

doi:10.19920/j.cnki.jmsc.2023.05.006

# 基于共词网络的群智知识挖掘方法<sup>①</sup>

——在应急决策中应用

徐选华<sup>1,2</sup>, 黄丽<sup>1\*</sup>, 陈晓红<sup>1,2</sup>

(1. 中南大学商学院, 长沙 410012; 2. 湖南工商大学前沿交叉学院, 长沙 410083)

**摘要:** 为了汇聚分布在社会网络中的群体智慧以合力化解特大公共安全突发事件重大风险, 提出了一种基于共词网络的群智知识挖掘方法, 并将其有效运用到突发事件应急管理当中. 首先, 提出了一种改进的共词网络构建方法, 用以实现社会网络公众群智知识的可视化分析, 从中获取应急决策属性参数信息; 同时, 为了确保数据质量、提高群智知识挖掘的有效性, 综合数据影响力以及数据涵盖的主题信息两方面因素, 构建了群智知识价值测度模型; 而后, 综合专家属性偏好与公众群体智慧, 提出了一种基于群智知识的动态属性信息生成办法. 最后, 以中国新冠肺炎疫情特大公共安全事件应急决策为例, 对本文所提方法进行了应用.

**关键词:** 应急决策; 共词网络; 群体智慧; 知识挖掘; 新冠肺炎疫情

**中图分类号:** C934 **文献标识码:** A **文章编号:** 1007-9807(2023)05-0121-17

## 0 引言

近年来, 我国特大公共安全突发事件(如吉林问题疫苗案件、天津港大爆炸火灾、长江客轮沉没、新冠肺炎疫情等)频繁发生. 这些事件均伴随着极强破坏性、高度突发性和不可预见性, 造成了巨大的财产、生命损失, 给社会稳定带来了重大威胁. 因此, 突发事件应急决策越来越受到各国政府和广泛学者关注<sup>[1-5]</sup>. 目前国内外还没有通用的应急决策模型, 现有模型多从多属性决策、大群体决策、案例推理、博弈论等经典方法出发. 如 Ding 等<sup>[3]</sup> 使用最差方法和毕达哥拉斯模糊不确定语言扩展了零和博弈模型, 以解决应急决策问题; Zheng 等<sup>[4]</sup> 提出基于案例推理与群体决策的动态方法以实现应急方案的生成与动态调整; 陈可等<sup>[6]</sup> 沿用相似案例推理方法, 提出事故等级生成方法, 以帮助应急决策群体把握事故态势, 科学制定应急措施.

特大公共安全突发事件应急决策视为复杂多

属性群体决策问题<sup>[7]</sup>. 决策属性是影响决策结果及其实践效果的关键要素. 目前决策领域内已经形成了诸多属性参数确定方法. 如基于信息熵优化、最大偏差法、非线性规划等统计模型的方法<sup>[3, 8]</sup>; 结合专家主观偏好的方法, 如 Liao 等<sup>[9]</sup> 将专家属性判断与专家赋权结合, 以提高决策参数取值的合理性; Sellak 等<sup>[10]</sup> 基于专家知识水平, 综合属性判断确定属性权重. 然而, 上述方法在应急环境下未必那么适用. 特大公共安全突发事件具有随机性、衍生性、不确定性等复杂特征. 此类危机情境下, 应急群体面临信息不完备、态势研判困难、决策风险认知不足、应急预案失灵等严峻挑战, 给危机处置带来了巨大风险. 因此, 亟待探究更具适应性的方法以有效应对复杂应急情境.

Patrick 等<sup>[11]</sup> 认为复杂情境下个体通常缺乏做出明智判断所需的专业知识, 而群体智慧能产生超越个体判断的聚集效应, 从而获得更加准确的判断. Wei 等<sup>[12]</sup> 将群体智慧用于新浪微博和

① 收稿日期: 2021-12-31; 修订日期: 2022-12-26.

基金项目: 国家自然科学基金资助重大项目(72293574); 国家自然科学基金资助基础科学中心项目(72088101); 国家自然科学基金资助项目(71971217).

通讯作者: 黄丽(1996—), 女, 江西吉安人, 硕士生. Email: yb961224@csu.edu.cn

Twitter 平台上的虚假新闻检测,帮助遏制虚假新闻传播,实验验证了群体智慧在协助基于机器的传统方法上的有效性和优越性. Hamada 等<sup>[13]</sup>研究表明群体智慧能有效适应复杂情境下的任务处理. 已有研究表明知识与数据协同驱动将为群体智能决策<sup>[14]</sup>提供新解法. 群体智慧的形成需要合适情境来聚集大量具有不同知识、背景的参与者. 当前,信息时代下 Web3.0、大数据技术的快速发展为集结群体智慧提供了有效支撑. 社交媒体平台凭借其开放性、实时性和用户导向性,正在吸引海量公众用户参与信息共享、聚合等知识工作、创造丰富的群体智慧知识<sup>[15-17]</sup>,利用群智知识辅助群体决策为有效应对公共危机、化解重大风险提供了创新方向. 例如:新冠肺炎疫情爆发后,依据掌握的危机信息与现场应急指挥行动,数以亿计的社会网络公众在“新浪微博”平台上通过图文、视频表达见解,由此形成了体量巨大的公众行为大数据. 这些与事件密切相关的数据反映了公众行为、态度和情感<sup>[18,19]</sup>,政府组织等应急人员从中挖掘群智知识、捕获公众诉求对于提高应急决策质量具有积极意义.

公众社会诉求具备巨大潜在应急管理价值,应当成为应急决策的重要参考<sup>[20,21]</sup>. 一方面,应急任务不仅对应急专家很重要,对公众更为重要. 公众作为公共安全突发事件全程的亲历者和核心利益主体,能否有效应对危机、化解风险牵涉到自身利益福祉;另一方面,各类突发事件应急环境比以往更加特殊、复杂,仅靠应急指挥专家组的知识与能力及时快速科学决策比较困难. 公众群体比任何独立个体都拥有更多解决问题的资源,引入大数据技术,基于海量数据挖掘群智知识,有利于应急人员更好把握事态发展、了解决策情况和公众诉求,从而提升应急决策的科学性. 因此,集成社交媒体平台上与危机事件紧密相关的公众数据,汇聚更广泛的群体智慧以合力应对复杂应急情境是必然选择,也是最佳选择.

社会网络分析、人工智能技术的发展为记录和分析突发事件期间的海量行为大数据提供了契机,为集结群体智慧助力应急管理提供了有效支撑<sup>[22,23]</sup>. 大数据时代下,政府公共危机应急决策由传统“闭门决策模式”逐渐向公众参与下的“数

据驱动智能决策模式”<sup>[24-28]</sup>转变,许多学者以行为大数据、群体智慧为研究对象,广泛开展数据驱动智能决策研究,促成了传统决策与新兴技术相结合的新态势. 如 Fu 等<sup>[16]</sup>探讨了公众关注主题演变与危机态势演变的潜在关联性,提出基于共词网络的社区演化方法,以分析危机情境下公众关注主题演变;Tsytarau 等<sup>[29]</sup>利用公众偏好大数据进行实时话题监测,输出公众实时情感变化以辅助危机应对;Wang 等<sup>[30]</sup>考虑属性之间的相关性,利用 TF-IDF 模型从文本评论中捕获用户偏好的关键要素,结合 Word2Vec 技术提取决策参数;Xu 等<sup>[31]</sup>运用空间向量模型和自组织映射方法,提出基于公众偏好大数据分析的应急决策方案动态调整方法;类似运用大数据、群体智慧辅助危机管理、群体决策的方法为公共危机应急决策提供了现实指导,但是仍然存在局限性:1) 相关方法较少考虑数据价值分析,方法对数据质量有较强依赖性. 信息技术的飞速发展极大拓宽了危机信息传播渠道,也带来了许多风险隐患. 例如各类真伪难辨的信息夹杂于亿万公众观点当中,无形中增加了危机处置维度,对应急管理的全面性和有效性提出了严峻考验;2) 部分数据驱动的相关方法忽视了决策专家观点,可能造成求解的决策参数不够符合现实应急需要;3) 相关研究对行为大数据的产生来源、形成机理分析较少,即,缺乏对公众行为的分析. 例如新浪微博用户之间的转发、评论、点赞等交互行为反映了博文在社会网络中的传播影响力,而以往一些相关研究忽视了博文影响力对数据分析的影响.

## 1 方法基础

### 1.1 TF-IDF 算法

Salton 提出计算文档中特征词权重的经典方法 TF-IDF (Term Frequency-Inverse Document Frequency), 术语频率 - 逆文档频率<sup>[32]</sup>. 其主要思想是单词术语只有在少量文本中大量出现才更有代表性.

**定义 1**<sup>[32]</sup> TF-IDF 通过词频运算得到特征词的重要程度.  $TF$  表示特征词在文档中出现的频率,  $IDF$  表示特征词出现在语料库与总文档比例

的对数,将二者相乘作为特征词的 *TF-IDF* 值

$$w_{td} = tf(t, D) \times \lg\left(\frac{N}{df_t}\right) \quad (1)$$

其中  $w_{td}$  为文档  $d$  中单词术语 (特征词)  $t$  的权重;  $tf(t, D)$  为术语频率, 表示单词术语  $t$  在文本  $d$  中出现的频率;  $\lg\left(\frac{N}{df_t}\right)$  为逆文档频率, 表示术语只有在少量文本出现才更具代表性;  $N$  指文档集合  $D$  中文档总数;  $df_t$  表示文档集合中包含单词术语  $t$  的文档数.

### 1.2 Louvain 社区检测算法

Louvain 算法是一种基于模块度的大型社区快速检测算法, 其优化目标是最大化整个网络的模块度. 该算法能够快速、灵活地实现层次化网络划分, 计算效率和效果都比较好, 被广泛应用于网络聚类分析<sup>[33, 34]</sup> 当中. 本文将使用 Louvain 算法实现群智知识挖掘.

**定义 2**<sup>[35]</sup> 模块度是评估社区网络划分质量好坏的指标, 表示社区的紧密程度.

$$Q = \frac{1}{2m} \sum_i \sum_j \left[ A_{ij} - \frac{k_i^{out} k_j^{in}}{2m} \right] \delta(c_i, c_j) \quad (2)$$

其中  $m$  表示所有边的权重和;  $A_{ij}$  表示节点  $i$  和节点  $j$  之间边的权重, 当网络不是加权图时, 边的权重均为 1;  $k_i^{out}$  表示所有从节点  $i$  引出的边的权重

之和;  $k_j^{in}$  表示所有指向节点  $j$  的边的权重之和;  $c_i$  表示节点  $i$  所属聚集, 如果  $i$  和  $j$  属于同一聚集  $c$ , 则  $\delta(c_i, c_j) = 1$ , 否则  $\delta(c_i, c_j) = 0$ .

**定义 3**<sup>[35]</sup>  $\Delta Q$  表示模块度的变化量

$$\Delta Q = \left[ \frac{\sum_{in} + k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i^{in}}{2m} \right)^2 \right] \quad (3)$$

其中  $\sum_{in}$  表示聚集  $c$  中所有边的权重和;  $\sum_{tot}$  表示与聚集  $c$  中的点相连的边的权重和;  $k_i$  表示与点  $i$  相连的边的权重和;  $k_{i,in}$  表示  $i$  与聚集  $c$  中的点相连的边的权重之和.

Louvain 算法实现聚类主要有两个阶段<sup>[35]</sup>.

1) 初始时, 将每个节点视作一个社区, 遍历每个节点, 考虑将其移至其邻居节点所属社区后模块度的变化量  $\Delta Q$ . 若最大变化量  $\Delta Q > 0$ , 将该节点移至该  $\Delta Q$  所对应的邻居节点所属社区中; 否则, 保持节点不动. 重复这一过程, 直到任何节点的移动都不能使模块度增大;

2) 将上一步形成的社区进行折叠, 即将每个社区视作一个新“节点”, 分别计算这些新生成“节点”之间的连边权重, 以及社区内所有节点之间的连边权重之和. 按照 1) 中步骤开始新一轮迭代, 直至  $\Delta Q$  低于预定阈值.

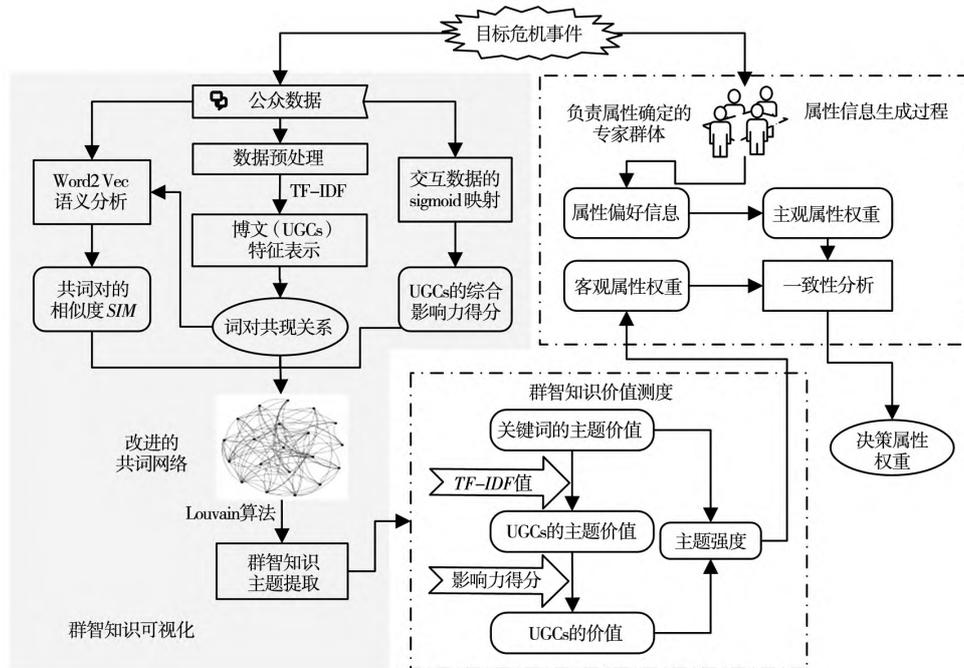


图 1 基于共词网络的群智知识挖掘方法及其应用过程整体框架

Fig. 1 The overall outline of group wisdom knowledge mining method based on co-word network and its applying process

## 2 方法原理

### 2.1 方法框架

当前群智知识挖掘技术通常聚焦于文本信息,针对公众行为与数据质量的分析较为不足. 社会网络上行为大数据体量巨大,但价值密度较低,分析所有涉及目标危机事件的公众数据显然不合理、且非必要. 已有研究表明数据反映的主题信息越丰富,其价值通常越高<sup>[36]</sup>. 此外,数据在社会网络中的影响力也能反映其价值高低. 而影响力高低某种程度上可以通过交互数据(转发量、评论量、点赞量)等信息体现,社会网络公众之间的交互行为(转发、评论、点赞等行为)决定了其交互数据的大小. 围绕目标危机的讨论活动中,具备高影响力的公众数据与其他相关数据产生交互的可能性更大,这类数据通常具有更高的交互数据指标. 因此,有理由相信公众行为分析与数据价值分析应当纳入基于行为大数据分析的群智知识挖掘流程.

本文考虑了公众交互行为和数据价值分析,以提高群智知识挖掘方法的可靠性. 方法及其应用过程整体框架如图 1 所示,包含以下内容:首

先,输入与待应对危机事件密切相关的社会网络公众行为大数据;而后,基于改进的共词分析方法生成共词网络,实现群智知识可视化,从中获取群智知识;同时,考虑到公众数据质量参差不齐,本文运用社会网络分析方法测度群智知识价值,从中提取应急决策属性参数;最后,综合专家属性偏好,利用一致性测度,构建了基于群智知识的动态决策属性信息生成方法,为大群体应急决策提供辅助支持.

### 2.2 群智知识的获取与处理

#### 2.2.1 数据获取

社会网络公众的非结构化行为大数据信息来源广泛、体量巨大,需要考虑数据获取和分析效率,以满足应急决策的及时性要求. 考虑去中心化的信息发布方式、平台开放性以及用户规模,本文选择“新浪微博”社交媒体平台作为群智知识的获取渠道<sup>[37]</sup>. 新浪微博中的社会网络公众行为通过微博用户的发布、转发、评论、点赞等活动进行描述<sup>[38]</sup>. 本文将以待解决的特大公共安全事件(案例分析部分指新冠肺炎疫情危机)的相关词为关键词,利用 Python 爬取公众分布在新浪微博上的行为大数据,相关信息如表 1 所示.

表 1 社会网络公众行为大数据的类别及其内容

Table 1 Types of social network public behavioral big data and its contents

类别	详细内容
博文数据	记作 UGCs( User Generated Contents, 用户生成内容),指公众发布的博文内容
交互数据	博文的转发量、评论量及点赞量
其他数据	时间戳(指微博的发布时间);发布人的用户名及用户 ID

为便于后续分析,同时提高知识挖掘质量,对行为大数据进行初步筛选,仅保留部分用户及其数据,用于构建待分析的博文语料库. 初步筛选原则如下.

1) 用户至少发布一条与目标事件相关的原创微博,且至少与其它相关微博产生一次交互(转发/评论/点赞);

2) 微博的交互数据不全为 0,即转发量、评论量、点赞量不全为 0.

#### 2.2.2 UGCs 特征表示

本文借鉴了大数据分析的一般性逻辑,提取

关键数据形成特征信息,并在特定应用背景下形成知识. 由于微博博文中存在大量生活化表达,简单的词频统计无法准确衡量博文文本向量中各术语的重要性. 而 TF-IDF 算法作为数据挖掘领域内的一种关键词提取技术,通过为文本向量中的术语分配分数,在有效区分高频词与关键词上表现较好,有助于确定术语在语料库中的重要性. 故本文将利用 TF-IDF 算法提取出能够代表博文的关键特征词,实现 UGCs 的特征表示,具体步骤如下.

**步骤 1** 对 UGCs 进行自然语言处理:执行分词、词性标注、去停用词、合并同义词<sup>②</sup>(运用 Py-

② 停用词及同义词列表见 <https://github.com/YB1224/ssq.git>.

thon-Synonyms 中文近义词工具) 等数据预处理操作, 构建待分析的 UGCs 语料库;

**步骤 2** 通过 TF-IDF 算法构造文本向量矩阵:  $TF-IDF = [(w_{td})_{T \times D}]^T$ . 其中,  $T$  为语料库内的术语总量,  $D$  为语料库内所有博文的数量;

**步骤 3** 依据文本向量矩阵, 计算术语  $t$  在所有博文中的  $TF-IDF$  值之和, 以衡量术语  $t$  在 UGCs 语料库中的重要性程度, 记为  $W_t$  ( $W_t = \sum_{d, d \in d_t} w_{td}, d_t$  表示包含术语  $t$  的博文集); 依据头尾断裂法<sup>[39]</sup>, 选择  $W_t$  值较高的术语作为博文的关键特征词, 实现 UGCs 的特征表示.

**2.3 基于共词网络的群智知识可视化分析**

关键词共现分析<sup>[33]</sup> 旨在利用两个或多个关键词在同一篇论文里同时出现的关系 (即, 共现关系) 来揭示学科研究主题或研究热点. 构建关键词共现网络 (简称共词网络) 有利于直观识别语义之间的关联关系. 同一主题下, 描述某一个主题的一系列关键词具有一定强度的语义关联关系, 且共同出现的可能性更大<sup>[16]</sup>. 基于这一思想, 本文将社会网络分析与关键词共现分析相结合, 构建共词网络, 以实现目标危机事件下公众群智知识可视化.

假设任意两个词只要共现 (即, 共同出现在同一博文中), 就存在语义关联. 首先, 以博文集中提取的特征关键词为基础, 按照关键词的共现情况, 生成初始共词网络. 本文将共词网络设置为无向赋权网络, 表示为  $N = (W, F)$ , 其中  $W$  表示关键词节点集合,  $F$  为边权集合; 若关键词  $w_i, w_j$  共现, 则二者构成共词对  $B-W_{ij}$ , 且两节点之间存在无向边;  $f_{ij}$  表示共词对  $B-W_{ij}$  之间的共现强度.

以往关于词共现分析的研究在确定网络边权时大多只考虑词对共现情况, 普遍做法是根据词对的共现频数计算 Jaccard 系数、Pearson 相关系数、Salton 余弦系数. 如丁晟春等<sup>[33]</sup> 和 Fu 等<sup>[16]</sup> 将词对共现频次作为网络边权; 李磊等<sup>[40]</sup> 采用“主题词 - 文档”加权频数矩阵来定义词共现矩阵. 上述方法本质都是直接基于词频数计算网络边权, 没有区分词对所属数据的质量. 为此, 本文对传统共词网络构建方法进行改进.

考虑共词对出现在不同博文中的差异性, 本文在确定  $f_{ij}$  时考虑了博文在社会网络中的影响

力. 公众之间不同交互行为对 UGCs 传播深度、广度的影响程度不同, 故不同交互行为对其影响力的贡献程度也应当不同<sup>[33]</sup>. 通过专家打分法设定不同的交互行为权重, 以综合衡量 UGCs 在社会网络中的综合影响力. 假设转发、评论、点赞三种交互影响力对综合影响力满足线性加和关系.

**定义 4** 某条博文数据  $u^c$  的综合影响力为

$$y_{u^c} = b_1 y_R^{u^c} + b_2 y_C^{u^c} + b_3 y_L^{u^c} \quad (4)$$

其中  $y_K^{u^c}$  ( $K = R/C/L$ ) 用来区分博文  $u^c$  基于三种交互行为的影响力;  $b_i \in (0, 1), i = 1, 2, 3$  依次表示转发、评论、点赞三种交互行为的权重, 且

$$b_1 > b_2 > b_3, \sum_{i=1}^3 b_i = 1$$

$$\begin{aligned} \text{令 } y_{ij} = y_{ji} &= \sum_{u^c, u^c \in UGCs_{ij}} y_{u^c} \\ &= \sum_{u^c, u^c \in UGCs_{ij}} (b_1 y_R^{u^c} + b_2 y_C^{u^c} + b_3 y_L^{u^c}) \end{aligned}$$

$f_{ij} = \varphi(y_{ij}) \times y_{ij}$  用以衡量所有出现共词对  $B-W_{ij}$  博文的综合影响力之和;  $UGCs_{ij}$  表示出现  $B-W_{ij}$  的博文集合. 显然, 共词网络的边权  $f_{ij}$  与博文的影响力正相关, 因此  $\varphi(\cdot)$  设置为单调递增函数.

令  $y_K^{u^c} = h(N_K^{u^c}), K = R, C, L$ . 显然, 转发量、评论量、点赞量 (记作  $N_K^{u^c}, K = R, C, L$  依次代表转发、评论、点赞行为) 越大, 博文在社会网络中的影响力越大,  $f_{ij}$  也越大; 反之,  $f_{ij}$  趋向于 0. 因此,  $h(\cdot)$  是单调递增函数. 此处的关键问题是如何定义函数  $h(\cdot)$ , 以实现博文交互数据到博文影响力的映射. 本文选择使用  $\text{sigmoid}(\cdot)$  函数<sup>[41]</sup>. 其主要优势在于:  $\text{sigmoid}$  映射将使大部分排列在中间位置 (即, 具有中等交互数据) 的博文数据受益, 获得相对更高的影响力得分.

**定义 5**  $y_i = \text{sigmoid}(x_i): x_i \in R \rightarrow y_i \in (0, 1)$

$$\text{sigmoid}(x_i) = 1 / (1 + e^{-u(x_i - v)}) \quad (5)$$

其中  $u, v > 0$  为待定参数,  $\text{sigmoid}(x_i)$  随  $x_i$  单调递增, 能够实现从实数集到  $(0, 1)$  的映射.

令  $y_K^{u^c} = h(N_K^{u^c}) = \text{sigmoid}(N_K^{u^c}), K = R, C, L$ . 使用  $\text{sigmoid}(x_i)$  函数分别将 UGCs 的转发量、评论量、点赞量规范化, 以度量不同交互行为对 UGCs 形成的影响力. 参考研究<sup>[41]</sup>, 通过语料库中交互数据的统计特征确定待定参数  $u$  和  $v$ . 以利用  $\text{sigmoid}(x_i)$  实现转发量  $N_R^{u^c}$  到基于转发行为的影响力  $y_R^{u^c}$  为例: 首先, 根据语料库中所有博文

的转发量大小(从低到高)对 UGCs 进行排列,选出排列在前的比例为  $y_1$  的 UGCs,归类为小组  $G_1$ ;按照转发量大小(从高到低)对 UGCs 进行排列,选出排列在前的比例为  $y_2$  的 UGCs,归类为小组  $G_3$ ;将剩余 UGCs 归类为  $G_2$ ;令  $x_1$  为  $G_1$  中博文转发量的最大值; $x_2$  为  $G_3$  中博文转发量的最小值.求解以下方程组确定  $u$  和  $v$  的取值.

$$y_1 = \text{sigmoid}(x_1), 1 - y_2 = \text{sigmoid}(x_2) \quad (6)$$

采用类似方法实现其它类型交互数据到 UGCs 影响力的规范化映射.由此,通过令  $f_{ij} = \varphi(y_{ij})$ ,本文在考虑语料库内部共现情况的基础上,纳入了 UGCs 影响力指标,以更准确描述词对间的共现关系.然而,引入影响力指标后得到的共现结果本质上反映的仍然是语料库内部的共现知识,并未考虑到语料库外部的语义相似度信息.事实上,两个关键词可能语义相似但并未出现在同一语料中,两个语义相似度不高的词却可能会在语料库中频繁出现,形成噪声共词对.因此,利用共词网络的社区检测算法进行群智知识挖掘可能会忽略部分关键词之间的语义关系.仅仅利用内部共现知识,执行基于词网络的社区检测算法不一定能正确得到群智知识主题.为此,有必要在考虑内部共现信息的基础上进一步结合外部的语义知识. Word2Vec<sup>[42]</sup> 是一种流行的词嵌入学习方法,可用于确定单词对或文本数据集之间的相似度.其主要思想是语义和句法相似的词在相同上下文中出现的可能性较高.本文拟利用 Word2Vec 方法计算关键词之间的语义相似度  $SIM(w_i, w_j)$ ,用以反映共词对的外部语义相似度信息,并将其纳入共词网络边权当中,以帮助获得更加合理的主题结果.

**定义 6** 综合内部共现信息与外部语义知识,共词对  $B-W_{ij}$  之间的共现强度  $f_{ij}$  可表示为

$$f_{ij} = \varphi(y_{ij}) = y_{ij} \times (1 + SIM(w_i, w_j)) \quad (7)$$

由此,围绕公众行为大数据,本文综合考虑博文影响力、词对共现情况与语义相似度信息,实现了基于共词网络的群智知识可视化表示.鉴于社区检测算法能够快速灵活地实现复杂网络中社区子网络的深度挖掘,本文选择利用 Louvain 算法实现群智知识主题检测.网络化分后得到的每个社区代表一个群智知识主题,主题名称由应急决策专家大群体人工归纳获得.为了提高主题社区检测

效果,拟对明显存在两个或两个以上的社区进行二次划分.

## 2.4 群智知识价值测度

大数据时代,突发事件背后的社会网络公众行为大数据已呈滚滚洪流之势不可阻挡,探索、挖掘其中蕴藏的高价值知识是获取公众情感、态度和诉求的重要渠道.由于微博用户具有超大规模性、社会性、异质性等复杂性,不同用户关注主题存在差异,对应生成的博文数据所反应的知识主题信息涵盖量也存在差异,这就导致 UGCs 质量参差不齐<sup>[15]</sup>.以往大部分数据驱动的主题挖掘研究方法缺乏对数据质量的评估,使得运用大数据辅助决策承担了较高不确定性风险.为了提高大数据辅助应急决策的执行效果,降低决策风险,有必要对基于行为大数据获得的群智知识进行价值测度.

社会网络公众对危机事件的关注点和关注度存在差异,其行为大数据所涵盖的群智知识主题信息也将有所不同.UGCs 包含的主题信息越丰富,某种程度表明与之关联的用户群体对危机事件的认知把握更全面,这些数据越可能具备高价值. Li 等<sup>[36]</sup>探讨了 COVID-19 流行期间新浪微博平台上公众数据的信息丰富度对公众参与决策的影响.基于其研究成果,本文将从 UGCs 反映主题信息的丰富度这一角度,测度群智知识价值.已知博文  $u^c$  经过预处理后通过一组特征关键词  $w_{u^c} = (w_1^{u^c}, w_2^{u^c}, \dots, w_Q^{u^c})$  表示,各关键词  $w_Q^{u^c}$  在  $u^c$  中对应的  $TF-IDF$  值记为向量  $\rho_{u^c} = (\rho_1^{u^c}, \rho_2^{u^c}, \dots, \rho_Q^{u^c})$ . 群智知识价值测度过程如下.

**步骤 1** 确定关键词的主题价值.共词网络各社区内的每个关键词节点在形成社区上所做的贡献不同<sup>[15]</sup>.本文拟测度关键词在社区形成中所做的贡献,用以衡量其主题价值.依据构建的共词网络,通过节点的加权重中心度得到关键词  $w_i$  的主题价值  $TV(w_i)$

$$TV(w_i) = \frac{\sum_{j, \delta(c_i, c_j) = 1} f_{ij}}{\sum_{i'} \left( \sum_{j, \delta(c_i, c_j) = 1} f_{i'j} \right)} \quad (8)$$

其中  $\delta(c_i, c_j) = 1$  表明关键词  $w_i$  与  $w_j$  属于同一主题社区  $k$ ,分子部分表示社区  $k$  内与关键词节点  $w_i$  相链接的边权之和,分母部分表示社区  $k$  内

所有边权之和. 因此,  $TV(w_i)$  也可解释为某一知识主题上各关键词的权重分布. 通过式(8), 可以根据关键词与社区内其余关键词的联系来判断关键词的重要性, 并清楚识别出各社区内的最重要关键词.

**步骤2** 确定博文的主题价值. 由于博文包含若干关键词, 在测度其主题价值时应当充分考虑各关键词的作用. 而博文特征表示过程中,  $TF-IDF$  值反映了各关键词在博文中的重要性. 因此, 本文将基于关键词的  $TF-IDF$  值及其主题价值  $TV(w_i)$ , 测度博文  $u^c$  的主题价值  $TV_{u^c}$ .

$$TV_{u^c} = \sum_{i, w_i \in u^c} \left( \frac{\rho_i^{u^c}}{\sum_{i', w_{i'} \in u^c} \rho_{i'}^{u^c}} \times TV(w_i) \right) \quad (9)$$

其中  $w_i \in u^c$  表示博文  $u^c$  的特征表示结果中包含关键词  $w_i$ .

**步骤3** 博文价值测度. 鉴于博文价值不仅

反映在其所包含的知识主题信息当中, 还与其在社会网络中的影响力有关. 因此, 本文将综合考虑博文的主题价值和综合影响力两方面因素, 通过式(10), 测度博文  $u^c$  的价值.

$$value(u^c) = (1 + TV_{u^c})^{(1+y_{u^c})} \quad (10)$$

注意到, 为保证  $value(u^c)$  是关于自变量  $TV_{u^c}$  和  $y_{u^c}$  的单调递增函数, 本文为两个自变量增加1.

**步骤4** 综合考虑博文价值以及关键词的主题价值, 计算知识主题社区  $k$  的主题强度  $TS^k$ , 实现群智知识价值测度.

$$TS^k = \sum_i \left( TV(w_{ki}) \times \sum_{u^c, w_{ki} \in u^c} value(u^c) \right) \quad (11)$$

其中  $w_{ki}$  表示主题社区  $k$  内所包含的特征关键词;  $w_{ki} \in u^c$  表示博文  $u^c$  的特征表示结果中包含关键词  $w_{ki}$ ;  $value(u^c)$  表示包含关键词  $w_{ki}$  的某条博文  $u^c$  的价值.

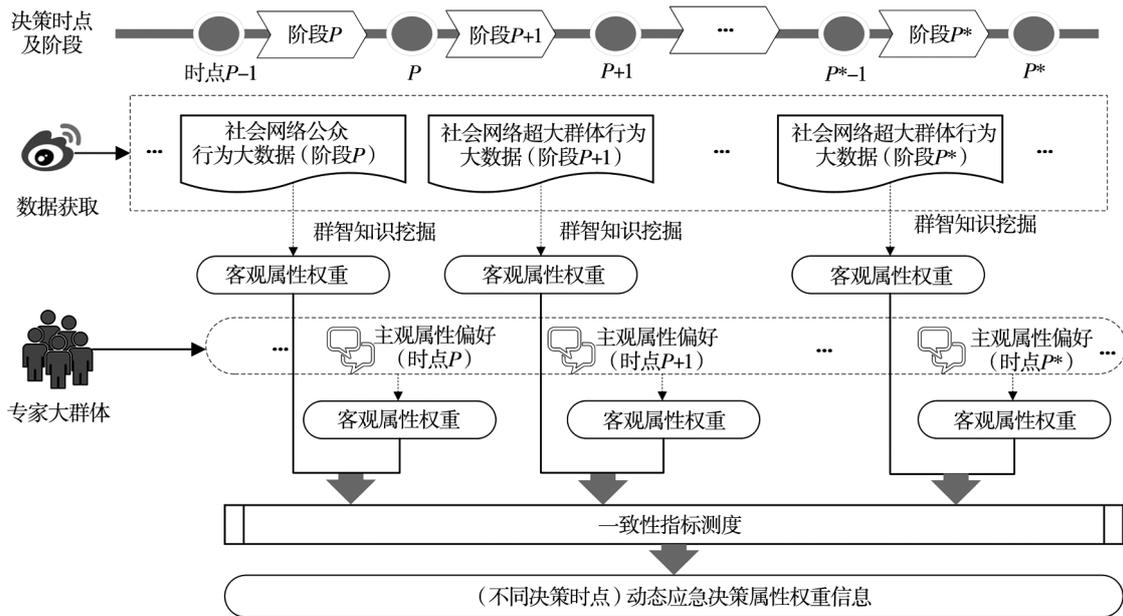


图2 基于群智知识的动态决策属性信息生成方法流程图

Fig. 2 Flow chart of dynamic decision attribute information generation method based on group intelligence knowledge

### 2.5 基于群智知识生成动态决策属性信息

特大公共安全突发事件爆发时, 需要立即召集各应急救援部门的专家组成大群体进行应急决策, 选择出最佳危机应对策略. 由于问题的特殊性、决策环境的复杂性, 突发事件应急决策视为复杂多阶段多属性群体决策问题, 此类事件的应急处置需要汇聚更多群体力量以合力应对. 事件爆

发时, 现场指挥部的应急行动与微博等平台上的公众行为大数据同时异步产生. 针对危机实时状态与应急决策处置效果, 社会网络公众在微博平台上发表个人见解, 由此形成的海量行为大数据既包含了公众对应急决策效果的实时反馈, 也包含了他们对下一阶段应急决策的诉求. 这些与危机应急紧密相关的公众行为大数据为凝聚群体智

慧、辅助应急管理人员制定和实施危机处置方案提供了必要支持。

然而,受到知识水平、决策经验等因素限制,仅仅依赖社会网络公众力量无法保障应急决策的科学性.此外,由于危机的突发性,公众行为大数据的形成与积淀可能无法满足应急决策的及时性要求.因此,应急专家组仍然是现实应急决策活动中的中流砥柱.危机问题的特殊性、复杂性以及专家在知识经验、风险偏好、职位地位等方面的异质性,使得不同专家针对同一危机应急事件往往持有不同决策观点和应对计划.为恰当反映应急管理的现实需求,同时充分发挥群智知识挖掘在应急管理中的有效作用,本文提出了一种基于群智知识的动态决策属性信息生成方法.方法流程如图2所示,从上一决策时点(如时点  $P-1$ )到当前决策点(如时点  $P$ )的公众行为大数据(即阶段  $P$  过程中产生的行为大数据),能够为专家提供决策参考,帮助确定当下决策时点(时点  $P$ )的决策属性信息.所提方法综合考虑了公众行为大数据与专家属性偏好,能够有效集结专家知识与公众群智力量,以合力应对危机处置.

假定负责确定应急决策属性权重的专家大群体(未必是进行应急处置方案选择的专家)集合:  $E = \{e_k\} = \{e_1, e_2, \dots, e_M\}$ ,第  $k$  个专家的权重记为  $\omega_k$ ,  $\sum_{k=1}^M \omega_k = 1$ ,假设各专家同等重要.专家群体依据群智知识可视化分析结果,经过协商,归纳、筛选出对危机处置与决策制定有用的  $N$  个知识主题社区,并将其作为应急决策属性信息.决策属性集合:  $C = \{c_j | j = 1, 2, \dots, N\}$ ,决策属性权重集合:  $w = \{w_j | j = 1, 2, \dots, N\}$ .

**步骤1** 对筛选出的  $N$  个知识主题的主题强度进行规范化处理,得到基于群智知识的客观属性权重

$$w_j^{(1)} = TS^j / \sum_{k=1}^N TS^k \quad (12)$$

其中  $j = 1, 2, \dots, N$  且  $c_j \in C$ .为便于与专家主观偏好区分,将式(12)得到的结果称为客观属性权

重,对结果进行降序排列,得到客观属性权重排序:  $O_1 = \{o_j | j = 1, 2, \dots, N\}$ .

**步骤2** 确定专家大群体的主观属性权重.

鉴于成对比较偏好在评估信息上的优势及其在表达专家评估上的应用广泛性,本文基于偏好关系<sup>[12]</sup>反映专家属性评估.首先,专家群体中的每个成员使用数值评分(介于  $0 \sim 1$  之间)对  $N$  个决策属性进行成对比较,  $N(N-1)/2$  次比较后得到专家的个体属性偏好比较矩阵.第  $k$  个专家的属性偏好比较矩阵:  $P^k = (p_{ji}^k)_{N \times N}$ .其中,  $p_{ji}^k$  表示专家  $e_k$  认为属性  $c_j$  重要的程度,且  $p_{ij}^k = 1 - p_{ji}^k$ .然后,以 Rank Aggregation 方法<sup>[43]</sup>为基础实现专家排序意见的聚合,通过加权平均算子聚合个体属性偏好,得到群体属性偏好比较矩阵:  $P_c = (p_{ji}^c)_{N \times N} = \left( \sum_{k=1}^M \omega_k p_{ji}^k \right)_{N \times N}$ .最后,计算属性  $c_j$  的偏好得分值:  $score(c_j) = \sum_{i, i \neq j} p_{ji}^c$ .将属性偏好得分值按照降序排列,得到基于专家偏好的主观属性权重排序:  $O_2 = \{o_j | j = 1, 2, \dots, N\}$ .依据属性偏好得分值,计算基于群体偏好的主观属性权重

$$w_j^{(2)} = score(c_j) / \sum_{j=1}^N score(c_j) \quad (13)$$

**步骤3** 一致性指标测度.为了更全面地考虑决策属性信息以满足现实决策的需要,有必要综合考虑基于群智知识的决策属性信息以及专家属性偏好,将二者比较融合,以提高最终属性结果的合理性.受 Tang 等<sup>[44]</sup>研究启发,本文将利用优势关系衡量属性之间的优先级,运用基于优先级的一致性指标,以确定最终的决策属性权重信息.定义属性  $c_j$  的优势类:  $[c_j]_s^{\geq} = \{c_i \in C | w_i \geq w_j\}$ ,属性  $c_j$  的优势类中包含所有重要性程度不劣于  $c_j$  的属性.令  $C/R_s^{\geq} = \{[c_1]_s^{\geq}, [c_2]_s^{\geq}, \dots, [c_N]_s^{\geq}\}$  表示各属性的优势类集合.基于客观、主观属性权重结果,分别计算两种决策情况(客观和主观情况)下各决策属性的优势类,依次记为  $[c_j]_{s_1}^{\geq}$  和  $[c_j]_{s_2}^{\geq}$ .定义基于优先级的一致性指标<sup>[44]</sup>.



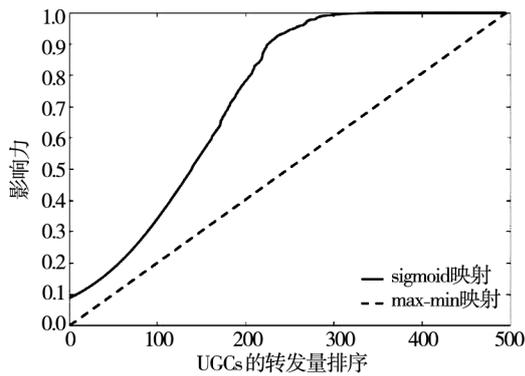


图 5a 交互数据(转发量)——影响力映射

Fig. 5a Interaction data (retweets): Influence mapping

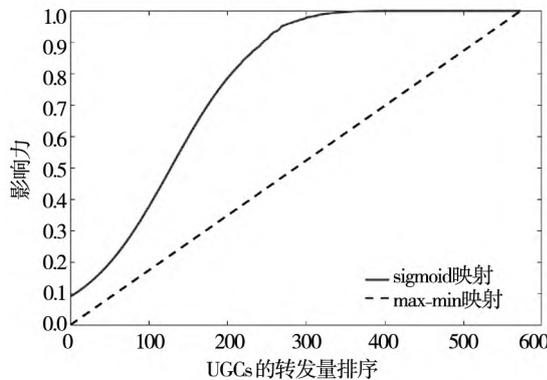


图 5b 交互数据(评论量)——影响力映射

Fig. 5b Interaction data (comments): Influence mapping

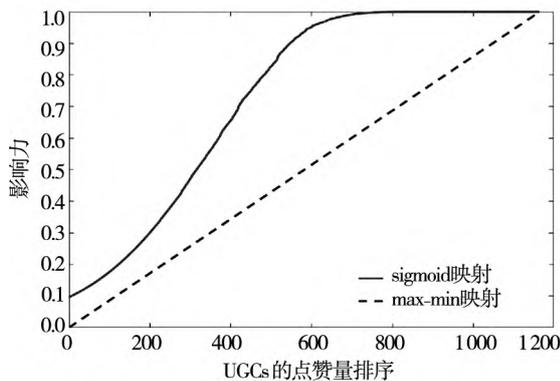


图 5c 交互数据(点赞同)——影响力映射

Fig. 5c Interaction data (likes): Influence mapping

**步骤 2** 依据关键词之间的共现关系,构建共词网络.首先,通过专家打分法确定转发/评论/点赞三种交互行为的权重,分别为 0.5,0.3 和 0.2;而后,通过 sigmoid 映射函数计算 UGCs 的影

响力.根据二八法则,令  $y_1$  和  $y_2$  取值 0.2,通过式 (6)求解参数  $u$  和  $v$ ,得到博文的影响力分布如图 5 所示.显然,与 max-min 规范化映射方法相比,通过 sigmoid 映射,具有中等交互数据量的 UGCs 将获得更高影响力,从而有利于避免影响力被具有高交互数据博文主导.

基于数据预处理结果,获取语料库内部的共现信息(即共词对集合);通过 Word2Vec 计算共词对的外部语义相似度  $SIM$ ;通过式(7)综合考虑内部共现情况与外语语义相似度信息,得到共词对之间的共现强度,由此构建共词矩阵.

为反映共现强度  $f_{ij}$  与博文影响力  $y_{ij}$  和词对语义相似度  $SIM$  的关系,对式(7)进行分析,比较了不同  $f_{ij}$  定义下的词网络划分结果,如表 2 所示,不同方式下的网络模块度有所差异.

表 2 不同共现强度下的网络划分模块度结果

Table 2 Results of the networked modularity under distinct co-occurrence intensities

方式	共现强度定义	模块度
1	$f_{ij} = y_{ij} \times (1 + SIM(w_i, w_j))$	0.527
2	$f_{ij} = y_{ij}^{(1+SIM(w_i, w_j))}$	0.426
3	$f_{ij} = (1 + SIM(w_i, w_j))^{y_{ij}}$	0.311

通常模块度超过 0.3 时,表明该网络具有明显社区结构<sup>[45]</sup>.显然,方式 1 得到的网络划分模块度最佳,因此本文选择方式 1 定义词对共现强度,由此构建的共词矩阵如表 3 所示.共词矩阵构建过程的部分源代码如表 4 所示.利用 Gephi 软件实现基于共词网络的群智知识可视化结果如图 6 所示.

**步骤 3** 对共词网络执行 Louvain 算法,进行社区检测,得到群智知识主题社区.通过标识各社区的关键词,归纳总结后人工标识了各社区的主题名称.主题社区的相关信息如表 4 所示.实验结果表面大多数关键词节点都能映射到与疫情危机相关的主题社区.



**步骤4** 基于 UGCs 特征表示结果,标识各社区内的关键词及与之相关的博文 UGCs,按照 2.4 节步骤测度群智知识价值.已知专家群体经多轮讨论协商后,最终确立了以下 4 个主题作为应急方案选择的考虑因素:人员伤亡( $c_1$ )、二次危机( $c_2$ )、社会影响( $c_3$ )、应急效果( $c_4$ ).由此,通过式(11)和式(12),得到基于群智知识的客观属性权重,结果如表 5 所示.

**步骤5** 专家群体在确定决策属性之后,于决策时点  $P$  提供个人主观属性偏好.协商一致后的群体属性偏好结果如表 5 所示.通过式(14),对客观、主观属性结果进行一致性测度,如表 6 所示,可知一致性指标为 0.937 5 ( $Con_{s1} =$

$\frac{1}{4}(\frac{1}{1} + \frac{3}{4} + \frac{3}{3} + \frac{2}{2}) = 0.937 5$ ).因此,通过式(15),得到最终的应急决策属性权重: $w = \{0.401, 0.197, 0.252, 0.150\}$ .

与此同时,为控制危机蔓延,来自不同领域的 20 位专家受邀参与决策时  $P$  的大群体应急方案决策.已知共有 5 种备选方案: $X = \{X_1, X_2, X_3, X_4, X_5\}$ ,方案主要内容如表 7 所示.专家群体利用数值评分(介于 0 ~ 1 之间)对各方案在各属性下的优劣进行评分;而后经过属性权重加权处理后得到方案最终得分.群体协商一致后得到的方案评价信息如表 8 所示.因此, $X_5$  为决策时点  $P$  的最佳应急方案.

表 5 客观属性权重(阶段  $P$ )与主观属性权重(时点  $P$ )

Table 5 Objective attribute weights (at the stage  $P$ ) and subjective attribute weights (at the time point  $P$ )

属性	主题强度	$w_j^{(1)}$	$o_1$	属性	$c_1$	$c_2$	$c_3$	$c_4$	score	$w_j^{(2)}$	$o_2$
$c_1$	414.171	0.400	1	$c_1$	-	0.635	0.751	0.496	1.881	0.409	1
$c_2$	206.844	0.200	3	$c_2$	0.252	-	0.264	0.435	0.699	0.152	4
$c_3$	259.934	0.251	2	$c_3$	0.311	0.542	-	0.663	1.205	0.262	2
$c_4$	153.271	0.148	4	$c_4$	0.404	0.386	0.428	-	0.814	0.177	3

表 6 最终属性权重信息(时点  $P$ )

Table 6 Final results on attribute weights (at the time point  $P$ )

属性	$o_1$	$o_2$	$Con_{s1}$	$w_j^{(1)}$	$w_j^{(2)}$	$w_j$
$c_1$	1	1	0.937 5	0.400	0.409	0.401
$c_2$	3	4		0.200	0.152	0.197
$c_3$	2	2		0.251	0.262	0.252
$c_4$	4	3		0.148	0.177	0.150

表 7 备选方案主要内容描述

Table 7 Main description of alternatives

方案集	方案主要内容
$X_1$	各城市间正常通行,来往需进行 14 天隔离;严格落实审批报告制度,加强重点地区来(返)人员排查;引导公众科学佩戴口罩,加强居民健康监测;学校开展在线教学;群众非必要不到中高风险区,不出省;调动省内资源,弥补医疗物资缺口;组建省市医疗服务队,建立急救医院.
$X_2$	设立防疫口,体征异常者立即隔离;非涉疫地区间来往正常通行;有中高风险区域旅居史居家隔离 14 天;加强进出中高风险地区的客运管控;学校开展在线教学;引导公众科学佩戴口罩,安全社交;调动省内资源,弥补医疗物资缺口;组建省市医疗服务队,建立急救医院;做好隔离资源储备.
$X_3$	构建核酸检测快速“采样圈”;禁止高风险地区来往通行;引导公众科学佩戴口罩,取消大众聚集活动;高校封闭管理或延迟开学;幼儿园、中小学开展在线教学;加强中高风险地区人员防疫工作,做好隔离场所、检测流调力量、应急物资准备;升级防控方案,全流程实战演练;寻求省内外支援,弥补医疗物资缺口;组建医疗服务队,建立急救医院.
$X_4$	设立防疫口;高风险地区城市管控处理,来往需提供 72 小时核酸阴性证明;非涉疫地区正常通行;通过社交媒体平台更新推送防疫动态,加强公众心理疏导;严格审批监管大众聚集性活动;学校开展在线教学;调动省内资源,弥补医疗物资等缺口;组建省市医疗服务队,建立急救医院.
$X_5$	科学划定风险区域,分级落实封闭封控;中高风险区域封闭管控,暂停大众聚集活动;高校封闭管理或延迟开学;幼儿园、中小学开展远程教学;构建核酸检测快速“采样圈”;有中高风险区域旅居史居家隔离 14 天;非涉疫地区间来往正常通行;加强医疗、检测、流调等领域专业队伍培训;寻求省内外支援,弥补防疫物资缺口;跨省区组建医疗服务队,建立急救医院.

表 8 应急决策方案评价信息及其排序结果  
Table 8 Evaluations of emergency decision alternatives and their ranking results

方案 \ 属性	$C_1$	$C_2$	$C_3$	$C_4$	方案得分	方案排序
$X_1$	0.456 9	0.413 0	0.511 3	0.674 6	0.494 4	$X_5 > X_3 > X_2 > X_1 > X_4$
$X_2$	0.495 8	0.524 0	0.503 8	0.592 7	0.518 0	
$X_3$	0.522 9	0.422 9	0.643 9	0.551 1	0.537 4	
$X_4$	0.517 9	0.432 3	0.413 4	0.542 8	0.478 4	
$X_5$	0.786 2	0.510 8	0.570 7	0.699 3	0.664 2	

3.3 方法比较

由于当前基于公众行为大数据分析的群智知识挖掘方法研究较为缺乏,为验证所提方法的有效性,选择了与本研究相关的两种方法进行比较.对比方法相关信息如表 9 所示.

为了更清晰地反映结果的差异性,比较了实验下属性权值排序结果,并且利用式(14)计算了三种排序结果两两之间的一致性,如表 10 所示.显然,由于方法考虑因素的差异性,三种方法得到

的属性权值有所差异.方法 1 与方法 3 下,属性权值排序结果完全一致;与运用 Word2Vec 技术的方法相比,方法 2 结果的一致性更低.这主要是因为 TF-IDF 算法获得的关键因素没有考虑单词语之间的语义相似性,容易导致数据维度灾难问题,降低数据分析结果的准确性.本文方法引入 Word2Vec 算法获得词对之间的相似度信息,降低了维度灾难对结果的负面影响,方法具有一定有效性.

表 9 对比方法相关信息

Table 9 Relevant information on comparison methods

方法	来源	方法描述
M1	本文方法	运用 TF-IDF 算法及 Word2Vec 算法;综合考虑公众数据在社会网络中的影响力、特征关键词的 TF-IDF 值及词对语义相似度;为便于方法比较,暂不考虑专家的属性偏好及属性赋权.
M2	Xu 等 <sup>[46]</sup>	运用 TF-IDF 算法确定关键词及其权重 $w_h$ ; 专家评估各关键词的损害严重程度 $S_h$ (本文令 $S_h$ 取等值);利用 $S_h$ 对 $w_h$ 加权得到属性的综合损害程度;属性权重与综合损害程度正相关.
M3	Wang 等 <sup>[30]</sup>	运用 TF-IDF 算法获取关键因素及其权重 $w_h$ ; 通过 Word2Vec 算法获取关键因素与每个属性之间的相似度 $S_h$ ; 综合 $w_h$ 与 $S_h$ 求解属性权重.

表 10 实验结果比较

Table 10 Comparisons of the experimental results

方法	权值	权值排序	一致性
M1	$w = \{0.400, 0.200, 0.251, 0.148\}$	$O_{M1} = \{o_1, o_3, o_2, o_4\}$	$Con_{M1, M2} = 0.854 1$ $Con_{M2, M1} = 0.875 0$ $Con_{M1, M3} = 1.000 0$ $Con_{M2, M1} = 1.000 0$ $Con_{M2, M3} = 0.875 1$ $Con_{M3, M2} = 0.854 2$
M2	$w = \{0.328, 0.252, 0.184, 0.235\}$	$O_{M2} = \{o_1, o_2, o_4, o_3\}$	
M3	$w = \{0.504, 0.157, 0.223, 0.116\}$	$O_{M3} = \{o_1, o_3, o_2, o_4\}$	

4 结束语

特大公共安全突发事件的特殊性、变异性和复杂性为此类事件应急决策带来了重大挑战,使得应急决策具有更高的科学性和有效性需求,需

要汇集更加广泛的群体智慧予以支持.复杂情境下应急决策存在属性信息不充分且依赖专家主观判定的问题,传统小群体决策技术难以提供有效支持.针对上述解决需求,本文聚焦社会网络公众行为大数据,以新冠肺炎疫情下的新浪微博数据为分析对象,基于社会网络分析、改进的共词网络

分析技术实现群智知识主题检测及其可视化,测度群智知识价值,并基于群智生成属性信息,由此实现了基于公众数据的群智知识挖掘及其在应急决策情境中有效运用。

研究贡献与创新点主要包括:1)利用改进的共词网络分析技术,实现群智知识可视化,为捕获公众诉求提供了创新方法.区别于传统词共现分析技术,本文综合数据影响力、词对共现关系和语义相似度确定共现强度以度量词对相似性,有利于提高主题检测的有效性;2)区别于以往未考虑公众数据及其价值分析的相关方法,本文充分运用社会网络分析,综合博文影响力以及博文涵盖主题信息两方面,构建群智知识价值测度模型,促进了高价值公众数据对应急决策的支持;3)综合公众行为大数据与专家属性偏好,提出基于群智知识的动态属性信息生成方法.该方法集成了专

家知识与公众群智,有利于协助应急人员将公共诉求转化为具体实践,提高决策的合理性;4)以“新冠肺炎疫情”为例的实验结果表明,本文将群智知识挖掘纳入应急决策流程,不仅提高了群智知识精炼效率,也能够有效适应突发事件态势转变造成的应急需求变化,从而提升应急决策的科学性。

本研究在运用群体智慧辅助应急决策上作了一项积极尝试,未来仍有一些问题值得研究:1)本文应用 Louvain 算法实现社区检测,一个关键词未必只属于一个知识主题社区.未来将考虑利用重叠社区检测算法进一步提高结果的准确性;2)特大公共安全突发事件应急决策是多阶段交互过程,应急决策需求随着事件态势演变而发生变化,与此同时,公众群智知识分布也会动态变化,未来将探讨群智知识演化情况,将其纳入应急决策流程。

#### 参 考 文 献:

- [1] 张 辉,刘远立,陈春花,等.全球性公共卫生危机治理:趋势与重点[J].管理科学学报,2021,24(8):133-146.  
Zhang Hui, Liu Yuanli, Chen Chunhua, et al. Global public health crisis governance: Trend and emphasis[J]. Journal of Management Sciences in China, 2021, 24(8): 133 - 146. ( in Chinese )
- [2] Ding X F, Zhang L, Liu H C. Emergency decision making with extended axiomatic design approach under picture fuzzy environment[J]. Expert Systems, 2020, 37(2): e12482.
- [3] Ding X F, Xue F, Liu H. A new approach for emergency decision making based on zerr-sum game with Pythagorean fuzzy uncertain linguistic variables[J]. International Journal of Intelligent Systems, 2019: 1667 - 1684.
- [4] Zheng J, Wang Y, Zhang K, et al. A dynamic emergency decision-making method based on group decision making with uncertainty information[J]. International Journal of Disaster Risk Science, 2020, 11(5): 667 - 679.
- [5] 胡 剑,朱 鹏,戚 湧.基于区块链的重大公共卫生事件下应急情报体系构建[J].情报理论与实践,2022,(5): 45.  
Hu Jian, Zhu Peng, Qi Yong. Construction of emergency intelligence system for major public health events based on blockchain[J]. Information Studies: Theory & Application, 2022, (5): 45. ( in Chinese )
- [6] 陈 可,黄燕霞,徐选华.基于突发事件相似度和区间一致性的复杂偏好大群体应急决策方法[J].控制与决策,2020,35(9): 2215 - 2224.  
Chen Ke, Huang Yanxia, Xu Xuanhua. Complex preference large group emergency decision method based on emergencies similarity and interval consistency[J]. Control and Decision, 2020, 35(9): 2215 - 2224. ( in Chinese )
- [7] Ding R X, Palomares I, Wang X, et al. Large-scale decision-making: Characterization, taxonomy, challenges and future directions from an Artificial Intelligence and applications perspective[J]. Information Fusion, 2020, 59: 84 - 102.
- [8] 万志远,刘勤明,叶春明,等.突发事件下的医院应急群决策模型研究[J].中国管理科学,2022,30(6): 9.  
Wan Zhiyuan, Liu Qinming, Ye Chunming, et al. Research on hospital emergency group decision model under unexpecter events[J]. Chinese Journal of Management Science, 2022, 30(6): 9. ( in Chinese )

- [9] Liao H, Qin R, Gao C, et al. Score-HeDLiSF: A score function of hesitant fuzzy linguistic term set based on hesitant degrees and linguistic scale functions: An application to unbalanced hesitant fuzzy linguistic MULTIMOORA[J]. *Information Fusion*, 2019, 48: 39 – 54.
- [10] Sellak H, Ouhbi B, Frikh B, et al. Expertise-based consensus building for MCGDM with hesitant fuzzy linguistic information[J]. *Information Fusion*, 2019, 50: 54 – 70.
- [11] Afflerbach P, VanDun C, Gimpel H, et al. A simulation-based approach to understanding the wisdom of crowds phenomenon in aggregating expert judgment[J]. *Business & Information Systems Engineering*, 2021, 63(4): 329 – 348.
- [12] Wei X, Zhang Z, Zhang M, et al. Combining crowd and machine intelligence to detect false news on social media[J]. *MIS Quarterly: Management Information Systems*, 2022, (2): 46.
- [13] Saiki J. Wisdom of crowds and collective decision-making in a survival situation with complex information integration[J]. *Cognitive Research: Principles and Implications*, 2020: 5.
- [14] 蒲志强, 易建强, 刘 振, 等. 知识和数据协同驱动群体智能决策方法研究综述[J]. *自动化学报*, 2022, 48(3): 17.  
Pu Zhiqiang, Yi Jianqiang, Liu Zhen, et al. Knowledge-based and data-driven integrating methodologies for collective intelligence decision making: A survey[J]. *Acta Automatica Sinica*, 2022, 48(3): 17. (in Chinese)
- [15] Shan S, Liu X, Wei Y, et al. A new emergency management dynamic value assessment model based on social media data: A multiphase decision-making perspective[J]. *Enterprise Information Systems*, 2020, 14(5): 680 – 709.
- [16] Fu X, Wang Y, Li M, et al. Community evolutionary network for situation awareness using social media[J]. *IEEE Access*, 2020, 8: 39225 – 39240.
- [17] Hamada D, Nakayama M, Saiki J. Wisdom of crowds and collective decision-making in a survival situation with complex information integration[J]. *Cognitive Research: Principles and Implications*, 2020, 5(1): 1 – 15.
- [18] Xu G, Yu Z, Chen Z, et al. Sensitive information topics-based sentiment analysis method for big data[J]. *IEEE Access*, 2019, 7: 96177 – 96190.
- [19] Bluttinger E J, Shahid S, Jarou Z J, et al. Translating COVID-19 knowledge to practice: Enhancing emergency medicine using the “wisdom of crowds”[J]. *Journal of the American College of Emergency Physicians Open*, 2021, 2(1): e12356.
- [20] 徐选华, 马志鹏, 陈晓红. 基于公众偏好大数据分析的大群体应急决策质量动态演化研究[J]. *中国管理科学*, 2022, 30(7): 10.  
Xu Xuanhua, Ma Zhipeng, Chen Xiaohong. Dynamic evolution research on emergency decision quality of large group basing on the public preferences big data[J]. *Chinese Journal of Management Science*, 2022, 30(7): 10. (in Chinese)
- [21] 李桂华, 林思妍. 基于社会诉求数据的智慧应急体系研究[J]. *情报理论与实践*, 2022, 45(5): 123 – 132.  
Li Guihua, Lin Siyan. Research on smart emergency system based on social demand data[J]. *Information Studies: Theory & Application*, 2022, 45(5): 123 – 132. (in Chinese)
- [22] Johnson M, Albizri A, Harfouche A, et al. Integrating human knowledge into artificial intelligence for complex and ill-structured problems: Informed artificial intelligence[J]. *International Journal of Information Management*, 2022: 64.
- [23] DiVaio A, Hassan R, Alavoine C. Data intelligence and analytics: A bibliometric analysis of human: Artificial intelligence in public sector decision-making effectiveness[J]. *Technological Forecasting and Social Change*, 2022: 174.
- [24] 汤志伟, 周 维, 王 研, 等. 大数据背景下公共危机事件公众参与行为研究——基于新冠肺炎疫情的实证分析[J]. *图书馆*, 2020, (8): 33 – 40.  
Tang Zhiwei, Zhou Wei, Wang Yan, et al. Research on public participation behavior in public crisis events under the background of big data: Empirical analysis of COVID-19[J]. *Library*, 2020, (8): 33 – 40. (in Chinese)
- [25] 陈国青, 曾大军, 卫 强, 等. 大数据环境下的决策范式转变与使能创新[J]. *管理世界*, 2020, (2): 95 – 105.  
Chen Guoqing, Zeng Dajun, Wei Qiang, et al. Transitions of decision-making paradigms and enabled innovations in

- the context of big data[J]. *Management World*, 2020, (2): 95 – 105. (in Chinese)
- [26] 伍之昂, 赵新元, 黄 宾, 等. 基于文献计量的大数据管理决策研究热点分析[J]. *管理科学学报*, 2021, 24(6): 118 – 126.
- Wu Zhiang, Zhao Xinyuan, Huang Bin, et al. Status and trends in big data drive managerial decision making on bibliometric[J]. *Journal of Management Sciences in China*, 2021, 24(6): 118 – 126. (in Chinese)
- [27] 王晰巍, 王楠阿雪. 数智驱动的重大突发事件应急情报管理: 新机遇, 新挑战, 新趋势[J]. *图书情报工作*, 2022, 66(16): 9.
- Wang Xiwei, Wang Nanaxue. Data intelligence-driven emergency information management of major emergencies: New opportunities, new challenges and new trends[J]. *Library and Information Service*, 2022, 66(16): 9. (in Chinese)
- [28] 王慧敏, 黄 晶, 刘高峰, 等. 大数据驱动的城市洪涝灾害风险感知与预警决策研究范式[J]. *工程管理科技前沿*, 2022, 41(1): 35 – 41.
- Wang Huimin, Huang Jing, Liu Gaofeng, et al. Research paradigm of big data driven urban flood disaster risk perception, early warning and decision[J]. *Frontiers of Science and Technology of Engineering Management*, 2022, 41(1): 35 – 41. (in Chinese)
- [29] Tsytsarou M, Palpanas T. Managing diverse sentiments at large scale[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(11): 3028 – 3040.
- [30] Wang L, Wang X, Peng J, et al. The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded rationality behavioural decision support model[J]. *Tourism Management*, 2020, 76: 103961.
- [31] 徐选华, 刘尚龙, 陈晓红. 基于公众偏好大数据分析的重大突发事件应急决策方案动态调整方法[J]. *运筹与管理*, 2020, 29(7): 11.
- Xu Xuanhua, Liu Shanglong, Chen Xiaohong. Dynamic adjustment method of emergency decision scheme for major incidents based on big data analysis of public preference[J]. *Operations Research and Management Science*, 2020, 29(7): 11. (in Chinese)
- [32] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval[J]. *Information Processing & Management*, 1988, 24(5): 513 – 523.
- [33] 丁晟春, 王鹏鹏, 龚思兰. 基于社区发现和关键词共现的网络舆情潜在主题发现研究——以新浪微博魏则西事件为例[J]. *情报科学*, 2018, 36(7): 7.
- Ding Shengchun, Wang Pengpeng, Gong Silan. Research on potential subject discovery of network public opinion based on community discovery and keyword co-occurrence: Sina Micro-blog Wei Zexi incident as an example[J]. *Information Science*, 2018, 36(7): 7. (in Chinese)
- [34] Han J, Teng X, Tang X, et al. Discovering knowledge combinations in multidimensional collaboration network: A method based on trust link prediction and knowledge similarity[J]. *Knowledge-Based Systems*, 2020, 195: 105701.
- [35] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 2008(10): P10008.
- [36] Li K, Zhou C, Luo X R, et al. Impact of information timeliness and richness on public engagement on social media during COVID-19 pandemic: An empirical investigation based on NLP and machine learning[J]. *Decision Support Systems*, 2022: 113752.
- [37] Li Y, Chen Y H, Liu T. Survey on predicting information propagation in microblogs[J]. *Journal of Software*, 2016, 27(2): 247 – 263.
- [38] Rosa R L, DeSilva M J, Silva D H, et al. Event detection system based on user behavior changes in online social networks: Case of the COVID-19 pandemic[J]. *Ieee Access*, 2020, 8: 158806 – 158825.
- [39] Jiang B. Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution[J]. *The Professional Geographer*, 2013, 65(3): 482 – 494.
- [40] 李 磊, 刘 继, 张竑魁. 基于共现分析的网络舆情话题发现及态势演化研究[J]. *情报科学*, 2016, 34(1): 5.

- Li Lei, Liu Ji, Zhang Hongkui. Topics identification and evolution trend of network public opinion based on co-occurrence[J]. Information Science, 2016, 34(1): 5. (in Chinese)
- [41] Li P, Zhao W, Yang J, et al. Let's CoRank: Trust of users and tweets on social networks[J]. World Wide Web, 2020, 23(5): 2877 – 2901.
- [42] Lilleberg J, Zhu Y, Zhang Y. Support vector machines and word2vec for text classification with semantic features[C]. 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI \* CC). IEEE, 2015: 136 – 140.
- [43] Xu Y J, Da Q L, Liu L H. Normalizing rank aggregation method for priority of a fuzzy preference relation and its effectiveness[J]. International Journal of Approximate Reasoning, 2009, 50(8): 1287 – 1297.
- [44] Tang M, Liao H, Mi X, et al. Dynamic subgroup-quality-based consensus in managing consistency, nearness, and evenness quality indices for large-scale group decision-making under hesitant environment[J]. Journal of the Operational Research Society, 2021, 72(4): 865 – 878.
- [45] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [46] Xu X, Yang X, Chen X, et al. Large group two-stage risk emergency decision-making method based on big data analysis of social media[J]. Journal of Intelligent & Fuzzy Systems, 2019, 36(3): 2645 – 2659.

## Collective intelligence knowledge mining method based on co-word network: Application in emergency decision-making

*XU Xuan-hua*<sup>1,2</sup>, *HUANG Li*<sup>1\*</sup>, *CHEN Xiao-hong*<sup>1,2</sup>

1. School of business, Central South University, Changsha 410012, China;

2. School of Frontier Crossover Studies, Hunan University of Technology and Business, Changsha 410083, China

**Abstract:** A method for collective intelligence knowledge mining based on co-word networks is proposed and effectively applied to emergency management of emergencies, to gather the collective intelligence distributed in the social network for addressing the major risks of public safety emergencies. Firstly, an improved approach for constructing a co-word network is proposed to realize the visual analysis of public group wisdom in social networks and obtain information concerning emergency decision attributes parameters. Meanwhile, to ensure the data quality and improve the effectiveness of collective intelligence knowledge mining, a collective intelligence knowledge value measurement model is established by considering the data influence and the covered topic information. After that, an approach for dynamic decision attribute information production based on collective intelligence knowledge is proposed by combining experts' attribute preferences with public group wisdom. Finally, the proposed method is applied to the mega public safety event of COVID-19 in China.

**Key words:** emergency decision-making; co-word network; collective intelligence; knowledge mining; COVID-19