

Research on Early Warning of Public Opinion Based on BP Neural Network

Qiong Zhao^{1,2}, Yuting Pan²

¹School of Information Management and Statistics, Hubei University of Economics, Wuhan Hubei

²School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan Hubei

Email: 77199651@qq.com

Received: Mar. 11th, 2020; accepted: Mar. 24th, 2020; published: Mar. 31st, 2020

Abstract

In the context of social stability, this paper builds a network public opinion early warning index system that includes 4 primary indicators and 16 secondary indicators. The BP neural network is used to establish a network public opinion early warning model to give a level early warning of network public opinion events, and to carry out the model empirical analysis. The results show that public sentiment theme, public sentiment, public sentiment participation, and public opinion dissemination trends have a large impact on the level of early warning. Then, the accuracy rate of online public opinion warning using BP neural network is as high as 70%, but the average accuracy of the warning 62.5%; there is still room for improvement. In general, the network public opinion early warning index system is feasible, and the BP neural network early warning model is effective, which can provide a theoretical reference for the network public opinion early warning method.

Keywords

Internet Public Opinion, Early Warning Indicators, BP Neural Network

基于BP神经网络的网络舆情预警研究

赵琼^{1,2}, 潘玉婷²

¹湖北经济学院信息管理与统计学院, 湖北 武汉

²中南财经政法大学统计与数学学院, 湖北 武汉

Email: 77199651@qq.com

收稿日期: 2020年3月11日; 录用日期: 2020年3月24日; 发布日期: 2020年3月31日

摘要

本文首先构建了包含4个一级指标和16个二级指标的网络舆情预警指标体系, 然后利用BP神经网络建立

文章引用: 赵琼, 潘玉婷. 基于BP神经网络的网络舆情预警研究[J]. 统计学与应用, 2020, 9(2): 224-236.

DOI: 10.12677/sa.2020.92025

网络舆情预警模型对网络舆情事件进行等级预警, 并对该模型进行了实证分析。结果表明, 舆情主题、舆情热度、舆情参与度、舆情传播态势这几个指标对预警等级的影响较大, 然后利用BP神经网络进行网络舆情预警的正确率高达70%, 但是预警的平均精确率为62.5%, 还有提升空间。总体上来说, 网络舆情预警指标体系具有可行性, 且BP神经网络预警模型是有效的, 可以为网络舆情事件的预警方法提供理论参考。

关键词

网络舆情, 预警指标, BP神经网络

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

近年来中国互联网技术的迅速发展催生了大量的社交网络平台, 例如新浪微博、微信、抖音短视频等等。这些平台拥有大量的网民用户, 并且用户可以通过这些平台自由发布个人评论和态度, 这样使得网络舆情由此产生。由于网络舆情具有虚拟性、突发性以及破坏性等特点, 很容易出现一些网民将网络平台作为随意发泄情绪的空间, 发布一些有悖于事实的言论导致网络空间的混乱甚至影响社会的稳定发展。因此建立一套合理的自动化网络舆情预警体系有利于协助相关人员及时关注网络舆情的走向, 更加准确地对网络舆情转播中的突发状况进行预警, 对维护和促进社会和谐稳定产生积极影响。

国外对于舆情的研究起步较早, 美国等国家基本都是围绕选举发展起来的, 1936年 Archibald Crossley 对民众舆情进行研究来预测当年美国总统大选结果[1]。进入互联网时代国外学者加速开展利用现代先进技术进行网络舆情预警分析。专家可通过使用人工智能算法理论进行数据挖掘工作。Antioine Naud 等利用文本主题聚类算法对舆情事件的热点主题进行文本倾向性分析, 及时发现民众情感倾向性进行及时预警[2]。Geng 利用层次分析法的极限学习机(AHP-ELM)预警建模来处理食品安全中复杂的食品检验数据方面的问题[3]。大多数国外的研究成果对于我国网络舆情安全预警方面有较大的参考价值, 但是国外大部分解析网络舆情方法均不对外公开, 很难获取和借鉴。

我国对于网络舆情预警模型的研究开始较晚。但是随着时代发展, 使用先进技术手段建模的文献逐渐增加。刘毅利用基于三角模糊数的模糊德尔菲法和模糊层次分析法, 建立了针对某一具体热点话题或公共事件的网络舆情预警指标体系[4]。王铁套等也是利用传统的模糊综合评价法建立网络舆情预警模型[5]。孙玲芳将遗传算法和BP神经网络和遗传算法相结合, 构造了基于BP神经网络和遗传算法的网络舆情危机预警体系[6]。张艳岩建立了基于支持向量机的网络舆情危机预警模型[7]。

梳理国内外文献可以发现: 从研究内容看, 我国对于社会稳定预警研究起步较晚且研究内容较少, 并且依据的理论基础大多数都从国外理论展开, 没有从我国的基本国情出发, 导致研究成果缺乏可操作性。从研究方法上看, 多数文献在网络舆情预警指标选择上采取定性描述的方式, 缺乏定量指标的设定。而定量分析应当是网络舆情预警研判的基础。从研究实证和影响看, 多数研究成果在对网络舆情预警指标设定之后缺乏实证分析和检验, 呈现出重理论、轻实证现状。本文首先使用改进的复合权重的计算方法, 将专家对网络舆情不同时间点的预警得分与百度指数相结合, 计算出最终的预警综合得分。其次采用BP神经网络在建立的网络舆情预警指标体系基础上构建完整的网络舆情预警模型对具体网络舆情事件进行定量预警。

1.1. 专家打分法

在使用 BP 神经网络构建网络舆情预警模型时设置期望输出值时引入专家打分法作为预警等级进行评判重要依据, 使得 BP 神经网络可以有效训练和学习。同时便于模型最终预警的评价和分析。

1.2. 复合权重得分法

通过对主观得分与客观得分进行赋权计算不同时间点的综合得分。本文将主观得分与客观得分采用公式(1)对不同时间点综合得分 S 的计算[8]。其中, 主观得分为专家打分数据, 客观得分为百度指数评价数据。

$$S_j = \sqrt{(\varepsilon_j W_j)^2 + [(1 - \varepsilon_j) \theta_j]^2}, j = 1, 2, 3, \dots, 25 \quad (1)$$

其中 W_j 为主观得分分数, θ_j 为客观得分分数, ε_j 为权重偏好系数。

为确定 ε_j 值, 可以采用最小二乘法建立规划模型, 可得到:

$$\begin{cases} \min S_j = \sqrt{(\varepsilon_j W_j)^2 + [(1 - \varepsilon_j) \theta_j]^2} \\ \text{s.t. } 0 \leq \varepsilon_j \leq 1 \end{cases} \quad (2)$$

通过求解可得到动态的权重偏好系数 ε_j 为:

$$\varepsilon_j = \frac{\theta_j^2}{W_j^2 + \theta_j^2} \quad (3)$$

1.3. BP 神经网络简介

BP 神经网络全称是基于误差反向传播算法的人工神经网络, 是神经网络学习算法的一种。误差逆传播(Back Propagation, 简称 BP)算法是目前最成功的神经网络学习算法。它包括输入层、隐层以及输出层。不同的层级之间互联, 相同层级的神经元不相连, 图 1 为 BP 网络结构图:

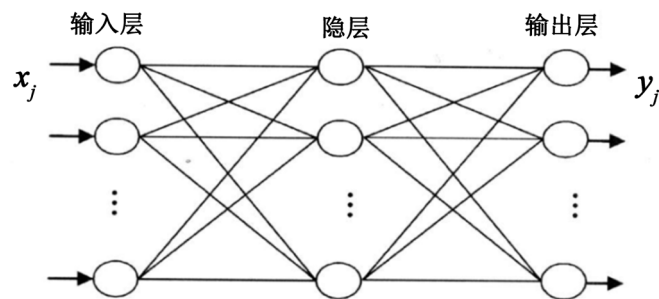


Figure 1. Structure diagram of BP neural network

图 1. BP 神经网络结构图

算法步骤主要分为以下几步:

1) 初始化网络连接权和阈值。输入样本 X 和期望输出 Y , 输入层第 I 个神经元与隐层第 H 个神经元之间的连接权为 v_{ih} , 隐层第 h 个神经元与输出层第 j 个神经元的连接权为 w_{hj} , 输出层第 j 个神经元的阈值为 θ_j , 隐层第 h 个神经元的阈值为 γ_h 。输入神经元个数即为设计的预警体系指标数为 d , 输出神经元个数为预警等级数量为 l , 隐层神经元个数为 q , 学习率为 η 。

2) 计算隐层输出, 其中 f 为隐层传递函数

$$m_h = f\left(\sum_{i=1}^d v_{ih}x_i + \gamma_h\right), h=1,2,\dots,q \quad (4)$$

3) 计算输出层输出。由隐含层输出, 计算预测输出层输出:

$$y_j = f\left(\sum_{h=1}^q w_{hj}H_h + \theta_j\right), j=1,2,\dots,l \quad (5)$$

其中 f 为输出层传递函数。

4) 进行误差计算并判断计算实际输出各节点的误差:

$$e_j = (Y_j - y_j) * y_j * (1 - y_j), j=1,2,\dots,l \quad (6)$$

计算隐层输出各节点的误差:

$$\delta_h = \left(\sum_j e_j * w_{hj}\right) * m_h * (1 - m_h), h=1,2,\dots,q \quad (7)$$

当误差小于预期设定的误差时返回步骤(1)进行下一个样本的学习, 若所有样本学习完成, 则训练结束。当误差未达到预定要求且学习次数未超限制时, 需要对权值和阈值进行更新。

5) 连接权值的更新。更新输出层的连接权值:

$$w_{hj}(N+1) = w_{hj} * N + \alpha * e_j * m_h, h=1,2,\dots,q \quad (8)$$

其中 N 表示迭代次数, α 表示修正系数, 且 $0 < \alpha < 1$ 。更新隐层的连接权值:

$$v_{ih}(N+1) = v_{ih} * N + \beta * \delta_h * x_i, i=1,2,\dots,d; h=1,2,\dots,q \quad (9)$$

其中 N 表示迭代次数, β 表示修正系数, 且 $0 < \beta < 1$ 。

6) 阈值的更新。更新输出层的阈值

$$\theta_j(N+1) = \theta_j * N + \alpha * e_j, j=1,2,\dots,l \quad (10)$$

更新隐层的阈值

$$\gamma_h(N+1) = \gamma_h * N + \beta * \delta_h, h=1,2,\dots,q \quad (11)$$

7) 将步骤(5)和(6)中修正的连接权值和阈值返回步骤(2)中进行继续训练学习, 直至训练误差达到预期要求或者循环次数足够大。

2. 数据处理与描述

2.1. 网络舆情预警指标的选取

网络舆情预警指标体系是不仅能够及时反映网络舆情现状和存在的问题, 并且可以及时对于突发事件作出及时预警。在结合相关网络舆情传播理论以及网络舆情的发展及变化规律的一些特点, 同时遵循科学性、可操作性、定性与定量相结合、灵敏性原则建立网络舆情预警指标体系。

在已有相关研究结果和网络舆情理论研究的指导下, 又通过向政府舆情管理部门和媒体单位的相关专家进行咨询, 经过多次整改遴选, 最终构建了 4 个一级指标和 16 个二级指标的网络舆情预警指标体系如表 1 所示。

Table 1. Network public opinion warning index system
表 1. 网络舆情预警指标体系

一级指标	二级指标
舆情热度 D_1	网络搜索量 E_1 媒体来源数量 E_2 媒体报道量 E_3
舆情参与度 D_2	原创微博发布量 E_4 微博转发量 E_5 微博评论数 E_6 微博点赞量 E_7 发布者影响度 E_8
舆情传播态势 D_3	事件搜索量变化率 E_9 原创微博发布量变化率 E_{10} 微博评论数变化率 E_{11} 微博转发量变化率 E_{12} 舆情内容直观度 E_{13}
舆情主题 D_4	舆情事件性质 E_{14} 舆情主题敏感度 E_{15} 舆情主题容忍度 E_{16}

2.2. 网络舆情预警指标含义

1) 网络搜索量是指在某个时间段之内网民使用百度搜索引擎搜索相应舆情事件的数量。该指标可以使用百度指数进行描述, 百度指数是将网民在百度的搜索量作为基础, 计算不同关键词在百度搜索中的频次进行加权计量得出。

2) 媒体来源数量是指在某个时间段内发布某舆情事件相关报道的网络新闻媒体的数量。通过综合统计, 国内使用频率较高影响力和公信力度较大的 10 个新闻媒体网站分别是: 人民网, 凤凰资讯, 央视网, 新浪新闻, 新华网, 中国网, 观察者网, 今日头条, 腾讯新闻, 搜狐新闻。在百度搜索引擎中对舆情事件进行搜索, 并将搜索出的报道标题结果进行分词处理, 与上述 10 个网站进行匹配, 计算出舆情事件出现在 10 个网站中的次数。

3) 媒体报道量是指某个时间段内某舆情话题相关的所有媒体报道的数量。

4) 原创微博发布量是指微博用户在某个时间段内发布某舆情事件的原创微博数量。通过新浪微博网页版搜索某舆情事件关键词, 选择对应时间段, 使用高级检索工具抓取原创微博数量。

5) 微博转发量是指在某个时间段内转发某舆情事件的微博条数。

6) 微博评论数是指在某个时间段内评论某舆情事件的微博条数。

7) 微博点赞量是指在某个时间段内某舆情事件的微博的点赞数量。

8) 发布者影响度是指不同舆情信息的传播者对于网络舆情事件发展的影响力。本文将此指标量化为发布者影响度 = $0.8 * \text{大V 博主数量} + 0.2 * \text{普通博主数量}$ 。

9) 事件搜索量变化率是指在一段时期内某个舆情事件搜索量的变化量。

10) 原创微博发布量变化率是指在一段时期内某个舆情事件原创微博的变化量。

11) 微博评论数变化率是指在一段时期内某个舆情事件微博评论的变化量。

12) 微博转发量变化率是指在一段时期内某个舆情事件微博转发量的变化情况。

13) 舆情内容直观度是指网民获取网络舆情事件进展的直观程度。据《2010 年中国互联网舆情分析报告》研究表明[9], 带有舆情事件相关图片、视频等较直观内容的信息传播速度更快。因此舆情内容直观程度可以通过发布的有关微博中带有图片和视频的数量占全部微博的比例来表示。

14) 事件性质是由不同网络舆情事件的类型决定的。不同的事件类型网民的参与程度也会有所不同。可以将事件分为自然灾害、事故灾难、公共卫生事件、社会安全事件、娱乐事件等。此定性指标可由专家打分的方式来量化不同事件性质的重要程度,得分在 0 和 1 之间,表示对应事件性质类型的重要程度,0 表示不重要,1 表示非常重要,0.5、0.7 分别表示重要、很重要。

15) 舆情主题敏感度是指网民对于某个舆情事件的关注和反映程度。0 表示不敏感,1 表示非常敏感,0.5、0.7 分别表示敏感、很敏感。

16) 舆情主题容忍度是指网民对于某一网络舆情事件的容忍程度的大小,其取值范围是[0,1]。

2.3. 网络舆情事件的概述

2019 年 3 月 30 日 18 时许,四川省凉山州木里县雅砻江镇立尔村发生森林大火由于火情等级较高,火灾波及范围和破坏力较大,在搜救和灭火过程中受自然地形和风力突变等原因发生山火爆燃事件,造成 27 名消防指战员和 4 名地方扑火人员壮烈牺牲。本文选取四川省凉山州木里县“3.30”森林火灾这一突发性自然灾害事件作为样本采集的舆情事件案例。事件具体进展如表 2 所示。

Table 2. Event progress

表 2. 事件进展情况

时间(2019 年)	事件进展
3 月 30 日	四川省凉山州木里县雅砻江镇立尔村发生严重森林火灾,火场地形复杂灭火难度较大
3 月 31 日	扑火人员转场时突发山火爆燃,30 名扑火人员失联
4 月 1 日	发布山火过火面积达 15 公顷左右,已确认 遇难人数达 30 人
4 月 2 日	勘查起火位置和原因,公布牺牲英雄人员名单并将遗体进行安置
4 月 4 日	西昌市人民政府设为哀悼日,当日新发现 1 名牺牲扑火人员,遇难人数升至 31 人
4 月 5 日	确定起火原因为雷击起火
4 月 7 日	凉山山火复燃,近 800 名消防人员参与扑救
4 月 8 日	继续采取地空配合扑救,直接扑打和处理余火

2.4. 数据来源与预处理

本文使用 python 软件对百度网页数据和微博平台数据进行爬取。爬取了 3 月 26 日至 4 月 21 日的数据进行实证分析。将 27 天分为 25 个时间点,前 15 个时间点为建模组用以模型的构建,其余 10 个时间点的样本数据划分为测试组。

采集的数据需要对原始数据进行数据预处理。依据本文建立的指标体系中的指标解释并采用 Excel 等计算软件对缺失值、极端值等进行处理得到的部分原始数据如表 3 所示。

Table 3. Part of the original data

表 3. 部分原始数据

时间 \ 指标	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
3.26	151	0	10	45	121	114	45	42
3.27	168	1	15	88	256	215	78	69
3.28	172	1	13	85	207	230	81	70
3.29	170	1	14	78	236	245	85	68

Continued

3.30	478	10	150	543	2564	859	895	88
3.31	461	8	265	1054	6675	1670	1689	136
4.1	1506	8	589	4593	10813	3896	4897	565
4.2	7324	7	765	3380	6790	1986	3698	372
4.3	4262	6	532	3547	6571	1769	2758	349
时间 \ 指标	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}
3.26	13.00	38.10	23.78	21.10	0.1	0	0.5	0.1
3.27	14.33	45.00	33.67	11.00	0.21	0	0.5	0.1
3.28	14.56	35.6	33.87	12.1	0.22	0	0.5	0.1
3.29	13.87	38.7	36.59	13.3	0.30	0.5	0.7	0.2
3.30	124.50	610.75	186.25	212.50	0.45	0.5	0.5	0.2
3.31	201.80	1,310.80	311.20	328.80	0.55	0.5	0.7	0.2
4.1	758.00	1,782.00	630.33	808.67	0.60	0.8	0.7	0.4
4.2	476.43	952.71	267.43	521.86	0.71	0.7	0.7	0.4
4.3	389.11	806.25	206.88	339.13	0.54	0.7	0.8	0.3

3. 实证分析

3.1. 样本预警等级的设定

关于网络舆情预警的级别暂时还没有统一的标准, 通过参考现阶段各类文献, 本文认为可以将预警等级进行进一步划分, 并使用不同颜色对严重程度进行表示。因此本文通过参考“社会治安预警等级标准”课题组关于社会治安预警等级标准的依据和预警等级的构造原则[10]最终将网络舆情预警等级分为特大警情(I级)、重大警情(II级)、较大警情(III级)、一般警情(IV级)、无警情(V级)共五个等级, 分别用红色、橙色、黄色、绿色和蓝色表示。不同预警等级的划分和得分区间如表4所示。

Table 4. Classification of early warning levels

表 4. 预警等级分类

预警等级	警情程度	综合得分区间	预警颜色
I	特大警情	(80,100]	红色
II	重大警情	(60,80]	橙色
III	较大警情	(40,60]	黄色
IV	一般警情	(20,40]	绿色
V	无警情	(0,20]	蓝色

使用 BP 神经网络建立网络舆情预警模型训练时需要输入舆情事件不同时间点的预警等级期望输出, 因此必须要提前对网络舆情事件不同事件点的预警等级进行科学研判。本文采用主观的专家打分法与客观百度指数评价数据相结合的方法对网络舆情事件不同时间点的预警等级进行划分。得到预警等级如表5所示。

Table 5. Comprehensive scores and warning levels of network public opinion events
表 5. 网络舆情事件的综合得分和预警等级

时间点	综合评分	预警等级	时间点	综合评分	预警等级
1	13.5000	V	14	41.1057	III
2	16.5000	V	15	49.4972	III
3	14.4999	V	16	43.2367	III
4	20.4995	IV	17	30.9157	IV
5	45.4453	III	18	31.9816	IV
6	61.4919	II	19	28.9876	IV
7	54.4592	III	20	33.9883	IV
8	81.1289	I	21	32.9953	IV
9	66.0464	II	22	50.4996	III
10	65.3972	II	23	28.4984	IV
11	61.3910	II	24	19.4492	V
12	51.5618	III	25	17.4756	V
13	48.4673	III			

将通过计算得到的综合得分和预警等级作为建立 BP 神经网络网络舆情预警模型中的期望预警等级。同时将最终计算出的综合得分作为模型评判的评价标准。

3.2. 模型构建

3.2.1. BP 神经网络模型结构的确定

1) 输入节点个数设定。本章建立的网络舆情预警指标体系中共设立 16 个预警指标, 故输入节点设置为 16。

2) 设置隐层节点。由于神经网络具有大量的并行分布结构和非线性动态特性[11]。隐层节点的设置没有明确的公式, 通过参考资料[12], 采用公式(12)来确定隐层节点个数:

$$N = \sqrt{m+n+a} \quad (12)$$

其中 m 为输入层节点个数, n 为输出层节点个数, a 为 1~10 之间的任意常数。当隐含节点数为 8 时训练精度较好, 因此最终确定隐含层的节点个数为 8。

3) 输出节点个数设定。预警等级分为五个等级, 因此设定输出节点个数为 5。

BP 神经网络预警模型是典型的“输入层 - 隐层 - 输出层”模型。其中输入层输入的样本是指标的样本实际值, 输出的是期望得到的预警结果, 即五种预警程度, 中间神经网络处理部分就相当于一个“黑匣子”。则输入节点数为 16, 输出节点数为 5。模型示意图如图 2 所示。



Figure 2. Schematic diagram of neural network model
图 2. 神经网络模型示意图

3.2.2. 面向 MATLAB 的 BP 神经网络模型的建立

1) 设置训练参数。采用的具体训练参数如表 6 所示。

Table 6. Training parameters of neural network
表 6. 神经网络的训练参数

训练参数	参数值
输入层 - 隐层传递函数	tansig
隐层 - 输出层传递函数	purelin
训练函数	traingdx

2) 数据归一化。将指标值进行数据预处理之后, 为使得各个指标具有相应的可比性需对其进行归一化处理。正向指标的归一化和负向指标的归一化分别如公式(13)和公式(14)所示。

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

$$x = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \quad (14)$$

对于本文设计的网络舆情预警指标体系的不同指标的性质进行分析, 不同指标的性质如表 7 所示。

Table 7. Index properties of early warning system

表 7. 预警体系指标性质

指标变量	指标名称	指标性质
C_1	网络搜索量	定量, 负向指标
C_2	媒体来源数量	定量, 负向指标
C_3	媒体报道量	定量, 负向指标
C_4	原创微博发布量	定量, 负向指标
C_5	微博转发量	定量, 负向指标
C_6	微博评论数	定量, 负向指标
C_7	微博点赞量	定量, 负向指标
C_8	发布者影响度	定量, 负向指标
C_9	事件搜索量变化率	定量, 负向指标
C_{10}	原创微博发布量变化率	定量, 负向指标
C_{11}	微博评论数变化率	定量, 负向指标
C_{12}	微博转发量变化率	定量, 负向指标
C_{13}	舆情内容直观度	定量, 负向指标
C_{14}	舆情事件性质	定性, 负向指标
C_{15}	舆情主题敏感度	定性, 负向指标
C_{16}	舆情主题容忍度	定性, 负向指标

将预处理之后的数据依据表 7 中指标的性质, 对原始数据进行归一化处理。

3) 数据输入。将训练样本以矩阵的形式输入到 MATLAB 中, 将训练模型中样本对应设定的预警状态作为期望输出变量。警情由低到高输出状态分别为 00001、00010、00100、01000、10000。则训练样本的期望输出如表 8 所示。

Table 8. Expected output of training samples
表 8. 训练样本期望输出

时间点	预警等级	期望输出
1	V	00001
2	V	00001
3	V	00001
4	IV	00010
5	III	00100
6	II	01000
7	III	00100
8	I	10000
9	II	01000
10	II	01000
11	II	01000
12	III	00100
13	III	00100
14	III	00100
15	III	00100

4) BP 神经网络的生成和训练。利用 MATLAB R2016a 中 BP 神经网络模拟工具包可以直接在对应操作框中输入训练样本和期望输出变量。

通过 32270 次的训练, 误差率为 $1.00e-05$, 达到目标范围内, 运行结果如表 9 所示。训练误差曲线变化图如图 3。

Table 9. Expected output of training samples
表 9. BP 神经网络运行结果

Progress	Epoch	Performance	Gradient	Validation Checks
数值	32270	$1.00e-05$	$7.69e-05$	0

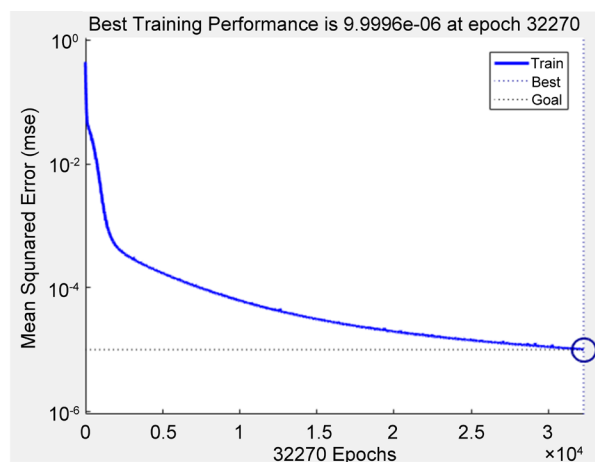


Figure 3. Error curve
图 3. 误差曲线图

3.3. 实证结果分析

3.3.1. 预警正确率

通过输入归一化后的训练样本, 通过模型对不同样本预警等级的划分判定 BP 神经网络预警模型的预警正确率。将得到的输出结果Y按照公式(15) [13]:

$$y_i = \begin{cases} 0 & Y_i \leq 0.1 \\ 0.5 & 0.1 < Y_i < 0.9 \\ 1 & Y_i \geq 0.9 \end{cases} \quad (i = 1, 2, 3, 4, 5) \quad (15)$$

实际输出结果与转换后的实际输出结果如表 10 所示:

Table 10. Summary of test sample output results

表 10. 检验样本输出结果汇总

时间点	BP 实际输出(Y)	期望输出	实际输出预警等级	结果
16	01000 (0.0226 0.9875 0.0000 0.0001 0.0031)	00100	II	不一致
17	00010 (0.0021 0.0001 0.0045 0.9768 0.0003)	00010	IV	一致
18	01000 (0.0087 0.9945 0.0034 0.0021 0.0346)	00010	II	不一致
19	00010 (0.0021 0.0003 0.0056 0.9563 0.0000)	00010	IV	一致
20	00010 (0.0014 0.0034 0.0001 0.9235 0.0021)	00010	IV	一致
21	00010 (0.0012 0.0026 0.9375 0.0019 0.0003)	00010	IV	一致
22	00100 (0.0012 0.0037 0.9235 0.0002 0.0018)	00100	III	一致
23	00100 (0.0031 0.0237 0.9153 0.0002 0.0001)	00010	III	不一致
24	00001 (0.0031 0.0003 0.0041 0.0005 0.9564)	00001	V	一致
25	00001 (0.0013 0.0021 0.0001 0.0004 0.9102)	00001	V	一致

通过结果汇总可以看出, 10 个测试样本中只有三个样本与期望输出存在差异, 其余样本均输出正确预警结果, 可看出使用 BP 神经网络建立网络舆情预警模型对网络舆情事件的预警等级的正确率达到 70%, 因此使用 BP 神经网络对网络舆情进行预警的正确率较高。

3.3.2. 预警精确率

使用 BP 神经网络模型对网络舆情事件预警的精确率统计如表 11 所示。

Table 11. BP neural network early warning accuracy rate statistics

表 11. BP 神经网络预警精确率统计

实际预警等级 \ 期望预警等级	期望预警 II 级	期望预警 III 级	期望预警 IV 级	期望预警 V 级	预警精确率(%)
实际预警 II 级	0	1	1	0	0
实际预警 III 级	0	1	1	0	50
实际预警 IV 级	0	0	4	0	100
实际预警 V 级	0	0	0	2	100

通过表 11 不同预警等级的预警精确率求算数平均数可以计算出使用 BP 神经网络的平均精确率 P_1 :

$$P_1 = \left(0 + \frac{1}{2} + 1 + 1 \right) \times \frac{1}{4} = 62.5\%$$

因此使用 BP 神经网络进行舆情预警的平均精确率达到 62.5%, 虽然超过 50%, 但是预警精确率还有待提升。

4. 结论及建议

本文通过建立网络舆情预警体系并使用 BP 神经网络建立网络舆情预警模型对 4.30 四川省凉山州森林山火事件进行实证分析可以得到以下结论。第一, 网络舆情预警指标体系一级指标包含舆情主题、舆情热度、舆情参与度和舆情传播态势, 说明这几个方面对于网络舆情的警情程度的判定影响较大。第二, 利用 MATLAB 构建 BP 神经网络舆情预警模型, 通过四川省凉山州山火实例数据验证了使用所建立的模型进行预警的正确率较高, 但是在精确度方面有待提升。第三, 通过实例不仅验证了利用 BP 神经网络建立网络舆情预警模型的有效性, 同时也证明了网络舆情预警指标体系的可行性, 可以利用指标体系中包含的指标因素对网络舆情事件开展及时治理工作。

基于以上结论, 本文就提高我国网络舆情预警的能力提出以下建议。第一, 发挥微博等媒体平台自律作用。为确保网络环境的安全秩序, 不同的网络媒体平台应当自发设立网络信息管理小组, 同制定相应的管理条例对发布信息的媒体机构和广大网民提出约束条件, 在保证网络空间言论自由的情况下保证网络空间的纪律性。微博平台通过从微博用户中招募“法官”来自发处理微博中的不实信息和不良评论, 采取封锁账号、禁言等方式进行惩罚。使得网络监管在政府刚性管理之外有了一个缓冲的平台, 推动了微博网友的自律, 具有较大的参考意义。第二, 使用人工神经网络进行网络舆情预警, 提升技术治理水平。通过先进的技术手段可以建立网络舆情预警模型, 对可能会影响社会稳定的舆情事件进行预判。本文通过实证分析证明了神经网络对网络舆情预警等级预测的准确性和实用性。因此在对网络舆情事件不同时间段的预警等级进行判断时可以采用 BP 神经网络实现预判。在大数据时代背景下, 通过大数据平台技术实现对网络舆情事件数据的自动收集、整理, 同时采用对应技术手段对数据进行分析 and 情感研究, 可以更加快速准确获取舆情信息, 弥补传统数据收集和处理办法的不足, 从而提高舆情处理效率。

致 谢

感谢匿名审稿人提出的宝贵修改意见。

基金项目

感谢国家社科基金青年项目《社交网络对社会稳定性影响的统计研究》(15CTJ006)对本研究的资助。

参考文献

- [1] Crossley, A. (1937) Straw Polls in 1936. *Public Opinion Quarterly*, **1**, 24-35. <https://doi.org/10.1086/265035>
- [2] Naud, A. and Usui, S. (2007) Exploration of a Text Collection and Identification of Topics by Clustering. *Intelligent Data Engineering and Automated Learning: 8th International Conference*, Birmingham, 16-19 December 2007, 115-124. https://doi.org/10.1007/978-3-540-77226-2_13
- [3] Geng, Z.Q., Zhao, S.S., Tao, G.C. and Han, Y.M. (2017) Early Warning Modeling and Analysis Based on Analytic Hierarchy Process Integrated Extreme Learning Machine (AHP-ELM): Application to Food Safety. *Food Control*, **78**, 33-42. <https://doi.org/10.1016/j.foodcont.2017.02.045>
- [4] 刘毅. 内容分析法在网络舆情信息分析中的应用[J]. 天津大学学报(社会科学版), 2006(4): 307-310.
- [5] 王铁套, 王国营, 陈越. 基于模糊综合评价法的网络舆情预警模型[J]. 情报杂志, 2012, 31(6): 47-51+58.
- [6] 孙玲芳, 周加波, 林伟健, 候志鲁, 许锋. 基于 BP 神经网络和遗传算法的网络舆情危机预警研究[J]. 情报杂志, 2014, 33(11): 18-24.
- [7] 张艳岩. 基于支持向量机的网络舆情危机预警研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2013.
- [8] 熊云峰, 陈章兰, 袁红莉. 基于复合权重的船舶性能灰色综合评价法[J]. 舰船科学技术, 2012, 34(8): 119-122.

- [9] 祝华新, 单学刚, 胡江春. 2010 年中国互联网舆情分析报告[R]. 人民网, 2010.
- [10] “社会治安动态预警研究”及“社会治安预警等级标准”课题组. 制定社会治安预警等级标准的初步构想[J]. 中国人民公安大学学报(社会科学版), 2010, 26(1): 8-12.
- [11] Jarvenpaa, S.L., *et al.* (2000) Consumer Trust in an Internet Store. *Information Technology Management*, **1**, 45-71.
<https://doi.org/10.1023/A:1019104520776>
- [12] 朱晨飞, 黄淑华, 何杭松, 康友春. 基于 BP_Adaboost 算法的网络舆情危机预警[J]. 中国公共安全(学术版), 2017(4): 95-101.
- [13] 薛圈圈. 基于 BP 神经网络的网络舆情危机预警研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2010.