Advances in Applied Mathematics 应用数学进展, 2021, 10(12), 4528-4534
Published Online December 2021 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2021.1012482

Hans汉斯

基于社交媒体的河南暴雨灾害负面情绪和主题分析

李梦楠，汪明艳*

上海工程技术大学管理学院，上海

收稿日期：2021年11月27日；录用日期：2021年12月17日；发布日期：2021年12月31日

摘　要

自然灾害对社会的整体运行和经济、民生等造成了重大的冲击，如何在灾害发生时高效地应对并实施救援是政府和救援组织面对的难题。本文对2021年7月发生在河南的暴雨灾害的微博数据进行了研究，通过使用VADER进行情感极性分析，并基于消极情绪信息进行了LDA主题模型挖掘，以此来展现暴雨灾害发生时的网络舆情，为政府和救援机构提供有效的救灾信息，为救助策略提供必要的信息支持，为减少灾害损失提供了帮助。

关键词

负面情绪，灾情主题分析，微博，暴雨灾害

Analysis of Negative Emotions and Themes of Rainstorm Disaster in Henan Based on Social Media

Mengnan Li, Mingyan Wang*

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Nov. 27th, 2021; accepted: Dec. 17th, 2021; published: Dec. 31st, 2021

Abstract

Natural disasters have a great impact on the overall operation of society, economy and people's livelihood. How to effectively respond to and implement relief in the event of disaster is a difficult

*通讯作者。

文章引用: 李梦楠, 汪明艳. 基于社交媒体的河南暴雨灾害负面情绪和主题分析[J]. 应用数学进展, 2021, 10(12): 4528-4534. DOI: 10.12677/aam.2021.1012482

problem faced by the government and relief organizations. This paper studies the microblog data of rainstorm disaster in Henan province in July 2021. VADER was used for emotional polarity analysis, and LDA theme model was mined based on negative emotional information. The study shows the network public opinion when the storm disaster occurred, to provide effective disaster relief information to the government and relief agencies, give necessary information support for rescue strategy, and provide help to reduce disaster losses.

Keywords

Negative Emotion, Disaster Themes Analysis, Weibo, Rainstorm Disaster

1. 引言

社交媒体的快速发展提供了一个机会，使学者们可以以一种新的方式来研究民众的关注点。微博拥有 5.11 亿月活用户，能够提供信息的反馈，并利用时间线向用户提供实时更新的内容。因此，微博作为一个可靠的数据源的研究，开发潜力是显而易见的。它提供了一个独特的视角来了解用户的关注点。微博无需繁杂的计算处理就可以公开获取稳定和准确的数据。从收集的数据中获取知识是灾情中态势感知的主要关注点。因此，探索微博数据的过程在帮助人们面对自然灾害方面发挥着至关重要的作用。在自然灾害期间，人们试图立即收集并分享信息，使他们远离可能的危险状况[1]。通过分析实时大数据发展高水平的数据感知能力有助于提高灾害管理的效率。当灾害发生时，人民的生命安全是处在最高优先级的。但是受灾人员的位置、受灾的严重性很少能够在第一时间被确定。信息获取的滞后无法满足动态的应急管理需求，致使耗费大量救援资源。社交媒体的即时性特点可以在一定程度上缓解信息获取的问题。很多手机厂商将应急消息的通知集成在手机系统中，当灾害发生时可以在第一时间将消息推送给用户。比如，当地震预警网监测到会发生引起强烈晃动的地震时，手机厂商通常会采用弹出框的方式来提醒用户，向用户推送预警信息，并播放警报[2]。可以利用社交媒体广泛的参与度，多元化的沟通渠道，大量的时空信息数据迎合灾害救援的要求，也因此进入了学者的研究视角之中。

问卷调查是分析民众意愿的传统方法，然而，这种时间成本很高的方法应该在灾害发生后才能实施。参与数据收集的被调查者人数较少，需要耗费相当长的时间。社交媒体的发展为追踪民众舆论提供了一个非常好的机会。实时的社交媒体数据可以帮助灾害管理人员在自然灾害期间更好地发展态势感知。本研究采用 VADER 情绪分析和 LDA 主题建模来揭示民众关注的重点。我们在微博上获取了大量数据，并追踪了 2021 年河南 720 暴雨期间人们的负面情绪信息并对此进行研究。

2. 文献综述

微博数据被广泛应用于商业拓展、健康追踪和政治治理等领域。其中包括使用数据挖掘和自然语言处理对地震的损伤检测和评估，制作一个报告地震相关事件的时空概率模型，开发基于微博频率的地震检测算法，在微博上应用基于机器学习的分类算法检测地震[3] [4]，使用定性方法分析地震后人们的行为，使用关键词分析跟踪地震中和地震后的社会态度，并分析微博中谣言动态[5]。民众对于灾情风险的感知有助于提升其应对风险的能力。个人可以利用社交媒体在虚拟的网络空间了解真实的应急信息，从而从

容地应对以减少损失。另一方面，每个人的信息都是一个节点，一个个节点构成了信息平面，可以向外传达各种信息数据，研究者也可以利用这些数据加以分析，从而得到科学的结论。

面对灾难时，民众情绪的改变也会影响对风险的认知[6]。作为风险传播和感知的信息源，社交媒体进行的信息获取和交流可能会使关于疾病的沟通变得复杂，因为情绪对公众风险认知或后续行为的形成具有重要的作用。借助社会风险框架的放大和情绪稳定的概念，使用主题分析法，学者 Alexandra 等利用推特数据深入了解面对灾难事件的个人的应对措施[7]。对于诸如自然灾害信息，Polaris 系统[8]用于分析和预测用户情感轨迹，可以实时分析大量社交媒体内容。Yuan 等[9]采用 LDA 主题模型计算基于主题情感表达和权重。研究结果表明，情绪稳定性较低的陈述与风险感知的增强有关，而情绪稳定性的提高则与风险感知的减弱有关。但少量的微博并不足以度量一个人的情绪。研究集中在社区层面，也未考虑个人层面的变化。对大多数群体来说，社交媒体报道与风险认知水平呈相关关系，而传统媒体报道只有在少数群体中具有相关性。可见，社交媒体的使用可以增加民众的风险感知。

这些研究跟踪了用户在暴雨期间的行为，使用相关关键词和地理标记推文分析微博上的危机相关信息[10]，找到有用信息，利用关键词分析确定水灾相关推文中的信息类型。过去关于自然灾害中使用社交媒体的研究为灾害管理提供了有用的分析，但自然灾害不可预测的性质为灾害应对者提供了巨大的动力，通过探索新的视角来不断改进灾害管理。因此，特别是当涉及到人们的负面情绪时，可以通过更好地将微博感知融入灾害管理实践，以此来提升灾害管理效能。

3. 研究案例

本文基于 7.20 河南暴雨灾害的微博数据，对相关微博数据进行了研究，该研究由三个部分组成：数据收集与处理、情绪分析和负面情绪主题内容分析。

3.1. 数据收集与处理

利用 Python 使用微博 api 检索、收集数据，时间从 2021 年 7 月 19 日晚 8 时到 31 日晚 12 时，一共获取了 7058 条有效微博。选择这个时间段是因为它是暴雨灾害发生的重要时期，具有一定的数据代表性。数据集中删除了转发和包含 url 的微博，并从收集到的微博数据中删除了停用词和 emoji 表情符号等无关数据。

3.2. 情绪分析

情绪分析揭示了微博文本数据里面的情绪极性。在本研究中，我们对微博的类别没有任何先验性标注，并以此寻找带有情绪色彩的微博。情感分析是一种计算方法，可以识别一段文本的正面或负面的情绪程度，本文使用 VADER 方法进行情感分析，VADER 是一种专门针对社交媒体的情绪分析方法[11]，可以帮助识别三种类型的微博文本：积极的、中性的和消极的。目前 VADER 已经被广泛应用于一系列情感分析研究中，这些研究使用微博或推特上发布的内容进行情绪分析，了解灾害恢复能力。为了衡量灾害期间负面微博情绪的程度，我们得到了负面情绪微博 1184 条，中性微博 2498 条，积极微博 274 条，其他无法分辨的微博(捐款、无关) 3102 条。当人们接触到不可预测的暴雨灾害时，会表现出负面情绪，这些情绪来自于有不希望结果的威胁的存在。

城市暴雨灾害发生后，可以通过对微博内容的分析，估计公众情绪，根据舆情进行相应的应急管理。2021 年 7 月的强降雨和随后的积水内涝发生之后，用户立即在微博平台表达了自己的情绪反应。从 7 月 21 日到 7 月 22 日，公众基本没有负面情绪。虽然也有一些负面词汇，如“惨”、“祈福”、“郑州加油”等，但这些词汇所占的比例非常小。相反，出现了一些积极的词汇，如“哈哈”。可以看出，这一

阶段公众的负面情绪并不明显，主要原因可能是在暴雨初始阶段，相关灾害影响并没有立即出现。从 7 月 23 日到 7 月 29 日，与前一阶段相比，公众开始表现出强烈的负面情绪：“可怕”、“怎么办”这类消极词汇的比例较大。有一些消极词汇的比例很低，比如“造成死亡”等。因此，公众在这一阶段表现出强烈的负面情绪。积极情绪数据为综合得分大于等于 0.05，消极情绪综合得分小于−0.05，其余为中性情绪数据。相关情绪数量分布如表 1 所示。

Table 1. Distribution of emotional polarity number
表 1. 情感极性数量分布

情绪 / 日期	积极	中性	消极	无关
7.21~7.22	46	564	32	588
7.23~7.29	59	1679	741	1763
7.30~7.31	169	255	441	751

3.3. 基于 LDA 灾情主题分析

接下来，我们将研究内容集中于 1184 条负面微博中检测主题。LDA (Latent Dirichlet Allocation)是一个有效的、被广泛使用探索主题模型的方法[12]。该方法是由 Blei 等人在 2003 年提出的一种典型的无监督主题概率模型。它可以以概率分布的形式给出集合中每个文档的主题，在对多个文档进行分析并提取其主题后，可以根据主题对文本进行分类或聚类。LDA 假设文档主题的先验分布是 Dirichlet 分布，即对于任一文档 i，其主题分布 θ_i 为：

$$\theta_i = \text{Dirichlet}(\vec{\alpha}) \tag{1}$$

其中，α 为分布的超参数，是一个 k 维向量。

如果在第 k 个主题中，第 v 个词的个数为：n_k^v，则对应的多项分布的计数可以表示为：

$$\vec{n}_k = \left(n_k^1, n_k^2, \cdots, n_k^v\right) \tag{2}$$

从狄利克雷分布 β 中取样生成主题 $Z_{i,j}$ 的词语分布 $\phi_{Z_{i,j}}$，从词语的多项式分布 $\phi_{Z_{i,j}}$ 中采样最终生成词语 $\omega_{i,j}$，因此整个模型中所有可见变量以及隐藏变量的联合分布是:

$$p\left(\omega_i, Z_i, \theta_i, \phi \mid \alpha, \beta\right) = \prod_{j=1}^{N} p\left(\theta_i \mid \alpha\right) p\left(Z_{i,j} \mid \theta_i\right) p\left(\Phi \mid \beta\right) p\left(\omega_{i,j} \mid \phi_{Z_{i,j}}\right) \tag{3}$$

最终一篇文档的单词分布的最大似然估计可以通过将上式进行积分和求和得到：

$$p\left(\omega_i \mid \alpha, \beta\right) = \int_{\theta_i} \int_{\phi} \sum_{Z_i} p\left(\omega_i, Z_i, \theta_i, \phi \mid \alpha, \beta\right) \tag{4}$$

LDA 使用了常见的吉布斯采样(Gibbs Sampling)算法进行计算：首先对所有文档中的数据进行遍历，为每一个随机分配主题，即 $Z_{m,n} = k \sim Mult\left(1/K\right)$，其中 m 表示第 m 个文本数据，n 表示文本中的第 n 个词，k 表示主题，m 文本中主题数量的和，k 主题对应的 t 词的次数，k 主题对应的总词数。对所有文本中的所有词进行遍历，假如文本 m 的词 t 对应主题为 k，即先拿出当前词，之后根据 LDA 中概率分布 sample 出新的主题。迭代完成后输出参数矩阵 φ 和文档 - 主题矩阵 θ。

$$\phi_{k,t} = \left(n_k^{(t)} + \beta_t\right) \Big/ \left(n_k + \beta_t\right) \tag{5}$$

$$\theta_{m,k} = \left(n_k^{(t)} + \alpha_k\right) \Big/ \left(n_m + \alpha_k\right) \tag{6}$$

LDA 模型可以用来获取文本内容的主题分布信息。LDA 使用统计分布将微博文本中的每个词以不同的属性分配给每个主题。例如，在给定的微博语料库中，LDA 将“暴雨”、“施工”和“绕行”分配到一个以“交通”为主的主题中。主题内容分析之后对抽取的主题进行分析。该方法对 5 个主题进行标签和分类，分析它们的频率，随后我们尝试解释主题背后可能的原因。在 2021 年 7 月 20 日至 2021 年 7 月 31 日里，我们手动标注了负面情绪话题。例如，如果“道路”、“损坏”、“街道”和“堵车”都在一个主题中，我们就把这个主题标记为“交通或基础设施损坏”。为了更详细地揭示暴雨灾害对微博用户的响应，我们对每个主题进行了总结。由于篇幅限制，我们只给出前 5 个常用术语。分类结果如图 1 所示。

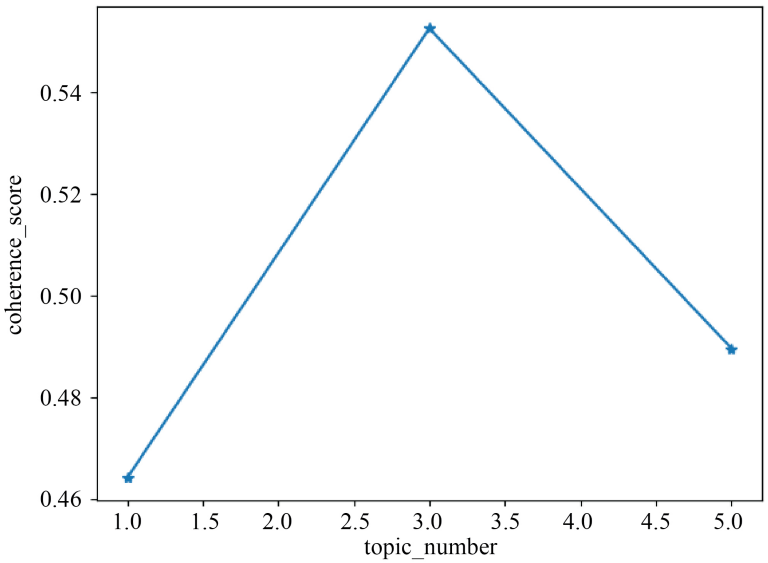

Figure 1. Results of classification
图 1. 分类结果

通过执行 LDA 模型，生成五个主题和每个主题对应的多个主题词，通过仔细观察和比较所有的主题和主题词，发现一些相关的主题可以归为一类。无意义的话题可以分为其他类别。随后，所有主题可分为 5 个类别：救援、受灾民众、交通或基础设施损坏、救灾物资、其他。每条微博都与话题一一对应，每个话题都可以被划分到相应的类别中，所以每条微博也可以根据话题的加标签结果被划分到一个类别中。根据以上分类过程，所有的微博被分为五类(由于“其他”类别话题与暴雨灾害无关，所以不在研究范围之内)。得到的结果如表 2 所示。

Table 2. Thematic distribution based on negative emotions
表 2. 基于负面情绪的主题分布

主题词 / 类别	1	2	3	4	5
救援	救援队	急需	郑州	辉县	联系
受灾民众	第一人民医院	老人	学生	孩子	孕妇
交通或基础设施损坏	积水	绕行	决堤	交通管制	倒塌
救灾物资	饮用水	食物	帐篷	救生艇	沙袋

在第一类的类别上，主题是“救援”。一些微博试图提高人们对暴雨灾害情况的防范意识，并建议公众不要外出，提醒他们应该注意安全。其他微博表明了明的救援目的或方式。除了城市名称“郑州”、“辉县”等外，还伴有“救援队”、“急需”、“联系”等。原因可能是由于过多的雨水淹没了灾民的居住房屋，导致这部分民众需要救援组织进行救援。

在受灾民众类别下，可以看到人们通过微博交流的主题词集中于“第一人民医院”。在水灾发生时，第一人民医院的大量病人因为病情原因无法快速转移，导致积水淹没之时被困在病床上，因此导致大量的求救微博发送关于转运病人的数据。“老人”、“学生”、“孩子”等群体因为行动不便或其他原因也出现了高频次的数据产生，这对生命救助来说是至关重要的。灾害管理中心在第一时间识别这类微博数据可以最大程度的制定精准救治的方案。

暴雨严重破坏了交通和基础设施，诸如“积水”、“绕行”、“决堤”、“交通管制”、“禁止通行”，公共场所：“房屋”、“倒塌”、“淹没”，这也在一定程度上表达了受灾严重民众想要获救的急切心情，还有救援组织赶往受灾地的紧迫性，对于救援组织来说是需要优先考虑给予帮助的。

在救灾物资类别中，受到关注的重要信息有“饮用水”、“食物”、“帐篷”，在统揽灾区形势后，“救生艇”、“沙袋”等救援物资也被提及，在以人民生命健康为首要救援任务的灾情面前，这些信息传播有利于减少不必要的对生命的安全威胁。

4. 结论和建议

本文将暴雨灾害期间微博用户的情绪按照积极、中性和消极进行区分，并使用消极情绪数据进行LDA的主题分类，得到五个类别。对相关类别可能状况进行了分析。本研究在一定程度上提供了暴雨灾害造成的破坏和灾民需求的宏观描述。这一结果可以帮助政府和救援机构更好地进行救灾资源分配并施以救助。

与其他研究相比，这项研究有几个优势。首先，将情绪分析和主题建模方法相结合。为政府和救援组织在灾害发生时提供一个良好的视角来了解整个发展趋势。其次，本研究专注于消极情绪的主题分析，而不是所有的情绪。这可以避免在救援的关键时间点上浪费不必要的资源。每个用户都是社交媒体的传播者，可以生成有价值的即时的数据，可以将态势感知更好地应用于灾害管理中。

在未来，我们希望能够分析其它情绪的微博，或分析更加细粒度的微博数据，这样可以辅助救援机构更加精准地进行灾害救援。VADER 方法目前主要应用于基于英文的情感识别研究，针对中文的训练数据还比较少，还有很大的研究空间。此外，将根据地理位置进行区域性关联研究，从空间角度探寻社交媒体在灾情发生时的作用。救援组织应该探索不同的方法来了解灾情：政府也可以追踪受影响最严重的地区，派遣救援队，进行救援行动；借助探索性的数据分析、主题建模可以有效评估灾害传播程度，识别受灾的微博用户，并利用它们向更多的灾民伸出援手；探索消极情绪细粒度识别，有利于救援组织合理、高效地进行自然灾害管理。

参考文献

[1] Castillo, C. (2016) Big Crisis Data: Social Media in Disasters and Time-Critical Situations. Cambridge University Press, Cambridge.

[2] 路鹃, 陈峦明. 公共危机事件中手机媒体的传播效果分析——以“7·21 北京特大暴雨灾害”为例[J]. 新闻界, 2012(20): 45-49.

[3] Sakaki, T., Okazaki, M. and Matsuo, Y. (2010) Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. *Proceedings of the* 19*th International Conference on World Wide Web*, Raleigh, 26-30 April 2010, 851-860. https://doi.org/10.1145/1772690.1772777

[4] Robinson, B., Power, R. and Cameron, M. (2013) A Sensitive Twitter Earthquake Detector. *Proceedings of the* 22*nd*

International Conference on World Wide Web, Rio de Janeiro, 13-17 May 2013, 999-1002. https://doi.org/10.1145/2487788.2488101

[5] Hashimoto, T., Shepard, D.L., Kuboyama, T., *et al.* (2021) Analyzing Temporal Patterns of Topic Diversity Using Graph Clustering. *The Journal of Supercomputing*, **77**, 4375-4388.

[6] Han, X. and Wang, J. (2019) Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. *ISPRS International Journal of Geo-Information*, **8**, 185. https://doi.org/10.3390/ijgi8040185

[7] Bec, A. and Becken, S. (2019) Risk Perceptions and Emotional Stability in Response to Cyclone Debbie: An Analysis of Twitter Data. *Journal of Risk Research*, No. 1, 721-739. https://doi.org/10.1080/13669877.2019.1673798

[8] Yoo, S.Y., Song, J.I. and Jeong, O.R. (2018) Social Media Contents Based Sentiment Analysis and Prediction System. *Expert Systems with Applications*, **105**, 102-111. https://doi.org/10.1016/j.eswa.2018.03.055

[9] Yuan, F., Li, M. and Liu, R. (2020) Understanding the Evolutions of Public Responses Using Social Media: Hurricane Matthew Case Study. *International Journal of Disaster Risk Reduction*, **51**, Article ID: 101798. https://doi.org/10.1016/j.ijdrr.2020.101798

[10] De Albuquerque, J.P., Herfort, B., Brenning, A., *et al.* (2015) A Geographic Approach for Combining Social Media and Authoritative Data towards Identifying Useful Information for Disaster Management. *International Journal of Geographical Information Science*, **29**, 667-689. https://doi.org/10.1080/13658816.2014.996567

[11] Hutto, C. and Vader, G.E. (2014) A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, **8**, 216-225.

[12] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, **3**, 993-1022.