

Building Real-Time Monitoring of Infectious Diseases and Corresponding Public Opinion Based on Naive Bayesian Algorithm and Crowd Sensing

Xuan Tang, Huajie Shen

Business School, Sichuan University, Chengdu Sichuan
Email: keeliatang@yeah.net

Received: Jul. 3rd, 2018; accepted: Jul. 13th, 2018; published: Jul. 20th, 2018

Abstract

[Background/Meaning] The Weibo users are regarded as perceptual units to form a Crowd-sensing network, to produce intuitive and visual description of the distribution, transmission and development of infectious diseases and corresponding public emotions. [Method/Process] The study is divided into five parts: first, crawl the microblog posts associated with the event; second, use the data filtering model to filter the set of posts that have low correlation with the event; meanwhile, query and detect the position based on OpenStreetMap to return to the real position set; then, based on JIEBA word segmentation system of Python and Bayesian algorithm, achieve keyword emotional prediction; finally, through the emotion feedback model μ and epidemic monitoring model ϑ , generate emotional curve and epidemic map, and make a case study. [Result/Conclusion] By this method (containing four models), epidemic emotion maps of infectious diseases and public emotion curves can be generated from Weibo. Among them, the map shows the epidemic situation and spread of each location set, and the emotion curve gives the public feedback changes. The study found that the more cases and the higher flow of population, the more negative the emotional performance. Excellent medical conditions and positive social media campaigns have a positive effect on public emotion.

Keywords

Crowd Sensing, Naive Bayes, Epidemic Monitoring, Emotional Recognition, Web Crawler

基于朴素贝叶斯分类与群智感知实现对传染病与相应舆论的实时监控

唐璇, 沈华杰

四川大学商学院, 四川 成都
Email: keeliatang@yeah.net

收稿日期: 2018年7月3日; 录用日期: 2018年7月13日; 发布日期: 2018年7月20日

摘要

[背景/意义]将微博用户看作感知单元, 形成群智网络, 从公众情感角度对传染病的分布、传播与发展进行简洁直观、可视化的描述。[方法/过程]研究分五个部分。首先, 爬取与事件相关的微博帖集; 其次, 利用数据过滤模型 α 过滤与事件关联度不高的帖集; 同时, 基于OpenStreetMap查询并检测位置以返回真实位置集; 然后, 基于Python的JIEBA分词系统, 进行分词和词频处理, 并基于朴素贝叶斯算法实现对关键词的情感预测, 得到情感集; 最后, 通过情感反馈模型 μ 与疫情监控模型 σ 产生情感曲线与疫情地图, 并进行案例研究。[结果/结论]通过这种方法(内含四个模型), 可以从微博上生成传染病事件的疫情情感地图与公众情感曲线。其中, 地图展示了各位置集的疫情严重程度与传播情况, 情感曲线给出了公众反馈随时间的变化。研究发现, 病例数越多与人口流动性较大的地区情绪表现更为消极, 优良的医疗条件和积极的社交媒体宣传对公众情绪有一定积极效果。

关键词

群智感知, 朴素贝叶斯, 疫情监控, 情感识别, 网络爬虫

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自2014年9月底, 埃博拉病毒在利比里亚之后, 全球范围对于传染病的担忧持续蔓延, 移民检查和医疗系统处理埃博拉病毒的能力受到广泛质疑。对于传统的公共卫生监测系统而言, 难以发现和监测传染病的传播规律以及公众对相关事件的情感变化。现有的监视方法, 如问卷和临床试验, 覆盖的人数与地理范围有限, 结果往往出现较大延误, 这对于疫情传播的控制与网络舆论的管理是不利的。为补充当前监视系统, 需开发一种新颖的模型。能够跟踪公众情绪的实时统计并基于公众情感变化提供疫情预警及监控, 帮助政府从多角度做出决策, 有效阻止或回应潜在的社会舆论及公共健康危机。

网络创造了前所未有的资源, 可用于追踪传染病对公共健康的威胁。微博是中国最大的社交平台之一, 是产生和分享实时内容与话题的重要媒介之一, 吸引了大量的主题社区[1], 舆论传播和内容分析[2]。研究将微博用户视为社交传感器[3], 用于监测传染病传播与公众舆论的对应情感变化趋势。

2. 文献综述

2.1. 群智感知

群智感知是一种利用移动端有效收集处理数据的创新物联网感知模式。该模式下, 大量用户的移动设备作为基础单元, 构成群智感知网络, 从而实现大规模的网络数据收集[4], 这种模式可以让普通人使用从社会群体的移动设备收集或生成的数据[5], 并可通过云端的异构众包数据来提取隐藏的信息。近年,

具有 GPS 与其他传感功能的交通工具逐渐普及至社会各角落, 这促进了群智感知模式的形成, 在网络营销领域, Jidong 等人[6]利用爬虫技术结合群众网络, 对电子香烟的销售做了大量的统计分析, 并得出了一些销售结论。在交通领域[7], 赵东等人利用传染路由机制(ERF)和二分喷射等待机制(BSWF)设计了新的数据转发规则, 用以解决汽车群智感知网络进行数据收集节能问题。旅游管理领域, Chen 等人[8]通过利用像一些大型照片网站来解决个性化旅游产品的推荐问题。在环境监测方面, Pankratius 等人[9]提出一种全新的架构, 使用移动设备建立了地球近空间环境的监测网络, 可利用现有的 GPS 信号预测电离层总电子含量。

2.2. 基于朴素贝叶斯的情感识别

在基于机器学习的情感分类中, 多数的研究成果表明机器学习方法中以朴素贝叶斯分类算法效果最好, 其原理是使用各类别的先验概率和类别出现特定特征的条件概率来预测出现这些特征的个体的所属类别。Ana Carolinade 等人[10]利用贝叶斯分类结合 BIGFIVE 模型实现了对微博用户的高精度人格预测。彭蔚喆等人[11]初步解决面向中文微博文本的主观句识别与情感分类问题, 他们使用朴素贝叶斯方法和支持向量机方法来识别主观句, 并得出朴素贝叶斯方法对主观句的识别效果更好的结论。张扬等人[12]提出了一种基于朴素贝叶斯模型的网络负面信息预警策略, 针对大规模网络碎片化信息的情感识别提高其执行效率。

2.3. 疫情监控地图

疫情监控地图的优越之处在于视觉效果与概括性, 能够直观地展示变化趋势或者传播范围。XiangJi 等人[13]基于 Twitter 平台, 利用三种机器学习的方法绘制了公共健康的监控地图, 用于确定公众关注的高峰与传染病的传播。Rosen 等人[14]提出了一个通过群智感知有效地监控企业 WiFi 网络的系统, 通过消费者移动设备来收集数据, 并生成一个性能高低地图。在噪声监测领域, 吴文乐[15]等人以智能手机为感知终端, 实现噪声污染数据的采集, 并在地图上用不同颜色表示噪声污染等级, 直观展示噪声污染情况。葛小三等人[16]利用网络数据挖掘技术, 通过网络热点事件关注群体地理分布的可视化表达, 为舆情综合管理提供决策支持。

2.4. 基于社交媒体的疫情舆论

基于社交媒体平台的飞速发展, 公众更加习惯于在互联网上发表自己对某焦点事态带有倾向性的观点, 通过对网络舆论进行挖掘分析, 可以更准确迅速地了解舆论走向, 特别是对于传染病一类公众关注度较高的事件。Broniatowski 等人[17]提出可以通过社交媒体对流感感染进行检测, 在季节和地理颗粒度上均显示出较好的性能; Kim 等人[18]通过研究灾害管理, 发现社交媒体对于发送警报、识别关键需求和聚焦响应具有重要作用; 梁芷铭等人[19]提出了基于情感本体的网络舆情观点挖掘模型, 有利于及时控制舆情倾向性。

2.5. 评述

通过对相关文献的现状梳理, 国内外基于数据挖掘与社交媒体在销售(Jidong, 2018)、旅游管理(Chen, 2013)、环境监测(Pankratius, 2014)、灾害管理(Kim, 2018)等领域的研究较多, 而针对于传染病传播与舆论控制的定量研究较少, 但相关研究仍为本文提供了一些可借鉴的方法, 同时成熟的数据挖掘算法对本文研究提供帮助。本文利用贝叶斯情感识别对微博平台上的“H7N9”案例评论数据进行分析, 并基于时间和地理双维度可视化展示公众舆情传播与发展态势, 帮助政府机构全面、及时、准确地了解疫情发展状况, 并基于社交媒体运作有效控制舆论。

3. 框架概述

本文研究过程主要分五个部分。第一部分, 通过搜索关键种子词集爬取与事件相关的一组微博帖子; 第二部分, 利用数据过滤模型 α 将与事件关联度不高的帖子过滤, 同时基于 OpenStreetMap 查询并检测位置以返回真实位置, 以提高实验效率与数据精度; 第三部分, 通过 Python 下的 JIEBA 分词, 对数据进行关键词和词频的处理, 得到一组与情绪相关的关键词词集; 第四部分, 基于朴素贝叶斯算法, 利用训练集进行训练并预测, 从而实现对关键词的情感赋值; 第五部分, 通过情感反馈模型 Φ 与疫情监控模型 φ 产生情感曲线与疫情地图, 并进行案例研究。文章研究框架如图 1 所示。

4. 模型概述

4.1. 数据过滤模型 α

1) 变量定义

令 $Y(y_1, y_2, \dots, y_n)$ 表示疫情事件集合, $S(s_1, s_2, \dots, s_n)$ 表示相关搜索种子集, $RM(rm_1, rm_2, \dots, rm_n)$ 表示相关的帖子集, $E(e_1, e_2, \dots, e_n)$ 表示采集数据集, $K(k_1, k_2, \dots, k_n)$ 表示关键词集, $T(t_1, t_2, \dots, t_n)$ 表示帖子时间集, $P(p_1, p_2, \dots, p_n)$ 表示帖子位置集, z_i , a_i 分别表示 rm_i 帖子的转发量和获赞数。

2) 输入与输出

输入(j): 给定疫情事件 Y 和搜索种子词集 S 返回的一组相关微博帖子。

输出(i): 数据集 E 被生成。数据集变量包含设置关键词 K , 时间集 T , 和位置信息 L , 帖子转发量 Z , 获赞数 A 。

3) 规则定义

数据过滤模型 α 由三条规则构成。由于在实际检索与爬取过程中会获得部分与事件关联度不高的帖子, 为提高实验效率与摘要质量, 文章建立了下述 3 条规则来过滤来自微博的无关帖子, 最终获得相关的帖子集(RM)。

规则 1:

$$\forall E_i \in E \rightarrow (\exists s_j \in S \wedge s_j \in rm_j)$$

发布应包含疫情事件的种子词, 反映事件的基本语义。

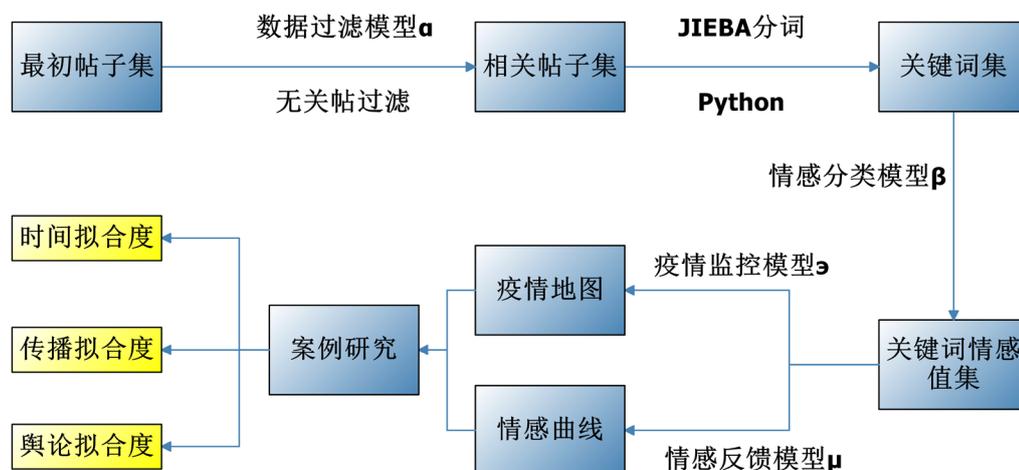


Figure 1. Research framework

图 1. 研究框架

规则 2:

$$\forall E_i \in E \rightarrow (\exists t_y \in T \wedge \exists t_{rm} \in T \wedge t_{rm} \in t_y)$$

相关发布对应的时间应在相关事件发生后。

规则 3:

$$\forall E_i \in E \rightarrow \{\exists z_j \in Z \wedge a_j \in A \wedge (z_j \geq 1000 \vee a_j \geq 1000)\}$$

发布应为社交平台认可的热门发布, 热门贴意味着社会群体对其有较高的认可度。

4) 模型过滤性能

文章将 F1-measure 值用作对过滤模型的主要评估方式。选择以“H7N9”为种子词进行搜索返回的 700 个帖子作为数据集。经过测试后发现, 合并 3 条规则后可以获得最高的召回率 92.0%, 约 8% 的相关发布被过滤, 说明并不是所有相关的发布都对事件描述有用。结合 3 条规则后获得了高于任何单条规则的 F1-measure 值, 高达 85.7%, 这表明模型对于无关帖的过滤性能是很好的。 α 模型过滤性能如图 2 所示。

4.2. 情感分类模型 β

1) 变量定义

令 $D(d_1, d_2, \dots, d_n)$ 表示分词集, $\Phi_i (1 \leq i \leq n)$ 表示关键词 d_i 的情感值, M, H, L 分别表示中性、积极、消极类情感, $f_i (1 \leq i \leq n)$ 表示分词 d_i 的出现频率。 ϕ 表示各集合情感值。

2) 输入与输出

输入(j): 由数据集 E 处理得到的分词集 D 。

输出(i): 各集合情感值 ϕ 。

数据源模型过滤性能

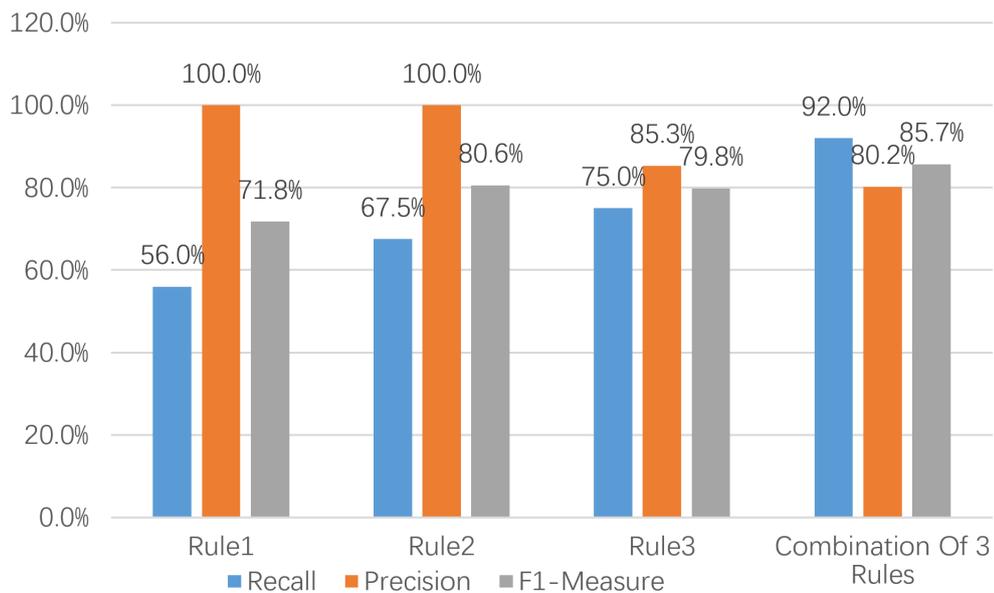


Figure 2. Filtering quality

图 2. 过滤性能

3) 规则定义

对于社会疫情事件, 情感值是一个重要的构成因素。情感值可以反映社交平台群体对于该事件的情感倾向, 文章的情感赋值基于朴素贝叶斯分类算法, 由以下 2 条规则构成。

规则 1:

$$\forall \phi_i \in \phi \rightarrow \{\exists k_j \in K \wedge k_j \in (L \vee H \vee M)\}$$

关键词根据情感属性分为积极类 H , 中性类 M , 消极类 L 三类, 分别赋值 1, 0, -1。

规则 2:

$$\forall \phi_i \in \phi \rightarrow \left\{ \phi_j = \frac{\sum_{k=0}^{n_j} (\phi_j * f_j)}{n_j} \wedge \phi_j \in [-1, 1] \right\}$$

集合情感值为所属集合的所有关键词情感值的加权平均数, 其取值范围为 $[-1, 1]$ 。

4) 模型性能

在对 2017 年的“H7N9”事件的相关帖集进行分词后, 从分词集中随机抽取了 191 个词语, 作为情感识别的测试集。

关键词根据情感属性分为积极类 H , 中性类 M , 消极类 L 三类, 分别赋值 1, 0, -1。对其进行情感定义, 并将其录入 WEKA 软件, 基于朴素贝叶斯算法建立分类模型, 模型性能如表 1 所示, F 值为 0.780, 分类精准度达 78.01%, 性能较好。

4.3. 情感反馈模型 μ

1) 变量定义

令 $MO(mo_1, mo_2, \dots, mo_n)$ 表示关键词月集合, ϕ 表示月集合情感值, \vec{E} 表示情感值变化曲线。

2) 输入与输出

输入(j): 关键词月集合 MO 、情感值 Φ 与词频 F 。

输出(i): 情感变化趋势曲线 \vec{E} 。

3) 规则定义

社会群体对于某事件的情感倾向会随事件的发展发生变化, 更具象化的体现是情感值随时间变化。

规则 1:

$$\exists e_i \in \vec{E} \rightarrow \{\exists e_j \in \vec{E} \wedge unit = (MONTH')\}$$

情感反馈曲线的变化以月为单位。

规则 2:

$$\exists e_i \in \vec{E} \rightarrow \left\{ \exists \phi_j \in \phi \wedge \phi_j = \frac{\sum_{k=0}^{n_j} (MO_j * f_j)}{n_j} \right\}$$

月情感值为月集合所有关键词情感值的加权平均数。

Table 1. Model performance

表 1. 模型性能

	Precision	Recall	F-Measure	AUC	AUPR	Correct	Incorrect
Result	0.721	0.780	0.729	0.833	0.770	78.01%	21.99%

4.4. 疫情监控模型

1) 变量定义

令 $PO(po_1, po_2, \dots, po_n)$ 表示关键词位置集合, $ep_i (1 \leq i \leq n)$ 表示 i 位置的疫情值, MA 表示疫情监控地图。

2) 输入与输出

输入(j): 关键词位置集 PO 、情感值 Φ 与词频 F 。

输出(i): 疫情监控地图 MA 。

3) 规则定义

社会群体对于某事件的情感倾向会随事件的发展发生变化, 更具象化的体现是情感值随时间变化。

规则 1:

$$\forall ma_i \in MA \rightarrow \{\exists po_j \in PO \wedge num(PO) = 34\}$$

疫情监控地图的共分为 34 个位置集, 分别对应中国的 34 个省级行政单位。

规则 2:

$$\forall ma_i \in MA \rightarrow \{\exists pm_j \in PM \wedge unit = (PROVINCE')\}$$

每个位置集的情感值为对于省级行政单位集合所有关键词情感值的加权平均数

规则 3:

$$\forall ma_i \in MA \rightarrow \{\exists \in PM \wedge \exists ep_j \in EP \wedge ep_j = (pm_j * 2 + 1) * 255\}$$

每个位置集的疫情值对应不同颜色, 颜色越绿表示情感越积极, 颜色越黑表示情感消极。这里引入 RGB 颜色模型, 其中 R、B 值固定, G 值随情感值变化, 经过实际测试, 研究发现疫情值与位置情感值的最佳换算公式如上, 经过该公式的计算使颜色变化更加明显。

规则 4:

$$\forall ma_i \in MA \rightarrow \{\exists e_j \in \ddot{E} \wedge unit = (YEAR')\}$$

疫情监控地图情感值单位为年, 表示一年中各位置集疫情严重程度。

5. 案例研究

研究选择了“H7N9”事件作为案例展示文中模型简洁、可视化、拟合度高等优点, 数据集(E)源自以“H7N9”为种子词返回的 700 条热门贴, 经过无关帖过滤、情感预测、情感曲线、疫情地图绘制后, 生成结果如图 3、图 4 所示。

1) 情感反馈模型结果

以纵坐标代表公众对于事件的情感值, 横坐标为时间轴, 各点分别对应各时间点公众的情感值。我们可以看出, 这种情感变化往往是非线性的, 所以, 时间与情感的关联性需要针对具体事件进行分析, 如图 3 所示。

2017 年 1 月~2017 年 3 月, 山东、玉林两例重症 H7N9 病毒感染患者被成功治愈, 公众对于“H7N9”的情感反馈逐渐上升, 由-0.248 上升至-0.164, 达到 2017 年的第一个极值点。

2017 年 3 月~2017 年 6 月, 国家卫生计生委发布公告, 多例患者已被治愈, H7N9 的传播已得到控制, 社交媒体用户的情感倾向持续上升, 由-0.164 上升至-0.047。

2017 年 6 月~2017 年 7 月, 夏季是各类传染病的高发季, 全国再次出现约 10 例确诊为 H7N9 病例, 公众的情感值小幅下降至-0.066。



Figure 3. “H7N9” event emotion curve

图 3. “H7N9”事件情感曲线

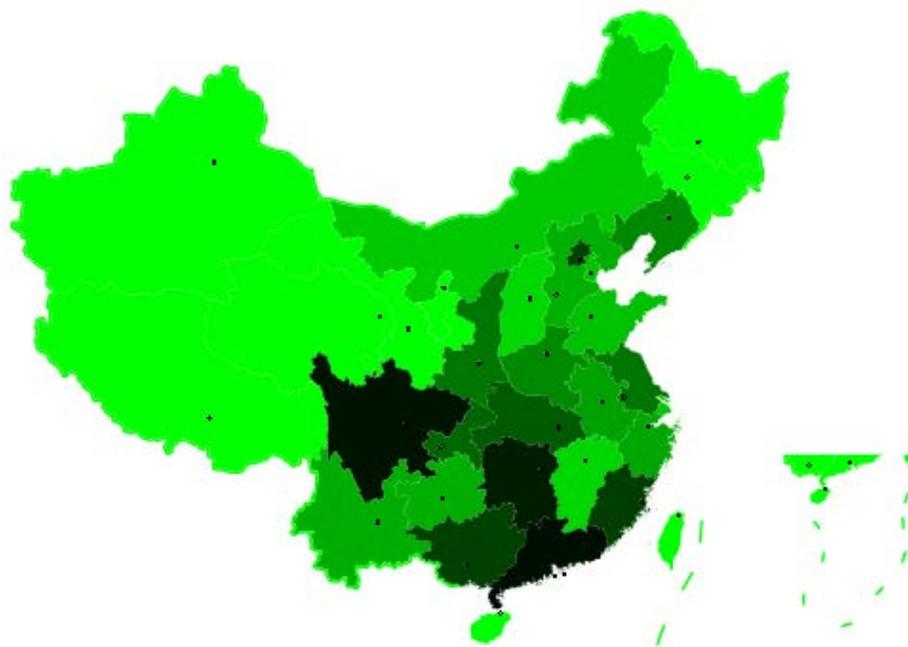


Figure 4. Epidemic emotion map

图 4. 疫情情感地图

2017年7月~2017年9月,复旦大学基础医学院应天雷课题组发现抗H7N9禽流感新型高活性全人源抗体,引发了微博的热烈讨论,群众情感值上升至-0.026,为2017年情感倾向的峰值。

2017年9月~2018年1月,国家卫生计生委发布《全国法定传染病疫情报告》,引发了微博用户的热烈讨论,公众情感值再次下降,至-0.207,后于2018年1月出现回升,至-0.076。

2017年1月~2018年2月,每年的12月至次年3月间,是人感染禽流感高发时段,此期间涉禽市场的阳性率高于水平,国内再次出现多例患者确诊为H7N9病例,社交用户的情感值持续走低。

基于情感曲线,研究认为通过模型 β 与 μ 生成的情感曲线与事件的真实发展有较高的拟合度,各极值点分别与疫情事件的转折点对应。

2) 疫情监控模型结果

本文共划分了 34 个位置集, 分别对应中国的 34 个省级行政单位, 每个位置集的疫情值对应不同颜色, 颜色越绿表示情感越积极, 颜色越黑表示情感消极, 如图 4 所示。

地理分布位置上, 沿海城市诸如广西、广东、福建等, 与一二线城市例如北京等, 表现出更为消极的情绪, 这与广东出现首例 H7N9 病例、沿海城市死亡人数较多有关。对数据进行统计后, 如表 2, 研究发现一二线城市人口流动性较大, 因此 H7N9 患者较多, 例如北京(13 例)处于内陆但几乎与广东(15 例)持平。

宏观上, 对情感值与案例数进行相关性检验后, 如表 3, 发现病例出现较多的城市公众的情绪表现越消极(皮尔森相关系数为 -0.832), 如四川(14 例)、福建(11 例)等城市同样表现较为消极, 与研究的假设情况相符, 病例出现越多的位置集讨论热度越高。

局部方面, 个别区域出现较为反常的情绪表现, 比如贵州(9 例)表现出与河北(1 例)较为相近的情感值, 经资料考证, 研究认为这和遵义一名重症患者成功被治愈与“贵州同城报”、“贵州微生活”等政务博主的积极报道与宣传有关。

基于地理疫情情绪地图, 病例数越多的地区情绪表现更为消极, 但是优良的医疗条件和积极的社交媒体宣传对公众情绪有一定正向引导效果。

Table 2. The number of cases corresponding to the emotional value of each location set (the position set of case number 0 has been omitted)

表 2. 各位置集情感值对应病例数(案例数为 0 的位置集已省略)

地点	情感值	真实病例数	地点	情感值	真实病例数
山西	-0.0713	1	辽宁	-0.22972	2
江西	-0.07927	1	陕西	-0.26634	7
内蒙古	-0.10566	1	江苏	-0.26762	6
上海	-0.10864	1	重庆	-0.30015	8
天津	-0.11611	2	湖北	-0.32229	9
山东	-0.11808	3	北京	-0.36897	13
河北	-0.15322	1	广西	-0.37124	7
贵州	-0.15567	9	福建	-0.37772	11
云南	-0.16056	2	湖南	-0.44218	19
浙江	-0.17119	2	厦门	-0.44623	5
安徽	-0.18141	1	四川	-0.46083	14
河南	-0.22898	2	广东	-0.46971	15

Table 3. Emotional value-case number correlation coefficient

表 3. 情感值 - 病例数相关系数

		情感值	真实病例数
情感值	Pearson 相关性	1	-0.832^{**}
	N	34	34
真实病例数	Pearson 相关性	-0.832^{**}	1
	N	34	34

**在 0.01 水平(双侧)上显著相关。

6. 结论

随着埃博拉等传染病在全球范围的传播, 传统的公共卫生监测系统饱受质疑, 随之出现的社会舆论浪潮让各国疲于处理。现有的传染病监测方法, 有覆盖面小、延时较长等特点, 难以发现和研究全新传染病的传播规律。文章提出的疫情监控方法侧重于从公众情感角度对传染病的分布、传播与发展进行简洁直观、可视化的描述, 通过这种模型(内含四个模型), 可以从微博上生成传染病事件的疫情情感地图与公众情感曲线。其中, 地图展示了各位置集的疫情严重情况与传播情况, 情感曲线给出了公众反馈随时间的变化。在案例研究中展示了模型的描述能力与较高的模型性能, 研究发现, 病例数越多与人口流动性较大的地区情绪表现更为消极, 优良的医疗条件和积极的社交媒体宣传对公众情绪有一定积极效果。研究的模型成果包含了事件基本内容、传播趋势与公众情感反馈, 能够向相关部门提供简洁有效的事件概述, 对于传染病疫情控制及舆论管理决策是有显著帮助的。

研究对于情感曲线的绘制处理仍有不足之处, 曲线的变化趋势较真实事件有一定的滞后时间, 文章认为可通过细化情感时间单位降低滞后程度。同时, 在爬取帖集的过程中, 研究发现, 图片、视频等多媒体集较文本能够更直观的描述疫情变化, 如果技术允许, 可将多媒体中的信息与文本整合, 更精准地实现情感分类预测。

参考文献

- [1] Hong, L., Gurumurthy, S., Gurumurthy, S., *et al.* (2012) Discovering Geographical Topics in the Twitter Stream. *Proceedings of the 21st International Conference on World Wide Web*, Lyon, 16-20 April 2012, 769-778. <https://doi.org/10.1145/2187836.2187940>
- [2] Lehmann, J., Gonçalves, B., Ramasco, J.J., *et al.* (2011) Dynamical Classes of Collective Attention in Twitter. *21st ACM International Conference on World Wide Web (WWW)*, New York, 251-260.
- [3] Sakaki, T., Okazaki, M. and Matsuo, Y. (2010) Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. *Proceedings of the 19th international conference on World Wide Web*, Raleigh, 26-30 April 2010, 851-860. <https://doi.org/10.1145/1772690.1772777>
- [4] 赵东, 马华东. 群智感知网络的发展及挑战[J]. 信息通信技术, 2014(5): 66-70.
- [5] Guo, B., Chen, H., Yu, Z., *et al.* (2015) FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing. *IEEE Transactions on Mobile Computing*, **14**, 2020-2033. <https://doi.org/10.1109/TMC.2014.2385097>
- [6] Huang, J., Kornfield, R., Szczypka, G. and Emery, S.L. (2014) A Cross-Sectional Examination of Marketing of Electronic Cigarettes on Twitter. *Tobacco Control*, **23**, iii26-iii30. <https://doi.org/10.1136/tobaccocontrol-2014-051551>
- [7] 赵东. 移动群智感知网络中数据收集与激励机制研究[D]: [博士学位论文]. 北京: 北京邮电大学, 2014.
- [8] Chen, Y.Y., Cheng, A.J. and Hsu, W.H. (2013) Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos. *IEEE Transactions on Multimedia*, **15**, 1283-1295. <https://doi.org/10.1109/TMM.2013.2265077>
- [9] Pankratius, V., Lind, F., Coster, A., *et al.* (2014) Mobile Crowd Sensing in Space Weather Monitoring: The Mahali Project. *IEEE Communications Magazine*, **52**, 22-28. <https://doi.org/10.1109/MCOM.2014.6871665>
- [10] Lima, A.C. and De Castro, L. (2013) Multi-Label Semi-Supervised Classification Applied to Personality Prediction in Tweets. *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*, Ipojuca, 8-11 September 2013, 195-203. <https://doi.org/10.1109/BRICS-CCI-CBIC.2013.41>
- [11] 彭蔚喆. 面向中文微博文本的情感识别与分类技术研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2014.
- [12] 张扬, 崔晨阳. 基于朴素贝叶斯模型的一种网络负面信息预警策略研究[J]. 图书馆杂志, 2014, 33(8): 78-82.
- [13] Ji, X., Chun, S.A., Wei, Z. and Geller, J. (2015) Twitter Sentiment Classification for Measuring Public Health Concerns. *Social Network Analysis and Mining*, **5**, 1-25. <https://doi.org/10.1007/s13278-015-0253-5>
- [14] Rosen, S., Lee, S.J., Lee, J., *et al.* (2014) MCNet: Crowdsourcing Wireless Performance Measurements through the Eyes of Mobile Devices. *IEEE Communications Magazine*, **52**, 86-91. <https://doi.org/10.1109/MCOM.2014.6917407>
- [15] 吴文乐, 郭斌, 於志文. 基于群智感知的城市噪声检测与时空规律分析[J]. 计算机辅助设计与图形学学报, 2014,

26(4): 638-643.

- [16] 葛小三, 付魁, 程钢, 马勇, 孙玉祥. 数据挖掘支持下的网络热点事件地理可视化研究[J]. 河南理工大学学报(自然科学版), 2016, 35(5): 655-659.
- [17] Broniatowski, D.A., Paul, M.J. and Dredze, M. (2013) National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic. *PLoS ONE*, **8**, e83672. <https://doi.org/10.1371/journal.pone.0083672>
- [18] Kim, J. and Hastak, M. (2018) Social Network Analysis: Characteristics of Online Social Networks after a Disaster. *International Journal of Information Management*, **38**, 86-96. <https://doi.org/10.1016/j.ijinfomgt.2017.08.003>
- [19] 梁芷铭, 周玫, 宁朝波. 基于情感本体的网络舆情观点挖掘模型构建——政务微博话语权研究系列之十一[J]. 情报杂志, 2014, 33(5): 143-147.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-1476, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: orf@hanspub.org