不同于一般的信息咨询(如图书馆参考咨询、“定题-委托”参考咨询等),其对时间更敏感,要求专家来源领域宽泛,工作流程科学高效。但目前应急参考咨询团队构建却存在诸多问题,集中体现为两点:①团队构建多依赖召集人的社会关系,缺乏客观准确的专家甄选程序。人际关系在社会网络中一直有着十分重要的作用。实际上,团队能否成功构建,很大程度上取决于相关研究领域人与人关系能否发现和成功连接^[2]。为领域内的应急需求组建咨询团队,召集人凭借对领域环境的熟知还能应对。跨领域合作则要求更加合理的知识及人员组成结构,召集人仅凭主观判定,在选人可能会造成一定误差,这些误差源自信息的不对称。因此应急专家咨询团队发现亟待寻找一种客观高效的方式。②专家特征表述模糊,缺乏程式化的描述符号。专家特征的揭示,需要综合考察其工作内容、研究成果、主持或参与项目等关联因素,而对此的描述却缺乏专业的逻辑符号,导致专家特征识别模糊甚至错误,进而出现专家知识结构和应急需求不兼容等后期问题,影响应急决策,造成严重的经济损失、社会影响和人员伤亡^[3]。

针对上述问题,笔者试图寻求一种科学高效的应急专家咨询团队构建方法,即以专家关联学术资源为单一识别对象,从专家个人、专家文本、专家引文三个分面入手,结合专家外部特征和内容特征,运用 D-S 理论和熵运算,设计了以“排名融合-特征融合”为流程的专家遴选方式,寄望提升应急参考咨询团队构建的工作效率。

2 相关研究综述

应急参考咨询团队构建的关键在于专家发现,而关联专家的发现过程等同于专家主题检索的过程。目前用于专家识别的资源多来自学术数据库、Web 页面和社交网络平台^[4]。围绕这些资源,相关专家学者开展了 4 类研究:①基于知识资源的专家识别,如夏立新等^[5]借助 XTM 绘制图书馆专家知识地图来改善专家自动化识别效果;Fang 等^[6]构建专家搜索引擎 FacFinder 来自动化甄选科研团队成员。②基于 Web 资源的专家识别,如巩军等^[7]以维基百科为背景知识构建专家个人知识地图,度量专家的知识构成和研究兴趣。③基于社会网络资源的专家识别,如廖开际等^[8]借助语义网络对专家知识发现

及表示方法进行了研究,有效地解决了专家识别过程中特征揭示简单、推荐结果不准等问题;Lin 等^[9]将文本分析和社会网络分析相结合,构建专家搜索软件 SmallBlue 来辅助企业进行专家特征识别。④综合多种资源的专家识别,如陆伟等^[10]、王曰芬等^[11]通过网络数据库、搜索引擎、专家推荐表等渠道获取专家个人及关系信息,分别完成了专家组织工作和系统构建;Moreira 等^[12]使用 D-S 理论和熵计算从专家作品、引文网络、专家简介获取的专家信息;Liu 等^[13]整合专家个人信息、专家声望和用户问答链接,分别为在线学术社区和 QA 网站设计了专家排名综合算法,并实证各自方法相较单一基准方法更全面准确。

综上分析并结合资源自身特点可知,知识资源揭示能力好但不全面;Web 资源量多但噪声大;社会网络资源融合了关系数据和属性数据,但依赖于专家沟通方式,内容稳定性不强。随着新兴计算机及网络技术的发展,专家传播自身观点的方式正趋于多元化,而不再仅局限于实物型文献和人际网络,基于多种资源融合的专家识别正逐渐成为该主题领域的研究热点,相关方法(如加权平均法、粗糙集、模糊集、D-S 理论、神经网络等)已在模式识别、人工智能、遥感测绘等领域获得广泛应用^[14,15],但大多基于某一平台展开,可移植性不强,不仅如此,多源信息权重设置方法还有待拓展,既要减少人为的经验估计,又要避免客观但过分复杂的算法^[4]。由于资源结构化及存量水平的差异,多源专家特征信息融合不一定要求资源类型一定要多样,从成熟度较高的单类型资源的不同分面去做融合也有很大的拓展空间^[12]。

3 应急参考咨询团队构建过程

从粗糙集角度来讲,无论哪一种形式的多源信息融合都是将相关知识进行分类排序的过程^[16],信息资源及其分面是分类的依据,专家与主题的关联程度是排序的依据,因此分面的数量决定了专家排名列表的数量。由于分面之间不存在优先级设置,属于并列关系,如上处置则在融合过程中极有可能会出现专家分面排名冲突,如两个专家 e_i 和 e_j ,在不同分面下的排名可能是相逆。传统的 CombSUM 算法不能解决上述情形^[12],D-S 理论^[17,18]有助于消除冲突,但又无法客观地分配每一分面下专家排名的可信度,故本文引入熵运算,结合 D-S 理论来实现专

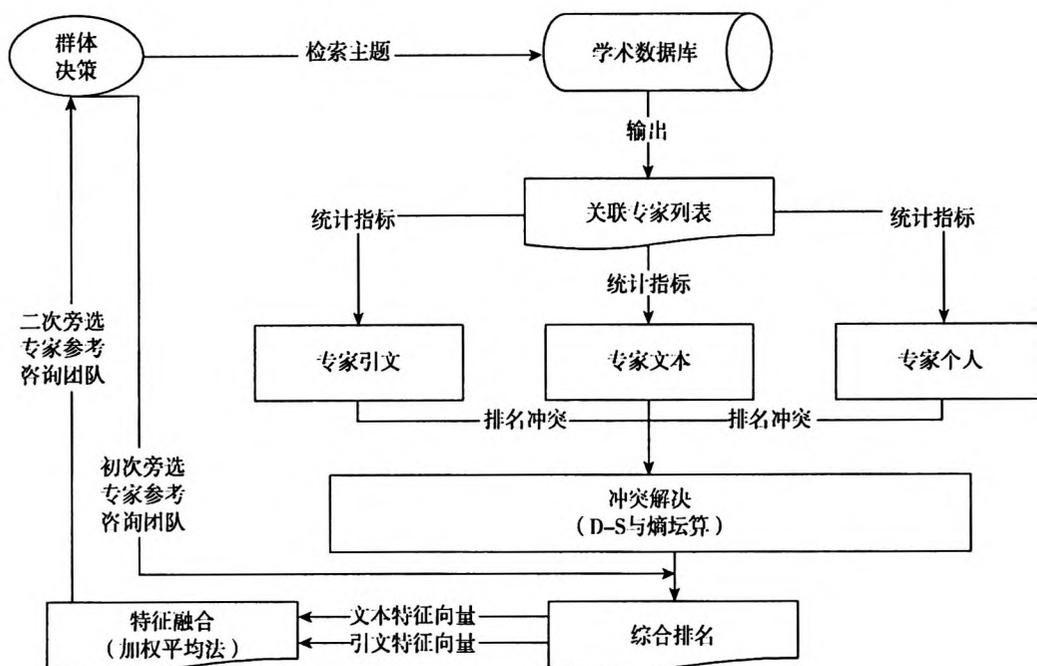


图1 应急参考咨询团队构建过程

家分面排名的融合,并在确认专家综合排名的基础之上,融合其文本特征向量和引文特征向量来辅助专家遴选,从而最终完成应急参考咨询团队的构建。整体过程参见图1。

3.1 关联专家列表获取

由于学术资源质量和结构化水平较高,客观地反映了研究者的方向和过程^[3],因此本文选择学术资源作为专家分面设计及特征识别的数据来源。在相关研究综述中,笔者已经指出关联专家列表获取等同于专家主题检索。因此,在对突发事件类型与情境进行需求分析之后,我们可借助逻辑运算符(AND OR NOT XOR)将其表达出来,其式可用 q 表示。紧接着,我们要选择一个具有较强代表性的学术数据库来检索学术资源。为保证检索结果的质量,我们可对学术资源发表的刊物范围进行限定或约束。最后,从检索记录集中提取“作者”字段,获取属性值集合。

学术资源与作者之间关系多重性表现为多对多,关系名为发表,但此时获取的作者集合并非都是具有“专家地位”的作者,还可能存在着“研究生等”(由于师生合著现象比较普遍),因此除对学术资源发表时间跨度进行限定以外,我们还需要提取更多字段(如“作者单位”、“作者身份”等),一方面为了规避同名作者造成的干扰,另一方面还需从作者集合中提取用于分析的专家集合。基于文献生产者分布

的一般规律,可推出专家应是该主题领域内发文量较高的群体,因此根据普莱斯公式^[19]等,可对关联专家列表进行清洗和剔除,获取具有“专家地位”的作者。此时选取的专家集合内成员还可能存在着合著情况,但考虑到合著文章中署名次序对文章贡献上的差异,我们需要尽量减小不同专家发表学术资源的重合比例,重合比例越高就意味着专家研究内容越相似,这对应急咨询团队知识结构系统性具有一定影响(排名融合后还进行特征融合也是考虑到该情形的出现),因此我们需要确保最终获取的关联专家与学术资源的关系多重性为一对多。

3.2 专家分面设计

3.2.1 专家个人分面

专家个人分面(Expert Profile Facet)揭示了专家发表成果的基本情况,其统计指标主要包括:①专家发表的包含 q 的文献总数;②专家每年发表的包含 q 的文献数的极值或平均值;③专家包含 q 的第一篇或最后一篇文献的发表时间。这些统计指标分别从数量和时间两个维度对专家个人信息进行了描述,但在实际操作中,还需注意:

(1)由源指标派生出的新指标也可以采用,如专家包含 q 的第一篇和最后一篇文献发表的时间间隔。

(2)并非所有的统计指标都要使用,但维度覆盖要尽量完备,多选择每一维度中相对重要的统计指标。

(3)很多统计指标的单位不一致,因此在分析过程中,为保证所有指标值处于同一数量级(0到1之间),我们还需要对数据进行标准化处理。常用的标准化方法包括 min-max normalization 和 z-score²⁰,本文采用前者,其数学形式参见式(1)。

$$x_{norm} = \frac{x - \min}{\max - \min} \quad (1)$$

式中, x 代表每一统计指标获取的实际值,min代表该统计指标样本的最小值,max代表该统计指标样本的最大值。 x_{norm} 代表标准化之后的数值。

3.2.2 专家文本分面

专家文本分面(Expert Text Facet)反映检索主题 q 与某一专家发表文献的关联强度,其大小可通过以下统计指标来评测:①词频(TF)、逆文档频率(IDF)、文档长度(DL)等;②BM25得分。仔细分析,我们发现上述两类指标实际源自两类信息检索模型:

(1)词频、逆文档频率源自向量空间模型,差别在于:在向量空间模型中,由于权重设置上的缺陷,通常情况下,二者被结合在一起使用。但在本文,当 q 包括多个检索词时,只需将每个检索词所获取的词频或逆文档频率求和,以获取该项指标的最终值。

(2)BM25是基于二值独立的概率检索模型(BIM)的一种扩展,广泛用于搜索相关性的评分,其计算过程可大致描述为:首先将 q 解析为多个检索词,然后计算每一个检索词与文献资源的得分,最后将所有检索词得分进行加权求和,从而得出BM25得分²¹。

3.2.3 专家引文分面

专家引文分面(Expert Citation Facet)是从专家-引文网络角度评测专家的相对影响力,主要统计指标包括:①专家发表的包含 q 文献的引文总数;②专家发表的包含 q 文献的引文平均值或极值;③专家 h 指数或 g 指数。专家-引文网络包括两类节点:专家和引用文献,它们之间存在引用关系,引文量是引用关系的属性指标,量化了专家在其研究领域的影响力,其值越高代表专家影响力越大,专家与同行之间的联系强度越高,专家发表资源的领域

认可度越高; h 指数是专家节点的属性指标,也从文献计量角度反映了专家的影响力^[22], g 指数是对 h 指数的补充和衍生,用于弥补 h 指数不能很好反应高被引论文的缺陷^[23]。

3.3 排名融合

3.3.1 理论基础

依托分面及其下的统计指标,我们对专家进行了排序,但不同分面获取的专家排名可能存在冲突。由于冲突是由多个并列排序指标引起的,因此不能简单地通过优先级设置来解决这一问题,为此,笔者引入D-S理论来实现排名融合。D-S理论又称证据理论,由Dempster首先提出^[17],后经Shafer系统化完善¹⁸。它设定存在一个识别框架(用 Θ 表示),该框架下的任意元素 A 都有一个基本概率分配(Basic Probability Assignment, BPA)。BPA是一个 $2^\Theta \rightarrow [0, 1]$ 的函数 mass (简写为 m),其满足条件: $m(\varphi) = 0$ 和 $\sum_{A \subseteq \Theta} m(A) = 1$ 。此外,对于 Θ 中的任意元素 A ,根据BPA还可以计算出它的信任函数 $\text{Bel}(A)$ 和似然函数 $\text{Pl}(A)$,它们共同组成了 A 的信任区间,用于表示已有证据对 A 的支持程度。在同一识别框架 Θ 下,可将不同证据对 A 的支持度通过D-S合并规则融合统一。对于 $\forall A \subseteq \Theta$, Θ 上的两个 mass 函数 m_1, m_2 的Dempster合成规则参见式(1)和式(2)。

$$m_1 \oplus m_2(A) = \frac{1}{K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \quad (2)$$

$$K = \sum_{B \cap C \neq \varphi} m_1(B) \cdot m_2(C) \\ = 1 - \sum_{B \cap C = \varphi} m_1(B) \cdot m_2(C) \quad (3)$$

式(2)和式(3)中的 B 和 C 分别表示 Θ 的任意元素。假定 Θ 的元素总数为 θ ,则在证据合成过程中,可以构建 $\theta * \theta$ 的对称矩阵来显式展现运算过程,矩阵行列标题来自 Θ 和 mass 值,矩阵元素值包括两部分:①行列元素对应集合的交集;②行列元素对应 mass 值的乘积。

3.3.2 专家排名 mass 值计算

根据D-S理论,每一分面下的专家排名都提供了一组证据,在进行排名融合之前,首先要建立一个相应的 mass 函数,获取每一分面下排名专家的 mass 值。结合以往研究^[12,24],本文采用两种方式计算 mass 值:CombSUM算法和信息熵算法。

(1) CombSUM 算法: 3.2.1 ~ 3.2.3 设计了不同专家分面及其统计指标, CombSUM 算法核心就是将上述不同统计指标获取到的数据进行标准化处理后求和, 和值作为不同分面下排名专家的 mass 值。在此, 笔者使用 $Fusion(A)$ 表示 CombSUM 算法求得的 mass 值, $A \subseteq \Theta$ 且为单元集合, 参见式(6)。

(2) 信息熵算法: 信息熵用来测度信息的不确定程度, 其数学形式参见式(4)。

$$H(s) = - \sum_{k=1}^n p_k * \log_2 p_k \quad (4)$$

式中, s 代表某一不确定信息, 该信息存在多种可能性情形; p_k 代表某一可能情形 $k (1 \leq k \leq n)$ 发生的概率。 $p_k * \log_2 p_k$ 代表情形 k 确认发生所需的信息量, 逐一加总各情形确认发生所需的信息量, 可得到 s 包括的信息量 $H(s)$ 。当 $p_k = 1$, 表明情形 k 确认发生所需信息量为 0, 不存在不确定性; 当 $p_k = 1/n$, 即每种情形发生概率一致时, $H(s)$ 最大, 其形式可表述为 $\text{Max}H(s) = -\log_2 p_k$ 。结合上述分析, 同时根据 Moreira 等研究成果^[12], 本文 p_k 计算参见式(5):

$$p_k = \frac{R_Indicator(i, e)}{T_Experts * T_Indicators} \quad (5)$$

式中, $R_Indicator(i, e)$ 代表针对某一专家 e , 在某一分面下与其相关的值不为 0 的统计指标 i 的个数 (未做标准化处理之前), $T_Experts$ 代表该分面下专家总数, $T_Indicators$ 代表该分面下统计指标的总数。如果我们设定 s 代表某一分面下的专家集合, Θ 代表所有可能结果的集合, A 代表 Θ 下的某一元素, 综合 CombSUM 算法和信息熵算法, $m_s(A)$ 的求解可用分段函数(6)表示:

$$m_s(A) = \begin{cases} Fusion(A) & A = \{A\} \\ \frac{H(s)}{\text{Max}H(s)} & A = \Theta \\ 0 & otherwise \end{cases} \quad (6)$$

3.4 特征融合

根据群体决策专家咨询团队构建的人员需求, 从综合排名结果中初次遴选一定比例的专家作为专家参考咨询团队的候选考察对象, 为了保证专家团队知识结构的系统完整, 降低同质化现象, 还需要提取专家内容特征, 以辅助进一步的专家遴选。

3.4.1 过程描述

在《多源专家特征信息融合研究》中, 笔者从传

感器工作过程出发, 依次论述了基于知识资源、Web 资源和社会网络资源的专家特征识别方法, 并围绕“资源均衡度”设计了基于多源信息的专家特征识别方法^[4]。资源均衡度体现了加权平均法在多源信息融合中的应用过程, 应用于本文, 其特征融合大致过程可描述为:

(1) 获取文本分面和引文分面对应的专家特征向量, 两种特征向量都采用文献关键词叠加的生成方式, 差别在于文本特征向量的来源文献为专家发文, 引文特征向量的来源文献为专家发文的引文。为简化实际运算过程, 来源文献也可选择专家高被引论文及其引文作为特征识别的关联资源。

(2) 评估文本分面、引文分面对应资源的分布情况, 其值用资源均衡度 σ 来度量^[25], 其数学形式参见式(7)。

$$\sigma = |p_1 - p_2| \quad (7)$$

式中, p_1 代表了文本分面对应资源占总资源的比例, p_2 代表了引文分面对应资源占总资源的比例。由此可知, σ 值域为 $[0, 1]$, σ 越小表示两类分面资源分布越均衡, $p_1 = p_2$ 时, σ 为 0, 资源分布完全均衡, 反之, σ 越大则资源分布越不均衡。

(3) 围绕资源均衡度 σ , 采用权重设置和优先级设置完成特征向量融合。预定义资源均衡分布的临界点参数 σ_i (作为形参, 其可根据实际需要进行调整), 当 σ 在 $[0, \sigma_i]$ 内, 通过权重设置法来生成专家综合特征向量, 专家综合特征向量 v 等于对应权重 (p_1, p_2) 与向量乘积之和。当 σ 在 $(\sigma_i, 1]$ 内, 表明各特征向量可信度存在较大差异, 通过优先级设置法来获取专家综合特征向量, 专家综合特征向量 v 等于最大 p 值所对应的特征向量。

(4) 参考专家综合排名以及综合特征向量, 完成应急专家参考咨询团队的构建。在应急处置过程中, 由于群体决策对质量、时效性以及流程规范性要求很高, 且基于学术资源的排名融合和特征融合要建立在大量数据分析的基础之上, 因此应急专家库的构建是保证“排名融合 - 特征融合”这一专家遴选过程科学高效实现的基础支撑因素。

3.4.2 比较分析

将直接特征融合与加入排名融合之后的特征融合(间接特征融合)进行比较, 笔者发现二者工作量差异明显, 主要表现为:

(1) 在获取关联专家列表的基础之上, 直接特

征融合从专家发布资源(学术资源、Web 资源和社会网络资源)的内容特征出发,逐一提取关键词、模板(规则)和标签来生成专家特征向量⁴,之后再利用加权平均法生成专家综合特征向量。假定关联专家列表获取专家集合容量为 n ,某一专家 e_i 发文资源为 x_i ,发文对应的引文资源 y_i ,Web 资源为 z_i (在本文其值为 0),则需要提取的特征向量个数 $\sum_{i=1}^n (x_i + y_i + z_i)$,获取这些特征向量后至少还需要 5 次融合才能获取专家综合特征向量。

(2)间接特征融合将资源的外部特征和内部特征相结合,通过更易获取的指标(如 TF、BM25、h 指数等),运用排名融合算法对不同分面下的关联专家进行了综合排名,在应急专家咨询团队构建过程中,召集人只需依照群体决策人员需求(假定为 $r, r < n$),选取排名比较靠前的若干位专家(大于或等于 r)进行专家特征提取和融合,需要提取的特征向量个数只有原来的 r/n ,工作量大大降低。

4 实验分析

本节以 SARS、森林火灾、地震为例,以寻找能够为各突发事件应急处置提供决策支撑的专家团队(假定各突发事件参考咨询团队需求人数为 3 人)为任务,依照第 3 部分进行了实验分析。

(1)数据来源:关联作者列表是进行专家排名融合研究的源数据,本文以 CNKI 为学术数据库,以突发事件所属管理领域的期刊为刊源,设计相应的

检索式,获取学术资源题录信息。以“SARS 应急管理”为例,刊源涵盖医药卫生科技领域的 185 类期刊,以“SU = SARS AND SU = 应急”为检索式,可用期刊文献共计 327 篇,关联作者 875 名,根据文献生产者分布规律,我们取各领域发文量较高的(结合普莱斯公式和需求人数设计)作者为团队候选考察对象,SARS 领域选择了 8 位,森林火灾领域选择了 11 位,地震领域选择了 5 位。

(2)分面指标选择及统计:以“发表期刊总数”、“发表期刊年限”为专家个人分面统计指标,以“BM25”、“TF”为专家文本分面统计指标,以“引文总数”、“H 指数”为专家引文分面统计指标。以上述三个领域 24 名专家为对象,统计不同分面下其关联资源的统计指标值,结果参见表 1。

(3)排名融合及对比分析:由 4(2)得出的 3 种专家分面排名可以看出,专家分面排名存在着两方面问题:①单一分面下专家 mass 值相同时,无法进行排名,如通过 CombSUM 计算可知,SARS 领域个人分面下的 YYH 和 WJF 的 mass 值都为 0.2029,森林火灾领域引文分面下的 ZC、SSD 和 CGY 的 mass 值均为 0.0542。②不同分面下专家排名次序出现相逆时,无法确认哪一种排名更合理,如地震领域个人分面和引文分面下 NGZ 排名高于 SGW,但文本分面下 SGW 排名高于 NGZ,类似情形在森林火灾和 SARS 领域也存在。为了消除上述问题,本文应用第 3 节论述的排名融合流程,经过两次排名融合,获取专家综合排名,结果参见表 2。

表 1 专家分面统计指标标准化结果

专家	指标	发文数	发文年限	BM25	TF	引文数	h 指数	
	SARS	YYH	1	1	0.4	0.0032	0.1098	0.2564
WJF		1	1	0.4	0	1	1	1
ZC		0	0	0.2	0.0475	0.9234	0.8462	0.6667
CQ		0	0	0	0.0265	0	0.2051	0.3333
YGF		0	0	0.2	0.0658	0.1083	0.2564	0.3333
LXL		0	0	1	1	0.0552	0	0
LYY		0	0	0	0.0265	0	0.2051	0.3333
CXT		0	0	0.4	0.3831	0.1062	0.0769	0

续表

专家	指标	发文数	发文年限	BM25	TF	引文数	h 指数	
森林 火灾	ZC	1	0.1111	0	0.7845	0.3415	0.5	
	FZK	0.5	0.3333	0.0947	0.4971	1	1	
	WJH	0	0.7778	0.2993	0.3452	0.4146	0.5	
	YJZ	0	0.1111	0.3261	0.9728	0.1463	0.5	
	ZC	0	0	0.3249	1	0.1951	0.5	
	JW	0	0.1111	0.3150	0.5690	0.2927	1	
	CJY	0	0	0.3249	1	0.1951	0.5	
	SSD	0	0	0.3249	1	0.1951	0.5	
	CS	0	1	0.3088	0.5126	0.2439	0.5	
	SLF	0	0.3333	0.2539	0.0628	0	0	
	PJB	0	0.6667	1	0	0.02439	0	
地震	NGZ	1	1	0	1	1	1	
	WXQ	0.0588	0.2	0.7825	0.2017	0.1910	0.3333	
	LYQ	0.0588	0	0.8819	0.5130	0.1738	0.3333	
	WNN	0	0.4	0.9442	0	0	0	
	SCW	0	1	1	0.9091	0.5579	0.5556	

注释:表中各项统计指标均为标准化处理后的结果,其中 BM25 和 TF 采用算数平均值,发文年限为派生统计指标,采用近似计算。

表2 综合排名统计结果

SARS		森林火灾		地震	
专家	综合排名 mass	专家	综合排名 mass	专家	综合排名 mass
YYH	0.1403	ZC	0.1232	NGZ	0.3246
WJF	0.2958	FZK	0.1466	WXQ	0.0649
ZC	0.1181	WJH	0.0981	LYQ	0.0635
GQ	0	YJZ	0.0624	WNN	0.0431
YGF	0.0323	ZC	0.0569	SCW	0.2080
LXL	0.1850	JW	0.0733	—	—
LYY	0.0020	CJY	0.0569	—	—
CXT	0.0730	SSD	0.0569	—	—
—	—	CS	0.1124	—	—
—	—	SLF	0.0252	—	—
—	—	PJB	0.0647	—	—

为对比本文综合排名方法相比基于单一分面排名方法的效果,笔者将表 2 综合排名结果与分面排名结果进行比较分析,评测指标为排名值重复率、排名一致率。排名值重复率定义为分面排名中专家 mass 值重复比例的均值(由于存在三个分面),其值越大表明排名采纳率越低;分面排名一致率定义为在不同分面排名下的顺序一致的专家对数,其值越大表明分面排名兼容性越强。

由表 3 对比结果可知,在应急各领域专家甄选时,综合排名下的专家 mass 值重复率都不高于分面排名(如 SARS 领域下,分面排名重复率为 50%,综合排名重复率为 0),这说明综合排名更具可实施性,通过一次排名被采纳率要高于分面排名。同样,排名一致率是为考察排名兼容性而设计的评价指标,各领域分面排名统计指标均低于综合排名,这说明不同分面排名之间的合理性无法度量,兼容性不好。综合排名重复率和一致率统计分析结果,综合排名融合了专家不同分面排名信息,其兼容性和一次采纳率要更高,同时所涵盖的专家信息也要更全面和合理。

表 3 综合排名与分面排名对比分析结果

领域	指标	排名重复率	排名一致率
SARS	分面排名	50%	50%
	综合排名	0	100%
森林火灾	分面排名	33.33%	16.36%
	综合排名	27.27%	100%
地震	分面排名	0	50%
	综合排名	0	100%

(4)特征融合及统计:

依据专家综合排名结果,根据应急专家参考咨询团队需求人数,选择排名 Top 3 的专家构建参考咨询团队,但为规避选取的专家知识结构同质程度过高(如森林火灾中的 ZC、SSD、CJY 的综合特征向量接近一致,参见表 4,源于三者来自同一单位,研究方向一致,合作比较紧密),还需要提取专家综合特征向量,以备二次甄选,以森林火灾领域为例,设 $\sigma_i = 0.7$,特征融合结果参见表 4。

表 4 特征融合统计结果(以森林火灾为例)

专家名	资源均衡度	设置方式	综合特征向量
ZC	0.3636	权重设置	森林火灾(5.4091),湖南省(4.3636),应急资源(1.3636),森林防火(0.6818),应急人员调度(0.6818)
FZK	0.7872	优先级设置	森林火灾(11),风险管理(7),应急管理(6),林火灾害(3),森林防火(3)
WJH	0.7143	优先级设置	森林防火(5),森林火灾(2),短波(1),跨端中继台(1),通讯盲区(1)
YJZ	0.4	权重设置	森林火灾(4.2),应急管理(2.1),预警监测(1.4),应急预案(1.4),应急处置风险(1.4)
ZC	0.5	权重设置	森林火灾(5.5),湖南省(4.0),应急资源(1.5),森林防火(0.75),扑救人员调度(0.75)
JW	0.4	权重设置	森林火灾(4.9),应急管理(2.8),林火灾害(1.4),风险管理(1.4),分级响应(1.4)
CJY	0.5	权重设置	森林火灾(5.5),湖南省(4.0),应急资源(1.5),森林防火(0.75),扑救人员调度(0.75)
SSD	0.5	权重设置	森林火灾(5.5),湖南省(4.0),应急资源(1.5),森林防火(0.75),扑救人员调度(0.75)
CS	0.5714	权重设置	森林防火(3.7857),森林火灾(2.3571),防火水位监测(1.5714),短波(0.7857),应急通讯(0.6429)
SLF	0.5	权重设置	森林防火(1.5),灾害预警系统(0.75),信息管理(0.75),Google Earth(0.75),雷击火(0.75)
PJB	0.2	权重设置	森林火灾(1.0),低温雨雪冰冻灾害(0.6),大火警示(0.6),森林大火(0.6),避险措施(0.6)

综合4(3)和(4)分析过程,确认各领域参考咨询团队的人选,SARS领域为:WJF,LXL和YYH;森林火灾领域为:FZK,ZG,CS;地震领域为:NGZ,SGW和WXQ。

5 结语

专家咨询团队在应急群体决策过程中发挥着重要作用,因此针对某一特定突发事件情景,选择合适准确的专家非常关键。目前揭示专家特征的资源类型多样,在已有基于多种资源的专家特征融合方法基础之上,我们尝试首先通过某一资源的不同分面来实现专家排名融合,继而实现内容融合。排名融合将D-S理论与熵运算结合在一起,解决了不同分面获取专家排名存在的冲突。在综合排名基础之上,内容融合对象就限定为了排名相对靠前专家。相比原来直接提取专家内容特征的方法^[4],新方法统计指标值易获取,内容特征计算量也较小。本文以三类突发事件专家甄选过程为例进行了演示,但在分析过程中,我们发现现有方法仍需进一步完善:①每一分面下的统计指标(条目和数目)需人工设定,缺乏有效的指标有效性评估方式;②文中所用mass值求解的组合公式简化了参与排名融合的元素数目,(θ 为关联专家列表数目,理论数为 2^θ ,实际数为 $\theta+1$),公式适用性还需进一步检验。

参 考 文 献

- [1] Witze A, Morello L, Turner M. Scientific advice: Crisis counsellors[J]. Nature, 2014, 512(7515): 360-363.
- [2] 程少川,李高,郑俊. 面向跨学科创新合作的知识推送方法研究[J]. 情报学报, 2013, 32(2): 148-153.
- [3] 李纲,叶光辉. 基于模糊匹配的跨学科专家团队发现算法研究[J]. 情报学报, 2014, 33(1): 68-76.
- [4] 李纲,叶光辉. 多源专家特征信息融合研究[J]. 现代图书情报技术, 2014(4): 27-34.
- [5] 夏立新,王忠义. 图书馆知识专家地图的XTM构建方法研究[J]. 中国图书馆学报, 2009, 35(180): 47-52.
- [6] Fang Y, Si L, Mathur A. FacFinder: Search for Expertise in Academic Institutions[R]. West Lafayette: Purdue University, 2008.
- [7] 巩军,刘鲁. 基于个人知识地图的专家推荐[J]. 管理学报, 2011, 8(9): 1365-1371.
- [8] 廖开际,叶东海,闫健峻,等. 基于加权语义网的专家知识发现及表示方法[J]. 情报学报, 2012, 31(1): 60-64.
- [9] Lin C Y, Ehrlich K, Griffiths F V, et al. SmallBlue: people mining for expertise search [J]. IEEE MultiMedia, 2008, 15(1): 78-84.

- [10] 陆伟,韩曙光. 组织专家的检索系统设计与实现[J]. 情报学报, 2008, 27(5): 657-663.
- [11] 王曰芬,王雪芬,杨小晓. 基于社会网络的科技咨询专家库的构建方案和流程设计[J]. 情报学报, 2012, 31(2): 116-125.
- [12] Moreira C, Wichert A. Finding academic experts on a multisensor approach using shannon's entropy [J]. Expert Systems with Applications, 2013, 40(14): 5740-5754.
- [13] Liu D R, Chen Y H, Kao W C, et al. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites [J]. Information Processing & Management, 2013, 49(1): 312-329.
- [14] 李辉,潘恺,张新. 基于模糊集合的多传感器信息融合算法[J]. 计算机工程与应用, 2013, 48(10): 128-131.
- [15] 张振海,王晓明,党建武,等. 基于专家知识融合的贝叶斯网络结构学习方法[J]. 计算工程与应用, 2014, 50(2): 1-4.
- [16] 杨雷,左文宜. 动态群体决策: 个体交互、知识学习和观点演化[M]. 北京: 科学出版社, 2011: 103-103.
- [17] Dempster A P. Upper and lower probabilities induced by a multivalued mapping [J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [18] Shafer G. A Mathematical Theory of Evidence [M]. Princeton: Princeton University Press, 1976.
- [19] Derek de Solla Price. Little Science, Big Science [M]. New York: Columbia Press, 1965: 53-53.
- [20] 苏娜,张志强. 基于z得分的科学计量学多关系融合方法研究[J]. 情报学报, 32(3): 244-250.
- [21] Robertson S E, Walker S, Jones S, et al. Okapi at TREC-3 [A]//Harman D K. Overview of the Third Retrieval Conference [C]. Gaithersburg: NIST special publication, 1995: 109-126.
- [22] Hirsch J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
- [23] Egghe L. Theory and practise of the g-index[J]. Scientometrics, 2006, 69(1): 131-152.
- [24] 贺玲玲,邓健. 信息检索融合的影响因素研究[J]. 图书情报知识, 2014(5): 81-90.
- [25] 吴贞东,向生建,曾德胜. 一种多维集合划分负载均衡资源优化分配算法[J]. 计算机应用, 2007, 27(5): 1208-1209.

(责任编辑 魏瑞斌)