

Risk Measurement of Large Group Emergency Decision-making of Major Floods Disaster Based on Social Network

Xuanhua Xu¹, Qianhui Zhang², Linlin Wang², Shanglong Liu², Xin Yang²

¹School of Business, Central South University, Changsha 410083, China

²Department of management science and Engineering, Central South University, Changsha
410083, China

基于社交网络的重大洪涝灾害大群体应急决策 风险测度研究

徐选华¹, 张前辉², 王麟麟², 刘尚龙², 杨欣²

¹中南大学商学院, 长沙 410083, 中国

²中南大学管理科学与工程系, 长沙 410083, 中国

Abstract*

Considering the relationship between the experts and the decision risk caused by the fuzziness of the expert preference information under the social network, a research method for the risk measurement of the large flood disaster emergency decision in the social network is proposed. First, we use large data technology to extract key words, then use TF-IDF technology to process large data, extract public concern and determine attribute weight. Secondly, the relationship between experts is taken into consideration in decision making, social network relationship of decision experts is constructed based on social network technology and expert decision is combined.

The decision risk is measured by the policy preference. Finally, through the relative score, the final group decision plan is obtained.

Keywords: risk measurement; large group; social network; emergency decision-making

摘要

考虑社交网络下重大洪涝灾害应急决策专家之间的联系和专家偏好信息的模糊性引起的决策风险的问题, 提出一种社交网络下重大洪涝灾害大群体应急决策的风险测度的研究方法。首先, 利用大数据技术进行关键词提取, 再采用 TF-IDF 技术进行大数据处理, 提取公众关心的事件属性并确定属性权重; 其次, 将专家之间的联系考虑到决策当中, 基于社交网络技术构建决策专家的社会网络关系, 并结合专家决策偏好对决策风险进行测度。最后, 通过相对得分得到最终大群体决策方案。

关键词: 风险测度; 大群体; 社交网络; 应急决策

*项目来源: 国家自然科学基金项目 (71671189); 国家自然科学基金重点项目 (71790615); 中南大学研究生自主探索创新项目 (502211806)。

作者简介: 徐选华, 男, 教授, 博士生导师, 从事大数据决策理论与方法、信息系统与决策支持系统、应急管理决策、风险分析与管理、复杂工程决策方法等研究; 张前辉, 男, 硕士研究生, 从事大数据决策理论与方法、应急管理决策、风险分析与管理的研究。

1. 引言

近年来,我国重大洪涝灾害多有发生,这类突发灾害具有复杂性、不确定性、突发性、和高破坏性特征^[1]。由于应急决策涉及面较广,参与决策的专家往往来自不同专业,拥有不同的职业背景,决策能力参差不齐,决策群体具有复杂大群体特征^[2]:(1)群体规模比较庞大,且决策成员来自不同领域,异构性强;(2)决策问题属性呈现多维性;(3)需要满足较高的共识要求。然而应急决策中决策成员之间不可避免的会出现意见偏差。从而大大提升了群决策的风险,而现有研究大多都没有考虑决策风险。风险是指不利事件发生的可能性以及不利事件所导致的损失,具有较大的隐蔽性、极高的随机不确定性和动态演变性。如何在尽可能短的时间内得出较为理想的最终决策方案,最大程度降低决策风险是亟待解决的问题。

随着计算机网络技术的发展,社交媒体在人们生活中扮演着越来越重要的角色。突发事件一旦发生,亿万网民会在短时间内通过社交媒体迅速参与其中,产生大量关于突发事件的大数据信息。将社交媒体大数据与应急决策相结合可以为决策提供更加可靠的信息,如结合灾害应急的相关知识,社交媒体数据可用于检测灾害事件的发生以及事件发生的状况^[3]。黄崇福^[4]提出利用计算机建立“智联网”概念,用以分析信息所存在的风险有利于更好的进行决策。Anping Pan^[5]指出受灾群众对灾害的风险感知有最直接真切的感受。Hristidis等^[6]强调灾害风险管理的关键是灾害大数据的管理和分析,并指出灾害管理的未来发展方向要依靠计算机专家对灾害大数据的挖掘和筛选。因此,将从社交媒体中提取的大数据信息运用到应急决策是很有必要的。

目前,群体决策往往是在社会网络框架下进行的,个人观点往往依靠亲朋好友或者有相同观点人群得到支撑。在群体决策中,决策者需要解决的一个问题就是如何消除不同领域的决策者之间存在的意见偏差。在社交网络条件下,群决策的互动过程应该考虑社会网络关系^[7,8,9]。

综上,本文首先基于社交媒体大数据分析技术,识别洪涝灾害发生后网民讨论较多的话题,通过大数据关键词提取技术,确定其属性权重。其次,基于社交网络,通过分析决策专家之间的关联,得到专家之间的信任程度,再结合决策专家的决策偏好信息对专家决策风险进行测度,以得到最合理的方案。

2. 问题描述与方法基础

2.1 风险性大群体应急决策

重大突发事件的大群体决策问题可以描述为:有 M 个应急决策成员组成的应急决策群体 $E=\{e_1, e_2, e_3, \dots, e_M\}$ ($M \geq 11$)构成大群体^[10]; $\lambda=\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_M\}$ 表示专家权重矢量,其中, $\lambda_i \geq 0$ 且 $\sum_{i=1}^M \lambda_i = 1$; $X=\{x_1, x_2, x_3, \dots, x_P\}$ 表示 P 个可行的替代方案; $C=\{c_1, c_2, c_3, \dots, c_N\}$ 为方案的 N 个属性; $W=\{w_1, w_2, w_3, \dots, w_N\}$ 表示的是属性的权重矢量,其中, $w_j \geq 0$ 且 $\sum_{j=1}^N w_j = 1$; $V^i=(\tilde{a}_{ij}^l)_{P \times N}$ 为成员 e_i 对第 l 个方案第 j 个属性的决策偏好矢量。

2.2 社交网络分析

社交网络分析(SNA)^[11,12]如今正成为研究企业等社会实体之间的一种有用的工具,它有助于我们研究包括中心性、声望、结构平衡和信任关系在内的位置属性和结构平衡^[8]。社交网络是由一组节点 E 和一组边 L 组成的社会结构,其中节点代表个体或组织,边代表节点之间相互的关系。则节点 i 即代表决策者 $E=\{e_1, e_2, e_3, \dots, e_M\}$ 中的 e_i , l_{ab} 表示连结决策者 e_a 和 e_b 的边,与节点 i 相连的节点个数为节点的度。被箭头所指向的点,表示被信任的专家。

2.3 区间直觉模糊集

定义 1^[13]设论域为一个非空有限集,定义: $\tilde{a}=\{<x, \tilde{\mu}_a(x), \tilde{\nu}_a(x)> | x \in X\}$ 表示区间直觉模糊集,其中,区间集合 $\tilde{\mu}_a(x)=[\mu_a^l(x), \mu_a^u(x)]$ 和区间集合 $\tilde{\nu}_a(x)=[\nu_a^l(x), \nu_a^u(x)]$ 分别表示 X 中的元素 x 属于集合 \tilde{a} 的隶属度和非隶属度,其中 $0 \leq \mu_a^u(x) + \nu_a^u(x) \leq 1$, $\tilde{\mu}_a(x) \subset [0, 1]$, $\tilde{\nu}_a(x) \subset [0, 1]$ 。显然当 $\mu_a^l(x) = \mu_a^u(x)$ 并且 $\nu_a^l(x) =$

$v_a^U(x)$ 时, 区间直觉模糊数退化为直觉模糊数。

我们令: $\pi_a^L(x) = 1 - \mu_a^U(x) - \nu_a^U(x)$, $\pi_a^U(x) = 1 - \mu_a^L(x) - \nu_a^L(x)$ 。称 $\tilde{\pi}_a(x)$ 为 X 中元素 x 属于 $\tilde{\pi}$ 的犹豫度区间。为简便起见通常将区间直觉模糊数表示为 $\tilde{a} = ([a, b], [c, d])$, 则犹豫度即表示为 $\tilde{\pi}_a = [e, f] = [1 - b - d, 1 - a - c]$ 。

定义 2^[14] 设区间直觉模糊数的距离为 $d(\tilde{a}_1, \tilde{a}_2)$, 模糊熵为 $E(\tilde{a})$ ^[15], 则:

$$d(\tilde{a}_1, \tilde{a}_2) = \frac{\sqrt{2}}{4} (\sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} + \sqrt{(c_1 - c_2)^2 + (d_1 - d_2)^2}) \quad (1)$$

$$E(\tilde{a}) = \frac{1}{4} (a^2 - b^2 + c^2 - d^2 + e^2 - f^2 + 2(e + f)) \quad (2)$$

2.4 社交网络中心度分析

定义 3 分别用 P_{in}^i 和 P_{out}^i 表示专家 e_i 的标准化点入度和标准化点出度, N_{in}^i 表示社交网络图中箭头指向 e_i 的个数, N_{out}^i 表示由 e_i 发出的箭头个数, 则:

$$P_{in}^i = \frac{N_{in}^i}{M} \quad (3)$$

$$P_{out}^i = \frac{N_{out}^i}{M} \quad (4)$$

$$\hat{d} = N_{in}^i + N_{out}^i \quad (5)$$

2.5 大数据文本处理和 TF-IDF 技术

大数据时代下, 如何从海量数据中提取有用信息成为一个热点问题。在国内, 新浪微博无疑是一个热点话题的信息库, 尤其是产生突发性热点话题时, 微博可以在短时间内收集大量的实时信息。运用 TF-IDF 特征提取技术提取突发事件发生时段的公众微博信息, 找到公众关注的属性信息, 有利于专家更好的进行突发事件应急处理。

TF-IDF 算法主要涉及两个关键技术, 即特征词条的选取和权重的选择。其主要思想是: 在特定的文档中, 一个词语出现的频率越高, 出现的范围越小, 说明该词区分文档内容属性的标识能力越强, 其权重自然就越大。通常采用如下公式:

$$TF - IDF = tf(T) \times idf(T) = tf_d(T) \cdot \log(N / df(T)) \quad (6)$$

词频 $tf_d(T)$ 为词条 T 在文本 d 中出现的频率; N 表示文本集合中所有文本总数; $df(T)$ 表示文本集合中有多少篇文本出现词条 T 。逆文档频率 $idf(T)$ 表示出现在少量文本中的词条更有区分度并且更加重要。

3. 社交网络下大群体应急决策风险测度

3.1 社交网络下大群体聚类

Louvain 聚类方法^[16]是一种用于提取大型网络社区结构的迭代两阶段算法, 它能揭示社会网络的完整分层社区结构。主要分为两个步骤: (1) 节点分区; (2) 分区组合。分区公式和分区衡量指标公式分别为:

$$\Delta Q = \left[\frac{\sum in + 2\hat{d}_{a,in}}{2m} - \left(\frac{\sum out + \hat{d}_a}{2m} \right)^2 \right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum out}{2m} \right)^2 - \left(\frac{\hat{d}_a}{2m} \right)^2 \right] \quad (7)$$

$$Q = \frac{\sum_{l=1}^k \sum_{a \in C} (A_{ab} - \hat{d}_a \hat{d}_b / 2m) \zeta(c_a, c_b)}{2m} \quad (8)$$

其中 m 表示网络关系图中的总边数, $\sum in$ 表示总的入度, $\sum out$ 表示总出度, \hat{d}_a 表示与节点 e_a 相连的边数。 A_{ab} 为关联矩阵, $\zeta(c_a, c_b)$ 反映 e_a 和 e_b 是否有关, 若有关则 $\zeta(c_a, c_b) = 1$ 否则 $\zeta(c_a, c_b) = 0$ 。

将决策成员 $E = \{e_1, e_2, e_3, \dots, e_M\}$ 聚集成 K 个聚集, 第 k 个聚集记为 C^k , 聚集 C^k 中的成员数为 n_k , 在对群体成员聚类以后, 依据聚类结果确定成员和聚集权重。由于同一聚集内成员偏好接近, 令其具有相同的权重, 且根据多数原则, 成员多的聚集, 包含了大多数的意见, 可赋予其成员较大的权重; 与之相对的, 成员少的聚集, 赋予其成员较小的权重, 则聚集 C^k 中成员 e_i 的权重为:

$$\lambda_i = \frac{n_k}{\sum_{k=1}^K n_k^2} \quad (9)$$

聚集 C^k 的权重为:

$$\mu_k = n_k \cdot \lambda_i \quad (10)$$

显然 $0 < \mu_k \leq 1$, $\sum_{k=1}^K \mu_k = 1$ 。聚集 C^k 的

偏好矢量为:

$$G_k = \frac{1}{n_k} \sum_{i=1}^{n_k} v_{ij}^i \quad (11)$$

群体偏好矢量为:

$$R = \sum_{k=1}^K G_k \cdot \mu_k \quad (12)$$

3.2 大群体决策风险测度

社交网络下的大群体决策可以方便快速的反映专家群体之间的意见差别,并直观反映专家的偏好情况。但社交网络下很多专家的意见容易受到其他专家的影响,同时不同群体之间的意见偏好差别更是给决策带来了很大的风险。本文主要从两个角度进行风险测度:(1)社交网络下的信任风险;(2)群体偏好的不确定性风险。

定义 5 专家信任风险系数反映的是在一个社交网络群体中,该节点的专家被其他专家所信任的程度。用 θ_1^i 表示专家 e_i 的信任风险系数,群体信任风险系数为 $\theta_1^{C^k}$:

$$\theta_1^i = P_{in}^i + P_{out}^i \quad (13)$$

$$\theta_1^{C^k} = \frac{1}{n_k} \sum_{i=1}^{n_k} \theta_1^i (i \in k) \quad (14)$$

定义 6 专家偏好风险系数反映了聚集偏好的不确定性。某一聚集的偏好风险系数越大,说明该聚集专家的不确定性越高。用 $\theta_2^{C^k}$ 表示聚集 C^k 的偏好风险系数,公式为:

$$\theta_2^{C^k} = \frac{1}{n_k} \sum_{i=1}^{n_k} E(v^i) \quad (15)$$

得到每个聚集的偏好风险系数后进行归一化处理,使得 $\sum_{k=1}^K \theta_2^{C^k} = 1$ 。

定义 7 群体风险由信任风险和偏好风险共同作用,反映每个决策聚集的风险程度,用 θ^{C^k} 表示:

$$\theta^{C^k} = \gamma \cdot (1 - \theta_1^{C^k}) + (1 - \gamma) \theta_2^{C^k} \quad (16)$$

$$R^\theta = \sum_{k=1}^K \theta^{C^k} \cdot G_k \quad (17)$$

其中 R^θ 为群体偏好风险矩阵, γ 表示信任风险和偏好风险的重要程度,本文认为两者对决策风险的影响同等重要,固取 $\gamma = 0.5$ 。

3.3 方案选择

定义 8^[12] $\tilde{a} = ([a, b], [c, d])$ 为区间直觉模糊数,称 $\Delta(\tilde{a}) = (a - c + b - d)/2$ 为 \tilde{a} 的得分函数,其中, $\Delta(\tilde{a}) \in [-1, 1]$ 。 $H(\tilde{a}) = (a + b + c + d)/2$ 为 \tilde{a} 的精确函数,其中, $H(\tilde{a}) \in [-1, 1]$ 。

$$\Delta(\tilde{a}) = (a - c + b - d)/2 \quad (18)$$

$$H(\tilde{a}) = (a + b + c + d)/2 \quad (19)$$

分别计算群体的偏好得分 $\Delta(\tilde{a})^1$ 和风险得分 $\Delta(\tilde{a})^2$, 两者的差值 $\Delta = \Delta(\tilde{a})^1 - \Delta(\tilde{a})^2$ 为方案最终得分,以得到最终的方案排序。

4. 方法步骤

本文针对应急决策方案选择问题,考虑社会公众在突发事件发生时所关注的事件属性,利用大数据技术提取关键属性并确定权重。在对专家偏好进行聚类以后,利用社交网络技术对每一个聚集的风险系数进行测度,然后算出偏好得分和风险得分,通过群体偏好得分与风险得分的差值对方案进行排序,最终得到理想的决策方案。

Step1 获得属性权重及专家社交网络关系:利用 IF-IDF 技术,对特定时段内与要解决的应急问题相关的微博文本信息进行挖掘,得到重要属性并通过(6)式获的属性权重矢量 W 。同时利用社交网络技术构建决策专家社交网络关系图。

Step2 偏好聚类:专家对不同方案的不同属性给出偏好信息,偏好信息以区间直觉模糊数的形式给出。对偏好信息进行聚类,依偏好信息将大群体分为 K 个聚集,利用(10)式计算出每个聚集的权重 μ_k ,并得到群体偏好矩阵。

Step3 风险测度:运用定义 5、6、7 中的公式得到每个聚集的风险系数,并进行归一化处理,通过风险系数得到群体偏好风险矩阵。

Step4 方案选择:通过定义 8,分别得到群体偏好得分 $\Delta(\tilde{a})^1$ 和群体风险得分 $\Delta(\tilde{a})^2$, $\Delta = \Delta(\tilde{a})^1 - \Delta(\tilde{a})^2$,得到最终的方案得分,并进行方案排序。

5. 案例分析

2017 年 6 月 22 日至 7 月 2 日,湖南省宁乡县发生了历年同期历时最长、范围最广、雨量最多、强度最大的强降雨,遭遇了宁乡有水文、气象记录 60 年以来最为严

重的洪涝灾害。20 名决策专家迅速对灾情进行分析，得到三个应急方案 $A = \{a_1, a_2, a_3\}$ 。

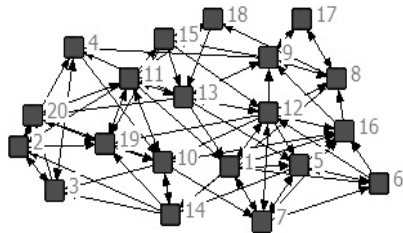


图 1 专家社会网络关系

Step1 利用 TF-IDF 技术对公众微博实时发布的灾情文本进行关键词提取，通过 (6) 式得到三个重要属性，且属性权重为 $W = \{0.2, 0.3, 0.5\}$ 。同时得到决策专家社会网络关系如图 1 所示。

Step2 决策专家对三个方案的三个属性分别进行打分，得到专家偏好矩阵，利用 (7)、(8) 式对决策专家进行聚类，得到四个聚类，记为 C_1, C_2, C_3, C_4 ，聚类结果如图 2 所示。利用 (10) 式得到聚集权重为 $\mu = \{0.5, 0.39, 0.08, 0.03\}$ ，群体偏好矩阵为 R 。

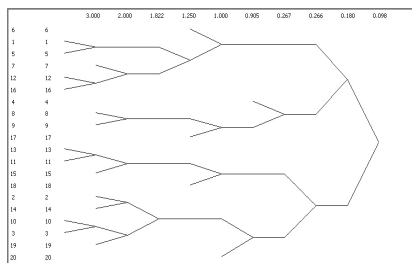


图 2 聚类结果

Step3 运用公式 (14)、(15)、(16) 分别得到每个聚类的信任风险系数、偏好风险系数、聚集风险程度，如表 1 所示：

表 1 风险测度结果

	C_1	C_2	C_3	C_4
θ_1	0.48	0.34	0.1	0.08
θ_2	0.26	0.32	0.27	0.15
θ	0.19	0.26	0.29	0.26

由每个聚集的聚集风险 θ 可得群体偏好风险矩阵 R^θ 。

Step4 通过获得的群体偏好矩阵为 R 和群体偏好风险矩阵 R^θ 。利用得分函数和精确函数，得到最终方案排序，如表 2 所示：

表 2 各方案得分

	偏好得分	风险得分	最终得分
方案一	0.3	0.58	-0.28
方案二	0.62	0.28	0.34
方案三	-0.34	0.6	-0.94

由表可知，最终选择方案二作为决策方案。

6. 案例分析

文章第 4 节的实例反映了本文所提方法的具体步骤，对社交网络媒体的信息进行关键词挖掘和分析是文章最大的工作量，这也是本文所提方法步骤中最耗时的阶段，在实际决策中将有赖于专业人才的数据处理能力。除此之外，后续的步骤只需将专家的偏好信息输入到计算机，通过设置好公式，可以快速得到最终结果，以最大程度满足应急决策的时间要求。

本文充分考虑了重大洪涝灾害发生时，大群体决策专家进行应急决策的高时间压力、交互性和复杂性。考虑到大群体意见难以在高时间压力下及时达成一致，提出了社交网络下重大洪涝灾害大群体应急决策风险测度的方法，考虑了决策专家主观不确定性和客观环境的风险，填补了社交网络应用到大群体风险测度的空白。同时应用大数据技术挖掘属性，通过对洪涝灾害发生时的社交网络媒体信息进行关键词挖掘、清洗、计算得到属性权重使决策更加全面科学。

同时本文还有很多地方需要完善，如未考虑专家的共识。如何将共识考虑到决策当中同时进行风险消解是下一步需要解决的问题。

参考文献

- [1] 计雷, 池红, 陈安. 突发事件应急管理. 北京: 高等教育出版社, 2006.
- [2] Xu X, Liang D, Chen X, et al. A risk elimination coordination method for large group decision-making in natural disaster emergencies. Human & Ecological Risk Assessment An International Journal, 2015, 21(5):1314-1325
- [3] Crooks A, Croitoru A, Stefanidis A, et al. #Earthquake: Twitter as a distributed sensor system. Transactions in Gis, 2013,

- 17(1):124-147.
- [4] Huang C. Internet of intelligences in risk analysis for online services. *Journal of Risk Analysis and Crisis Response*, 2011, 1(2):110-117.
- [5] Pan A. A Study on residents' risk perception in abrupt geological hazard. *Journal of Risk Analysis & Crisis Response*, 2012, 2(1):44-55.
- [6] Hristidis V, Chen S C, Li T, et al. Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 2010, 83(10): 1701-1714.
- [7] Liu Y, Liang C, Chiclana F, et al. A trust induced recommendation mechanism for reaching consensus in group decision making. *Knowledge-Based Systems*, 2016, 119(C):221-231.
- [8] Pérez L G, Mata F, Chiclana F. Social network decision making with linguistic trustworthiness-based induced OWA operators. *International Journal of Intelligent Systems*, 2015, 29(12):1117-1137.
- [9] Pérez L G, Mata F, Chiclana F, et al. Modelling influence in group decision making. *Soft Computing*, 2016, 20(4):1653-1665.
- [10] 宋光兴, 杨槐. 群决策中的决策行为分析. *学术探索*, 2000 (3): 48-49.
- [11] Robert AH, Mark R. *Introduction to Social Network Methods*. Department of Sociology University of California Riverside, 2005.
- [12] Wolfe A W. Social network analysis: methods and applications. *American Ethnologist*, 1995, 24(48):136-137.
- [13] 徐泽水. 区间直觉模糊信息的集成方法及其在决策中的应用. *控制与决策*, 2007(2):215-219.
- [14] 徐选华, 蔡晨光, 杜志娇, 王佩. 基于区间直觉模糊数的多属性多阶段冲突型大群体应急决策方法. *运筹与管理*, 2016, 25(4):12-22.
- [15] 屈克, 汤磊, 荣文静. 区间直觉模糊集熵的构造及其基本性质. *重庆文理学院学报(自然科学版)*, 2010, 29(3):21-24.
- [16] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008, 2008(10):155-168.