应急专家发现路径融合模型探究*

叶光辉 （华中师范大学信息管理学院 武汉 430079）

摘 要 文章鉴于目前应急专家发现存在的问题，提出构建应急专家发现路径融合模型。从融合介入时间节点将模型划分为三个子模式：应急专家关联资源整合在评估资源分布态势的基础之上，设计基于资源均衡度的路径加权融合流程；应急专家特征融合引入工程化建模逻辑，构建从特征选择到特征关联，再到特征融合的分析模式；应急专家发现结果融合在不改变原有发现渠道的同时，通过设计分面指标及分面排名融合算法，获取各路径推荐的综合结果。结合典型场景及业务需求构建应急专家发现路径融合模型，可为应急处置提供全面有效的智力支撑。不同融合模式的横向比较还未进行深入探讨与分析。

关键词 应急决策 专家发现 特征融合 排名分析 资源均衡度

A Probe into the Emergency Experts Discovery Path Fusion Model

Ye Guanghui （School of Information Management of Central China Normal University, Wuhan，430079）

Abstract In view of the current problems in emergency experts finding, the author proposes to build path fusion model. The model can be divided into three sub-modes from the time node of fusion intervention as follow: on the basis of assessing the distribution of resources, path-weighted fusion process based on resource balance degree has been designed in resource integration related to emergency experts; Incorporating engineering modeling logic, the analysis model from feature selection and feature correlation to feature fusion has been built in emergency experts feature fusion; While not changing the original path discovery channel, the baseline finding results are merged by designing the facet index and faceted ranking fusion algorithm to obtain the comprehensive recommendations. Combining the typical scenarios and business requirements, the construction of the path fusion model for emergency expert finding can provide comprehensive and effective intellectual support for emergency response. The horizontal comparison of different fusion modes has not been thoroughly discussed and analyzed.

Keywords crisis decision-making， expert finding， feature fusion， ranking analysis， resource balance degree

1 引言

应急决策（Crisis Decision-Making）是一种非常规的、非程序化的决策，决策主体在有限的资源、人力、时间和信息等不利因素下，对突发事件态势做出研判和快速行为的过程。与常规决策相比，应急决策具有决策目标动态权变、决策信息严重不对称、决策环境复杂多变、决策步骤非程序化等特点。作为一种非结构化的决策，应急决策采用定量分析模型不易，因此专家咨询法应用尤其广泛。融合了专家群体的知识与经验，应急专家咨询法可最大程度地降低方案施行过程产生的风险，提升决策的质量，其鲜明特征体现在：参与应

* 本文系国家自然科学基金项目“基于‘排名分析-特征识别’的应急决策专家发现”（批准号:71740012）、湖北省自然科学基金“基于社会化标签挖掘的智慧城市‘印象云’构建模式研究”（编号: 2018CFB387）和中央高校基本科研业务费项目“基于社会化标签挖掘的城市画像研究”（编号: CCNU18QN040）的研究成果之一。

急咨询的专家来源更分散，应急咨询的质量很大程度上取决于应急专家发现的效率和效果，而传统信息咨询已行业化和规范化，其咨询人员多来自机构内部。

从本质上讲，应急专家发现是决策需求与专家特征相似度匹配的过程，匹配的前提是决策需求与专家特征的最终表达形式要一致。知识经济时代，专业知识体系构成了领域内的知识共同体——专家团队，但共同体中的专家节点总体上并不一定具有明晰的架构。这一方面说明研究正趋于集群化[1-2]，另一方面也表明专家特征具有模糊性和时变性。现有应急专家发现通常借助4类网络：

（1）人际网络。人际网络是人际关系的总和，而人际关系是专家社会关系的重要组成部分。作为一个开放复杂的巨系统，应急管理具有多尺度、多变性、多主体和多因素的特征，需要大量不同领域的专家共同参与到决策咨询中[3]。为领域内的应急任务发现专家，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，在选择专家时可能会造成一定误差，这些误差源自信息的不对称[4-5]。

（2）文献网络。文献网络是文献资源及其关系的总和，目前国内外围绕它开发的较有代表性的系统包括：FacFinder专家搜索引擎[6]、SmallBlue专家搜索软件[7-8]、科技咨询系统[9-10]、专家检索系统[11-12]、基于本体的专家定位系统[13]和C-DBLP（http://c-dblp.cn/）等。除此之外，各类文献系统也已成为定位专家的重要途径，目前国内三大学术数据库（CNKI、万方和维普）都有专家检索入口，但数据量并不大，且专家信息组织偏重外部特征，对内容特征的揭示不深。

（3）组织网络。通过电子政务等工程，目前应急管理各级行政部门很多建设了专家库、专家名录等结构化数据，但这些数据的采集往往依靠行政命令，缺少应急相关的客观知识体系辅助，专家特征表达规范程度不高，存在一词多义、异词同义等自然语言问题，决策需求与专家特征的完全匹配将造成部分相关专家的漏检[14]，进而影响到应急决策咨询团队知识结构的系统性。除此之外，现有专家库、专家名录更新周期往往较长，直接影响到应急专家发现的全面性。

（4）开放网络。开放网络关联到的专家资源数量大、质量参差不齐，且部分信息来源于文献数据库。这些资源多采用检索模型来处置专家识别过程[15-16]，开放网络资源专家特征（如名称、研究方向等）常以关键词形式包括在非结构化或半结构化的文本中，需要采用文本分析方法获取属性数据。鉴于应急决策质量的高要求，从开放网络发现的专家可信度仍需要进一步验证，而可信度大小可作为结果相关度判断的标准之一，为决策主体认可与否专家提供参考[17]。

现有四种专家发现路径都各自存在一定的局限性：通过人际网络发现专家不够客观，且不利于跨学科专家发现；通过组织网络层层上报的专家特征表达不够规范，且更新周期比较长，需要构建客观知识体系进行规范；文献网络发现的专家多来自学术机构，覆盖面不足，尤其是对“小众专家”识别力度不够；开放网络关联到的专家资源丰富，但质量差异明显，需要设计资源可信度评测算法，以提升专家发现质量。鉴于应急专家发现基准路径存在的问题，本研究提出应急专家发现路径的融合模型。

2　应急专家发现路径融合模型概述

应急专家发现路径融合模型以单一应急专家发现路径分析为基础，可采用“增量分析+逐步迭代”方式来进行路径融合，也可应用元组织方法将各类路径整合成特征关联网络。前者借鉴工程化设计逻辑来串联应急专家发现各个环节，后者采用文本网络化技术来构建由客观知识联系形成的异质网络。两者融合的策略与融合结果的表现形式都存在一定差异，这主要源于不同学科对融合内涵的认知不同：遵照应急专家发现系统开发模式，增量迭代模型强调将差异化的应急专家发现路径转换为统一的过程，基准发现路径不展现在最终融合结果中；元组织模型采用网络科学方法解决信息组织问题，注重不同发现路径的形式聚合，而非内容融合，最终融合结果包含基准发现路径。两种融合方式各有优劣，前者信息融合彻底，不存在对专家特征的误读，因为在逐级迭代过程中已经解决了不同路径发现的同一专家特征间存在噪声和冲突问题；后者严格上讲只是以专家为连接点，将不同路径发现的专家特征序列进行综合展现。本质上讲，后者是前者融合的基础，两者共同展现了信息融合的两个不同层级：形式聚合与内容融合。本文将在上述融合模式分析的基础之上，设计应急专家发现路径融合模型，如图1所示。

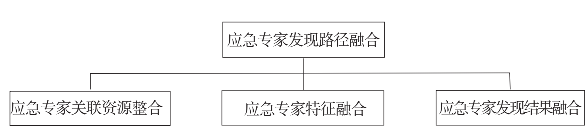

图1　应急专家发现路径融合架构

从图1可直观看出，应急专家发现路径融合可依照三类模式来展开：

（1）应急专家关联资源整合。引言以应急专家关联资源类型为分类标准，将应急专家发现路径分为4类，这些资源共同组成了应急专家特征提取的语料集，应急专家关联资源整合就是要把揭示同一专家特征的多源异构资源进行有机地整合，形成规范的未解析的应急专家发现数据集。

（2）应急专家特征融合。特征融合是在对不同分面资源解析的基础之上，利用文本分析、统计分析等算法，识别并提取能够表达应急专家特征的词序列，常见的数学表达形式为特征词向量。之后依据预定好规则对特征词进行选择，根据特征词之间存在的客观知识联系，采用耦合分析建立基于特征关联的应急专家网络，进一步对该网络进行社群结构识别，还可发掘知识相似或知识互补的专家群。

（3）应急专家发现结果融合。结果融合导向应急专家参与的典型场景及业务需求，通过突发事件分类与情景分析刻画出应急情景特征，计算出应急情景特征与应急专家特征之间的匹配度，获取不同路径推荐的专家排名后，借助排名融合算法生成专家综合排名，从而根据应急专家咨询团队人数及知识结构需求确认最终结果。

需要注意的是，上述三类模式的融合并非递进关系，而是并列关系：资源整合是在资源描述形式统一的基础之上，依据资源分布状况与信源可信度，综合确认不同资源在揭示应急专家特征的权重，采用加权计算方式获取最终结果；特征融合是在对应急专家特征识别、提取和表达的基础之上，添加特征关联环节，构建“专家–特征–专家”的异质网络；鉴于大规模在线数据更新给资源保管带来了极大挑战，结果融合强化从末端提升应急专家遴选的效率，引入分面排名融合算法来简化大规模资源整合的难度。

3 资源加权：应急专家关联资源整合流程

应急专家关联资源整合包括资源结构化处理、资源权重设计和基于资源均衡度的专家路径融合三个环节，其中资源加权是资源整合流程设计的核心逻辑。各环节具体内容为：

（1）资源结构化处理。揭示应急专家特征的资源类型多样，质量参差不齐，因此需要通过去重、清洗、合并等一系列规范化处理步骤进行预处理。同时这些资源还存在一词多义、表意模糊等自然语言问题，因而需要引入客观知识体系，以对资源语义进行合理映射和规范描述，结合元分析、主题分析等方法，将非结构的资源转换为结构化的特征向量。

（2）资源权重设计。随着应急专家对外沟通渠道的日趋多样化，适用于专家发现的资源既包括传统的学术文献，还包括社交网络资源和Web资源。不同渠道获取的资源可信度存在着一定差异，如果不依据信源可信度对资源揭示专家特征的能力进行权重设置，可能会导致融合后获取的应急专家特征出现较大噪声。权重设定既可以依托专家经验评估，还可以依据资源分布状况。为避免主观评价可能产生的随意性，提出使用资源均衡度指标来设计资源权重。资源均衡度是评估资源分布状况的综合指标，其数学表达可以描述为：在假定应急专家关联资源完备的情形下，设定各渠道资源单位数r_i，i表示资源类别号，资源单位数即为同类型资源提取的特征向量数目，如文献网络统计通过文本识别获取的特征向量（与文献单元相对应）数目，Web网络统计基于模板、规则和统计方式获取的特征向量（与文本单元对应）数目，社交网络统计专家–资源和专家–实体关系类型的组例数目。结合上述参数，资源均衡度b_degree的数学形式可表述为式（1）[18-19]。b_degree值越大表示资源分布越不均衡，揭示应急专家特征的资源类型越单一，反之亦然。

$$\text{b\_degree}=\sqrt{\sum(r_i-r_j)^2}\ (i,j\in(1,2,\ldots,n),i\neq j) \qquad (1)$$

（3）基于资源均衡度的专家路径融合。b_degree表示不同应急专家关联资源类型具有一定差异，需要构建不同的路径融合策略：加权融合策略和优先融合策略。在设定资源均衡分布临界值的基础之上，当b_degree小于临界值，采用加权融合策略，融合后的应急专家特征向量等于各资源类型权重与各基准特征向量乘积之和；当b_degree大于临界值，采用优先融合策略，优先融合是权重融合的极端情形，其过程是各资源类型权重大小，融合后的应急专家特征向量等于权值最大资源类型所对应的特征向量，此时其他特征向量作为对照组，有助于降低在资源分布极不均衡情形下简单加权计算所造成的信息噪声。

4 特征关联：应急专家特征融合流程

应急专家特征融合流程包括特征选择、特征关联和特征融合，其中特征关联是特征融合实现的关键。各环节具体内容表现为：

(1)应急专家特征选择。突发事件处置的典型场景为应急专家发现指明了特定的业务需求,但无论哪一类需求都存在共性的特征选择方法。特征选择是从预处理好的特征向量中选取有意义的特征词输入到机器学习的模型和算法中进行训练,选择将重点依据特征发散程度和与需求的相关程度来综合评断。前者重点评判特征词对样本的区分程度,常见的测度指标为方差,其值越接近0,表明特征词对样本的区分程度越弱。后者从预测分析角度计算特征词与需求的匹配程度,常见分析指标包括相关系数法、卡方检验、互信息法等。同时为最大程度降低特征选择的计算量,特征选择还可建立在特征词分类基础之上,基础分类维度可设计为三类:属性特征词、内容特征词、上下文特征词。属性特征词说明专家的基本信息,内容特征词表达专家知识背景、研究专长及影响力,上下文特征词说明专家应对突发情形的情景信息。

(2)应急专家特征关联。揭示应急专家特征的语词间存在着客观的知识联系,发掘并细化这种语义关系将为应急专家网络的构建提供基础支撑。常采用的特征关联方法有:一是统计共现分析,提取应急专家特征向量共同包含的特征词,以这些特征词为连接点,构建“专家-特征-专家”网络,进一步通过矩阵自身与其转置的相乘,可获取“专家-专家”网络;二是语义内容分析,引入领域相关词表,判别特征词间的语义类型,建构当前特征词间的语义组织网络。两种特征关联构建方法各有优劣,统计共现分析直观易操作,但特征词关联的类型相对有限;语义内容分析能够更为准确全面揭示语词间的语义关系,但需要面向特定主题的词表支撑。为兼顾统计数值计算和语义细化分析,现实应用中往往以统计共现分析所建立的矩阵为语义分析基础数据,结合领域词表、专家判定和机器学习算法对矩阵数值进行语义修正,进而服务于精细化的专家检索及社群结构识别。

(3)应急专家特征融合。特征选择只是特征处理的第一步,特征关联也只实现了应急专家形式上聚合,后续将依据特征词规模进行降维处理和融合分析。特征选择必须要进行融合模型训练,但受限于特征词规模,模型训练往往会导致计算量大,训练时间长的问题,因此需要采用基于L1惩罚项的分析、主成分分析(PCA,无监督的降维方法)、线性判别分析(LDA,有监督的降维方法)等方法将原始的样本映射到维度更低的样本空间中,这将极大提升应急专家特征向量融合的计算效率。设计应急专家特征融合方法的核心在于改进和组配现有数据挖掘、模式识别等典型算法及模型,最大程度降低特征叠加过程中产生的信息噪声和消解不同资源提取特征间的冲突。从特征选取到降维,再到特征融合,这一系列分析共同构成了应急专家发现的特征工程,该工程既要强调特征在应急专家显式表达过程的关键作用,也要保证各环节模型算法的效果和效率。

5 排名推荐:应急专家发现结果融合流程

应急专家发现结果融合包括分面排名指标设计和分面排名融合两个环节,各环节内容主要为:

(1)分面排名指标设计。结果融合最大程度地保留了原有的应急专家发现路径,在无差别获取应急专家集合的基础之上,通过分面排名指标的设计,给出面向应急处置特定主题领域的专家排名。排名计算需要综合的指标包括三个维度:一是专家个人分面,主要涉及与应急相关的个人基本信息,如机构、城市等地理信息,这将为基于位置邻近的推荐提供基础数据;二是专家文本分面,主要涉及专家专长领域、专家研究主题、参与应急处置经历等,专长领域与突发事件分类体系要紧密对照,专家研究主题要标记时间戳,以体现专家关注主题的迁移脉络,应急处置经历反映了专家实践工作能力,配合应急处置效果评估,能够最大程度简化应急专家发现流程;三是专家影响分面,主要涉及传统的学术评价和效果评价,学术评价关联到专家发表作品的载体影响力以及引用扩散情况,效果评价强调理论研究成果的技术转化情况、指导应急处置的效力等。

(2)分面排名融合。上述三个分面下都需设计定性与定量相结合的指标体系,这样才能在突发事件情景分析的基础之上,综合给出不同分面下的专家排名。获取的应急专家分面排名间可能存在一定冲突,必须设计冲突消解算法来进行分面排名融合。目前应用较多的排名融合算法是专家系统中的不确定性决策方法D-S证据推理[20-21],在之前研究中《应急参考咨询团队构建模式研究》[22]中,我们已对D-S证据推理模型在应急专家发现及团队组建中的应用过程进行了系统分析和效果验证。除此之外,还可采用加权平均、TOPSIS等算法对分面排名进行融合,最终目的都是为了获取综合排名结果,所以信息融合算法需要设计相对繁复的融合步骤,以消除分面排名间的不一致。分面排名融合效果评测指标可以选取排名值重复率和排名一

致率：排名值重复率定义为专家分面排名值重复比例的均值，其值越大表明排名采纳率越低；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。

6 结语

目前应急专家发现路径主要借助4类网络，但每一类网络都存在尚需完善的环节，借此本文试图构建应急专家发现路径的融合模型。从融合介入的时间节点，融合模型又被分为应急专家关联资源整合、应急专家特征融合和应急专家发现结果融合。资源整合以权重设计为指导原则，根据揭示专家特征的资源分布态势，可设计基于资源均衡度的路径融合流程；特征融合以特征关联为衔接纽带，借鉴工程化建模思想，意图设计从特征选择到特征关联，再到特征融合的一整套应急专家发现流程；结果融合在不改变既有应急专家发现路径的基础之上，设计分面排名指标，构建分面排名融合算法，寄期望从各路径获取一致性的推荐结果集。资源整合和特征融合可通过构建应急专家库来实现集中化管理，但在前期资源建设与特征工程设计环节需要投入较大成本；结果融合最大程度地保留了现有专家发现渠道，技术层面不存在较大问题，但需要注意应急专家发现与管理主体之间的组织协调。

同时我们也注意到，目前信息融合方法的应用对象还多集中于社会化专家，较少关联到应急管理领域，但相关研究方法对应急专家发现路径融合模型设计具有很大参考价值。同时路径融合各个环节都普遍设置了多类评测指标，为解决融合过程的噪声和冲突，各融合模式往往借鉴和改造其他学科较为成熟的方法，改造后的融合结果总体来讲都比基准方法效果要好，但改进后不同融合模式的横向比较还未进行深入的探讨与分析。

参考文献

[1] Wuchty S, Jones B F,Uzzi B. The increasing dominance of teams in production of knowledge [J].Science, 2007, 316(5827): 1036-1039.

[2] Börner K, Dall'Asta L, Ke W, et al. Studying the emerging global brain: analyzing and visualizing the impact of co-authorship teams [J].Complexity, 2005, 10(4):57-67.

[3] 范维澄.国家突发公共事件应急管理中科学问题的思考和建议[J].中国科学基金,2007,21(2):71-76.

[4] 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.

[5] 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.

[6] Fang Y, Si L, Mathur A. FacFinder: Search for Expertise in Academic Institutions[R]. West Lafayette: Purdue University, 2008.

[7] Ehrlich K, Lin C Y, Griffiths-Fisher V. Searching for experts in the enterprise: combining text and social network analysis [C]. Proceedings of the 2007 International ACM Conference on Supporting Group Work. New York: ACM, 2007: 117-126.

[8] Lin C Y, Ehrlich K, Griffiths-Fisher V, et al. SmallBlue: people mining for expertise search [J]. IEEE MultiMedia, 2008, 15 (1): 78-84.

[9] 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.

[10] 王曰芬,王雪芬,颜端武.基于社会网络的科技咨询专家库的原型系统设计与实现[J].情报学报,2012,31(3):250-258.

[11] 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.

[12] 陆　伟,刘　杰,秦喜艳.基于专长词表的图情领域专家检索与评价[J].中国图书馆学报,2008,36(186):70-76.

[13] 刘　萍,叶　燕.基于本体的高校专家定位系统研究[J].情报学报,2010,29(5):813-819.

[14] 夏立新,程秀峰,叶光辉.基于词词关联矩阵改进的模糊检索研究(二)——叙词表控制下词间关联度的改进算法[J].情报学报,2013,32(3):270-280.

[15] Balog K, Azzopardi L, De R M. Formal models for expert finding in enterprise corpora [C]. Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2006: 43-50.

[16] Liang S, De Rijke M. Formal language models for finding groups of experts [J].Information Processing & Management, 2016, 52(4):529-549.

[17] 王　平,程齐凯.网络信息可信度评估的研究现状及述评[J].信息资源管理学报,2013(1):46-52.

[18] 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, 30(4):27-33.

[19] 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].计算机工程,2007,27(5):1208-1209.

[20] Dempster A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339.

[21] Shafer G. A Mathematical Theory of Evidence [M]. Princeton: Princeton University Press, 1976.

[22] 叶光辉, 李　纲, 武　川. 应急参考咨询团队构建模式研究[J]. 情报学报, 2015, 34(7):734-742.

[作者简介]叶光辉，男，1986年生，华中师范大学信息管理学院讲师。

收稿日期：2018-06-15