·理论探索·

基于 BP 神经网络的突发传染病舆情热度趋势预测模型研究

曾子明[1,2] 黄城莺[1,2]
（1. 武汉大学信息资源研究中心，湖北 武汉 430072；
2. 武汉大学图书情报实验教学中心，湖北 武汉 430072）

〔摘 要〕[目的/意义] 研究突发传染病舆情热度的发展趋势，能够为制定舆情引导策略提供参考，具有重要的理论意义。[方法/过程] 本文首先构建微博舆情热度评价指标体系，基于信息熵确定各个指标的权重，然后对求得的舆情热度趋势值进行分类，在此基础上，建立基于 BP 神经网络的突发传染病舆情热度趋势预测模型。以新浪微博为例，选取“MERS 病毒卫生突发事件”的舆情热度数据进行实例分析，预测该突发传染病事件的发展趋势，从而验证模型的可行性。[结果/结论] 实验结果表明，该模型能有效预测突发传染病舆情热度趋势，进而为舆情管控提供决策支持。

〔关键词〕BP 神经网络；舆情热度；突发传染病；微博；预测模型
DOI: 10.3969/j.issn.1008-0821.2018.05.006
〔中图分类号〕G206.2 〔文献标识码〕A 〔文章编号〕1008-0821（2018）05-0037-08

Research on Public Opinion Heat Trend Prediction Model of Emergent Infectious Diseases Based on BP Neural Network

Zeng Ziming[1,2] Huang Chengying[1,2]
（1. Center of Information Resources Research，Wuhan University，Wuhan 430072，China；
2. Library and Information Experimental Teaching Center，Wuhan University，Wuhan 430072，China）

〔**Abstract**〕[Purpose/Significance] It is of great theoretical significance to study the development trend of public opinion in emergent infectious diseases，which can provide reference for making public opinion guidance strategy. [Method/Process] The paper first constructed the index system of microblog public opinion and evaluated the weight of each index based on the information entropy，and then classified the obtained public opinion heat trend，on the base of which，it established the public opinion heat trend prediction model of emergent infectious diseases bases on BP neural network. Taking Sina microblog as an example，it analyzed the public opinion heat data of “MERS virus” to predict the development trend of the emergent infectious disease event，and verified the feasibility of the model. [Result/Conclusion] The experimental results showed that the model could effectively predict the trend of public opinion in emergent infectious diseases，and then provide decision support for public opinion control.

〔**Key words**〕BP neural network；heat of public opinion；emergent infectious diseases；microblog；prediction model

据中国互联网络信息中心（CNNIC）发布的《第 40 次中国互联网络发展状况统计报告》显示，截至 2017 年 6 月，我国互联网普及率较 2016 年底提升了 1.1 个百分点，达到 54.3%，超过全球平均水平 4.6 个百分点[1]。在互联

收稿日期：2017-12-23
基金项目：教育部人文社会科学重点研究基地重大项目“大数据资源的智能化管理与跨部门交互研究——面向公共安全领域”（项目编号：16JJD870003）。
作者简介：曾子明（1977-），男，教授，博士生导师，研究方向：大数据资源的数据挖掘与智能化管理研究。黄城莺（1993-），女，硕士研究生，研究方向：大数据资源的数据挖掘与智能化管理研究。

网迅猛发展的环境下，国际范围内频繁发生的突发传染病引起人们高度关注，如“埃博拉病毒”、“塞卡病毒”、“SARS 病毒”、“MERS 病毒”等。微博作为传播媒介的代表，具有互动性高、传播速度快等特点，突发传染病事件借助新浪微博等社交平台不断发酵，往往迅速演化为网络舆情。以 2014 年西非爆发的埃博拉病毒为例，在 2014 年 2 月 1 日至 10 月 31 日期间，新浪微博平台上共产生 23 万多条包含“埃博拉”关键词的微博[2]。在描述突发传染病舆情的诸多要素中，舆情热度体现了人们对于舆情事件的关注程度，日益受到政府以及学术界的广泛关注。

突发传染病舆情同其他类型突发事件舆情相较而言，具有爆发性、演变不确定性、负面倾向性等特征。由于涉及公众的健康和生命安全，社会公众高度关注致病原因、每日新增病例数、死亡率、治愈情况等与之相关的信息。相关消息一旦发出就会掀起网络舆情的浪潮，引起整体舆论环境的波动，其舆情发展的管控已经成为应急管理的一个重要组成部分。基于此，本文结合徐旖旎[3]对媒体奇观网络舆情热度趋势分析以及赵磊、王松等[4]对舆情热度趋势仿真模型的研究思路，将突发传染病舆情热度趋势预测问题转化为模式分类问题，并尝试引入 BP 神经网络对舆情热度趋势做预测。首先，文章构建面向微博的舆情热度评价指标体系，基于信息熵确定各个指标的权重，再利用加权求和的方法得到热度值，然后求出舆情热度趋势值并进行分类，接着引入 BP 神经网络理论，从新浪微博收集“MERS 病毒卫生突发事件”相关数据，对突发传染病的舆情热度趋势进行预测，探讨该方法的可行性和有效性。

1 研究现状

1.1 网络舆情热度研究

网络舆情热度研究是一门涉及情报学、统计学、传播学等多学科交叉融合的研究领域。当前，国内外学者针对网络舆情热度的研究分为定性研究、定量研究以及定性定量相结合方法。其中，定性研究包含网络舆情热度发展演变规律、特征、热度评价指标体系建立等，定量研究包含最优化模型、系统动力学模型、马尔可夫链模型等。Lean Yu 等[5]提出了以网络公民、意见领袖、政府以及大众媒体四大主体为代表的网络舆情传播模型，并通过 4 起典型的危险化学品泄漏事件进行案例研究，验证模型的有效性；张行钦等[6]使用百度指数，研究了“乙肝疫苗”事件网络舆情热度演变规律；Jeffrey R Lax 等[7]根据突发事件网络舆情生成过程、热度涨落的影响因素，提出了较为成熟的舆情热度指标。曹学艳等[8]引入突发事件应对等级，构建了网络舆情热度评价指标体系；王慧军等[9]通过最小化舆情热度的负面作用与监控成本之和，研究了政府对舆情热度的最优监控问题；袁国平等[10]借助系统动力学的流图模型，通过 Vensim PIE 软件进行模拟仿真，从事件公共度、事件敏感度、网民质疑度、政府公信力 4 个方面分析对网络舆情热度的影响；屈启兴等[11]给出了基于微博的企业网络舆情热度的计算公式，在此基础上提出基于马尔可夫链的舆情热度趋势分析模型；王新猛[12]构建了针对政府负面网络舆情热度趋势的马尔可夫链预测模型。Xue Gang Chen 等[13]运用粗糙集理论降低网络舆情指标体系的属性，并通过层次分析法确定指标权重，从定量和定性的角度出发，提出一种网络舆情趋势预测与评价的新方法。

1.2 我国突发传染病舆情研究

我国对于突发传染病舆情的研究始于 2003 年的 SARS 事件。目前，相关研究主要包括舆情传播规律、舆情监测预警和舆情引导治理 3 个方面。安璐等[2]以埃博拉（Ebola）有关的微博为调查对象，利用 LDA 模型和 SOM 方法比较分析了 Twitter 和 Weibo 平台上相关微博的热点主题类别，揭示其演化模式和时序发展趋势的异同点；靳松等[14]以 H7N9 禽流感事件为研究对象，通过采集到的数据生成 H7N9 信息传播网络拓扑结构图，并基于其邻接矩阵，系统分析传播网络的要素和内部簇结构特性；安璐等[15]以“塞卡病毒”的微博数据为研究样本，利用潜在狄利克雷模型识别微博内容的主题特征，同时结合用户特征和发布时间特征，构建决策树模型，对突发传染病微博影响力进行预测；杜洪涛等[16]以新浪微博社区中 MERS 疫情数据为样本，研究如何改善突发性传染病舆情中的公共管理沟通问题；翁士洪等[17]以 H7N9 事件为例，探讨微博谣言的产生机制，并结合现有文献资料和现实状况，提出针对性的治理对策。

综上所述，虽然国内外已经有诸多学者对网络舆情热度展开研究，涵盖了从演变规律、评价指标到预测模型的多个方面，但是突发传染病舆情的相关研究仍然处于发展阶段，还需要深入研究该领域及相关技术方法。BP 神经网络由于具有强大的自学习、自适应的能力，擅长于模式识别、分类、数据拟合等问题的解决，被广泛应用于应急需求预测、微博转发量预测、冬小麦耗水预测等方面。因此，本文将基于 BP 神经网络，对突发传染病的舆情热度趋势进行预测研究。

2 微博舆情热度评价指标体系构建

2.1 舆情热度评价指标

建立一个科学合理的评价指标体系是衡量微博舆情热度的基础，并非指标越多越好，关键在于能否定量化反映微博舆情热度的实质。本文借鉴文献[11，12]构建的网络舆情指标，从原创微博发布量（A）、转发量（B）、评论量（C）、点赞量（D）等 4 个指标来描述微博舆情热度。这些数据以天为单位进行统计，其与时间的对应关系见表 1。

表 1 突发传染病微博舆情数据统计表

时间（T）	原创微博 A_1	原创微博 A_2	…	原创微博 A_n
1	$b_{1,1}$ $c_{1,1}$ $d_{1,1}$	$b_{1,2}$ $c_{1,2}$ $d_{1,2}$	…	$b_{1,n}$ $c_{1,n}$ $d_{1,n}$
2	$b_{2,1}$ $c_{2,1}$ $d_{2,1}$	$b_{2,2}$ $c_{2,2}$ $d_{2,2}$	…	$b_{2,n}$ $c_{2,n}$ $d_{2,n}$
…	…	…	…	…
m	$b_{m,1}$ $c_{m,1}$ $d_{m,1}$	$b_{m,2}$ $c_{m,2}$ $d_{m,2}$		$b_{m,n}$ $c_{m,n}$ $d_{m,n}$

第 i 天的原创微博发布量由公式（1）表示：

$$A_i = n \tag{1}$$

第 i 天的转发量由公式（2）表示：

$$B_i = \sum_{j=1}^{n} b_{i,j} \tag{2}$$

第 i 天的评论量由公式（3）表示：

$$C_i = \sum_{j=1}^{n} c_{i,j} \tag{3}$$

第 i 天的点赞量由公式（4）表示：

$$D_i = \sum_{j=1}^{n} d_{i,j} \tag{4}$$

由以上公式可以推出第 i 天微博舆情热度 $H(i)$ 的表达式为：

$$H(i) = W_1 \cdot A_i + W_2 \cdot B_i + W_3 \cdot C_i \tag{5}$$

其中，W_k（$k=1,2,3,4$）分别为原创微博发布量、转发量、评论量、点赞量的权重，为了更加便于计算，A_i、B_i、C_i、D_i 需根据公式（8）进行归一化处理。

第 i 天微博舆情热度趋势值的表达式为：

$$\overline{H(i)} = H(i+1) - H(i) \tag{6}$$

2.2 计算舆情热度评价指标权重

本文利用信息熵[18]确定各项指标的权重，计算步骤如下：

1）设有 m 个评价对象（天数），n 个评价指标，构造的原始指标矩阵为 $X=(x_{i,j})_{m\times n}$，由公式（7）表示：

$$X=\begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix} \tag{7}$$

2）一般而言，不同评价指标的类型、量纲等往往存在差异，为了消除这些差异带来的影响，将其转化为无量纲、方向一致的标准指标值，本文采用极值法[19]对评价指标进行无量纲化处理：

效益型指标：$$y_{i,j}=\frac{x_{i,j}-\min(x_{i,j})}{\max(x_{i,j})-\min(x_{i,j})} \tag{8}$$

成本型指标：$$y_{i,j}=\frac{\max(x_{i,j})-x_{i,j}}{\max(x_{i,j})-\min(x_{i,j})} \tag{9}$$

其中，$\max(x_{i,j})$、$\min(x_{i,j})$ 分别为指标评价值的最大值和最小值。

3）计算第 j 项指标下第 i 个评价对象指标值的比重，由公式（10）表示：

$$p_{i,j} = \frac{y_{i,j}}{\sum_{i=1}^{m} y_{i,j}} \tag{10}$$

4）计算第 j 项指标的熵值为：

$$e_j = -k\sum_{i=1}^{m} p_{i,j}\ln p_{i,j},\ k = \frac{1}{\ln m}\ (k>0, 0\leqslant e_j \leqslant 1) \tag{11}$$

5）进一步对 $1-e_j$ 归一化，得到第 j 项指标的熵权值为：

$$w_j = \frac{1-e_j}{\sum_{j=1}^{n}(1-e_j)} = \frac{1-e_j}{n-\sum_{j=1}^{n} e_j}\ (0 \leqslant w_j \leqslant 1 \text{ 且 } \sum_{j=1}^{n} w_j = 1) \tag{12}$$

3 BP 神经网络

BP 神经网络（Back Propagation Neural Network）基于梯度下降策略，通过反向传播来不断调整网络连接的权值和阈值，直到输出值与真实值的误差减少到可以接受的范围或预先设定的学习次数为止。BP 神经网络由输入层、隐含层、输出层组成，本文选取单隐层的三层 BP 神经网络来实现突发传染病舆情热度趋势的预测。其中输入向量为原创微博发布量、转发量、评论量、点赞量 4 个元素，所以输入层的节点数为 4。输出向量为微博舆情热度趋势值，本文将预测问题转化为模式分类问题，将微博舆情热度趋势值分为 6 类：$C1$ = 急速上升 = $\left[\frac{\overline{H(i)}\max}{2}, \overline{H(i)}\max\right]$，$C2$ = 明显上升 = $\left[\frac{\overline{H(i)}\max}{4}, \frac{\overline{H(i)}\max}{2}\right]$，$C3$ = 缓慢上升 = $\left[0, \frac{\overline{H(i)}\max}{4}\right]$，$C4$ = 缓慢下降 = $\left[\frac{\overline{H(i)}\min}{4}, 0\right]$，$C5$ = 明显下降 = $\left[\frac{\overline{H(i)}\min}{2}, \frac{\overline{H(i)}\min}{4}\right]$，$C6$ = 急速下降 = $\left[\overline{H(i)}\min, \frac{\overline{H(i)}\min}{2}\right]$，其中 $\overline{H(i)}\max$、$\overline{H(i)}\min$ 分别为微博舆情热度趋势值的最大值和最小值。在此基础上，分别用二进制 001、010、011、100、101、110 表示微博舆情热度趋势值的类别，所以输出层的节点数为 3，输出状态为：001、010、011、100、101、110，分别对应 6 种类别。对于隐含层节点数而言，若节点过多，则会致使网络复杂化甚至出现过度拟合的情况，若节点过少，则会致使结果不收敛，目前并没有一个理想的解析式可以用来确定合理的隐含层节点数，本文采用经验公式（13）得到隐含层节点数的估计值：

$$N = \sqrt{m+n} + \alpha \tag{13}$$

其中，m 为输入层节点数，n 为输出层节点数，α 为[1,10]之间的常数。BP 神经网络结构见图 1。

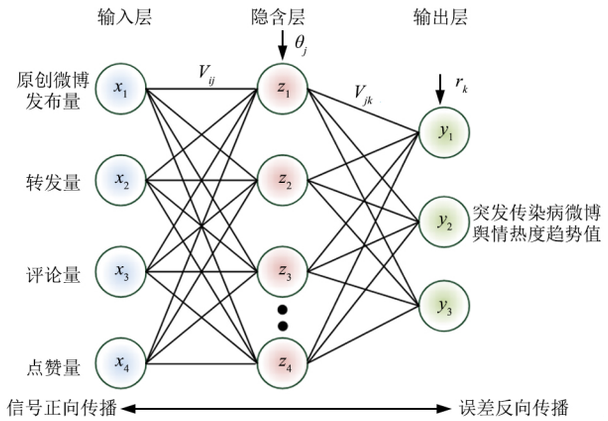

图 1 BP 神经网络结构图

BP 神经网络标准学习步骤[20]如下:

输入样本为 x_i，对应的隐含层的输出和输出层的输出分别为 z_j 和 y_k；V_{ij}是输入层和隐含层间的权重，V_{jk}是隐含层和输出层间的权重；θ_j、γ_k 分别为隐含层和输出层阈值。

a. 将连接权值 V_{ij}、V_{jk}以及阈值 θ_j、γ_k 随机初始化为[-1,+1]之间的值。

b. 根据公式（14）计算隐含层输出 z_j。

$$z_j = f\left(\sum_{i=1}^{m} V_{ij}x_i + \theta_j\right) \quad j = 1,2,\cdots,N \tag{14}$$

其中，N 为隐含层节点数；f 为隐含层激励函数，本文所选函数为 $f(x)=\dfrac{e^x-e^{-x}}{e^x+e^{-x}}$。

c. 根据公式（15）计算输出层输出 y_k。

$$y_k = f\left(\sum_{j=1}^{N} V_{jk}z_j + \gamma_k\right) \quad k = 1,2,3 \tag{15}$$

其中，f 为输出层激励函数，本文所选函数为 $f(x)=\dfrac{e^x-e^{-x}}{e^x+e^{-x}}$。

d. 根据公式（16）计算误差:

$$e_k = t_k - y_k \tag{16}$$

其中，t 表示期望输出，y 表示实际输出。

e. 根据公式（17）、（18）更新网络连接权值 V_{ij}、V_{jk}。

$$V_{ij} = V_{ij} + \eta z_j(1-z_j)\,x_i\sum_{k=1}^{3} V_{jk}e_k \tag{17}$$

$$V_{jk} = V_{jk} + \eta z_j e_k \tag{18}$$

其中，η 为学习率。

f. 根据公式（19）、（20）更新阈值 θ_j，γ_k。

$$\theta_j = \theta_j + \eta z_j(1-z_j)\sum_{k=1}^{3} V_{jk}e_k \tag{19}$$

$$\gamma_k = \gamma_k + \eta e_k \tag{20}$$

g. 网络进行学习训练，使得实际输出尽可能地接近期望输出，直至达到最大训练次数或满足误差精度要求。

4 实验及结果分析

本研究通过 Excel 2007 软件完成描述性统计以及图形绘制，利用 MATLAB R_2016a 神经网络工具箱构建突发传染病舆情热度趋势预测模型。研究分为突发传染病舆情时间跨度选择、舆情热度数据收集、舆情热度数据预处理及清洗、舆情热度数据归一化处理等阶段，具体的流程见图 2。

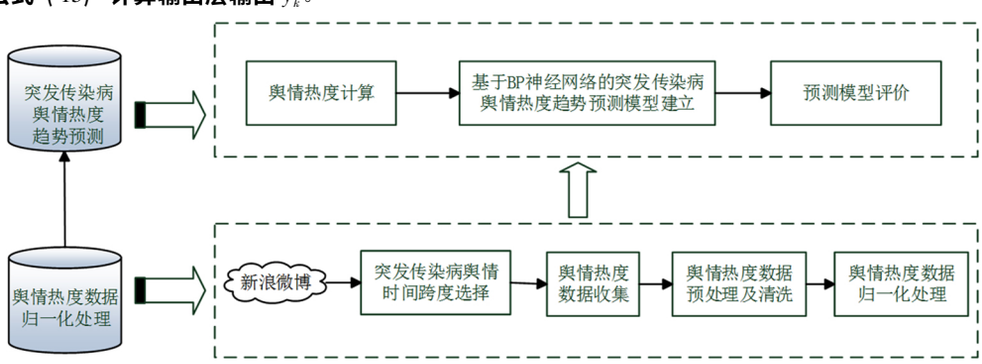

图 2 突发传染病舆情热度趋势预测流程图

4.1 突发传染病舆情时间跨度的选择

本文选取 2015 年上半年人民网舆情监测室广受关注的“MERS（中东呼吸综合征）病毒卫生突发事件”作为研究对象。根据人民网、中国新闻网关于中东呼吸综合征的新闻报道得出该突发传染病事件的进展见表 2。

结合百度指数搜索指数的网民关注度时间变化趋势，截止到 2015 年 7 月 31 日，该事件基本平息，网民关注度降到与突发传染病爆发前持平的状态。因此，选取 2015 年 5 月 28 日至 2015 年 7 月 31 日为研究时间段。

表 2 MERS（中东呼吸综合征）进展表

时间	MERS 事件进展报道
2015. 05. 28	广东出现首例 MERS 疑似病例，韩国发热男子入境已被隔离;
2015. 05. 31	广东官方：接诊首例 MERS 患者救护车司机被感染是谣言;
2015. 06. 11	1 名中国人在韩国感染 MERS（首位）;
2015. 06. 13	韩国 MERS 疫情出现首例第三代人传人感染者;
2015. 06. 26	中国首例 MERS 患者痊愈回国;

表 2（续）

时间	MERS 事件进展报道
2015. 07. 03	韩国 MERS 存活的第一代患者全部治愈，确诊病例增至 184 例；
2015. 08. 04	韩国 MERS 疫情致 36 人死亡后，韩国总统宣布撤换卫生部长；
2015. 10. 01	韩国最后 1 名 MERS 确诊患者判定为完全治愈；
2015. 11. 25	韩国最后 1 名 MERS 确诊患者死亡（10 月 11 日出现发热症状）；
2015. 12. 24	韩国宣布 MERS 疫情结束，此次疫情共 186 人感染，38 人死亡；

4. 2 舆情热度数据收集

本文利用 Gooseeker 爬取新浪微博上包含“MERS”词条的所有原创微博条目，具体字段包括：发布时间、博主名称、原创微博内容、微博网页、转发量、评论量、点赞量，共收集数据样本56 043条。

4. 3 舆情热度数据预处理及清洗

对收集到的原创微博进行逐条筛选，剔除 232 条广告、重复的1 255条记录、以及其他与 MERS 无关的 401 条微博，得到有效数据54 155条，累计转发量1 316 856次，累计评论量572 759次，累计点赞量991 963次。然后，以天为单位，根据公式（1）~（4）整理汇总每日微博舆情热度数据，得到 65 条结果见 3。

表 3 每日微博舆情热度数据

序号	时间	原创微博发布量	转发量	评论量	点赞量
01	2015. 05. 28	465	28 131	10 958	7 284
02	2015. 05. 29	1 308	95 938	19 061	27 052
03	2015. 05. 30	3 330	109 948	33 919	121 904
04	2015. 05. 31	9 431	213 551	90 099	188 940
05	2015. 06. 01	4 759	29 486	20 399	40 555
06	2015. 06. 02	2 817	188 797	50 634	84 709
07	2015. 06. 03	2 996	57 244	37 601	71 484
08	2015. 06. 04	2 842	155 920	53 580	105 945
09	2015. 06. 05	2 660	74 825	36 065	72 362
10	2015. 06. 06	1 471	15 991	10 993	25 581
11	2015. 06. 07	1 284	7 725	6 476	12 230
12	2015. 06. 08	1 509	30 667	19 079	28 190
13	2015. 06. 09	1 611	26 475	17 557	22 286
14	2015. 06. 10	1 726	34 413	13 977	12 324
15	2015. 06. 11	1 711	36 632	15 101	17 821
…	…	…	…	…	…

数据来源：新浪微博。

4. 4 舆情热度数据归一化处理

本文中原创微博发布量、转发量、评论量、点赞量均为效益型指标，根据公式（8）进行无量纲化处理，见表 4。

表 4 微博舆情热度数据无量纲化

序号	时间	原创微博发布量	转发量	评论量	点赞量
01	2015. 05. 28	0. 047993205	0. 131688996	0. 121426748	0. 038404328
02	2015. 05. 29	0. 137502654	0. 449225207	0. 211381121	0. 143046196
03	2015. 05. 30	0. 352197919	0. 514833217	0. 376325226	0. 645145068
04	2015. 05. 31	1	1	1	1
05	2015. 06. 01	0. 503928647	0. 138034382	0. 226234749	0. 2145243
06	2015. 06. 02	0. 297727755	0. 884078467	0. 561884568	0. 44825341
07	2015. 06. 03	0. 316733914	0. 268023471	0. 417200457	0. 378246899
08	2015. 06. 04	0. 300382247	0. 730117401	0. 594589194	0. 560666134
09	2015. 06. 05	0. 281057549	0. 350354265	0. 400148758	0. 382894591
10	2015. 06. 06	0. 154809938	0. 074838087	0. 121815295	0. 135259461
11	2015. 06. 07	0. 134954343	0. 036128893	0. 071670423	0. 064585969
12	2015. 06. 08	0. 158844765	0. 143564936	0. 211580946	0. 149070197
13	2015. 06. 09	0. 16967509	0. 123934045	0. 194684666	0. 117817385
14	2015. 06. 10	0. 181885751	0. 161107235	0. 154941773	0. 065083558
15	2015. 06. 11	0. 180293056	0. 171498682	0. 167419709	0. 094181916
…	…	…	…	…	…

4. 5 舆情热度计算

根据公式（10）~（12）计算得到原创微博发布量、转发量、评论量、点赞量对应的权重（保留小数点后 5 位）：$W_1 = 0.20364$，$W_2 = 0.28456$，$W_3 = 0.23797$，$W_4 = 0.27383$。根据公式（5）计算微博舆情热度，结果见表 5。

从图 3 可以看出，MERS 微博舆情热度的演变具有快速爆发、回落相对缓慢的特点，大致经历了萌动、加速、成熟、衰退 4 个阶段，基本符合网络舆情生命周期的特点。

表 5 MERS 微博舆情热度

序号	时间	MERS 微博舆情热度
01	2015. 05. 28	0. 086660857
02	2015. 05. 29	0. 245312423
03	2015. 05. 30	0. 48446897
04	2015. 05. 31	1. 00005
05	2015. 06. 01	0. 254490092
06	2015. 06. 02	0. 568681963
07	2015. 06. 03	0. 343643907
08	2015. 06. 04	0. 56398168
09	2015. 06. 05	0. 357021939
10	2015. 06. 06	0. 118854669
11	2015. 06. 07	0. 072507156
12	2015. 06. 08	0. 164377249
13	2015. 06. 09	0. 148416243
14	2015. 06. 10	0. 137580468
15	2015. 06. 11	0. 151151954
…	…	…

萌动期（5 月 28 日~5 月 30 日），公众开始关注该突发传染病事件，舆情热度上升明显，此后，网络舆情热度进入短暂的加速期（5 月 30 日~5 月 31 日），然后维持在成熟期（6 月 1 日~6 月 5 日）。尤其是在衰退期（6 月 6 日后）前期，出现明显的波动，体现为在 6 月 18 日和 6 月 26 日微博舆情热度剧增，出现两个小高峰，经过分析原创微博原文发现在这两日，“为抢救韩国 MERS 患者 广东 15 天花掉逾 800 万元”与“韩国籍 MERS 患者出院”两大话题引起网友广泛关注，致使微博舆情热度上升。

根据公式（6）计算 MERS 微博舆情热度趋势值，再由上文关于类别的计算方法，将微博舆情热度趋势值分为 6 类，$C1$ = [0. 25779052，0. 51558103]，$C2$ = [0. 12889526，0. 25779052]，$C3$ = [0，0. 12889526]，$C4$ = [-0. 18638998，0]，$C5$ = [-0. 37277995，-0. 18638998]，$C6$ = [-0. 745559908，-0. 37277995]，计算结果见表 6。

4. 6 BP 神经网络参数设置

隐含层和输出层的传递函数均采用双曲正切 S 型函数

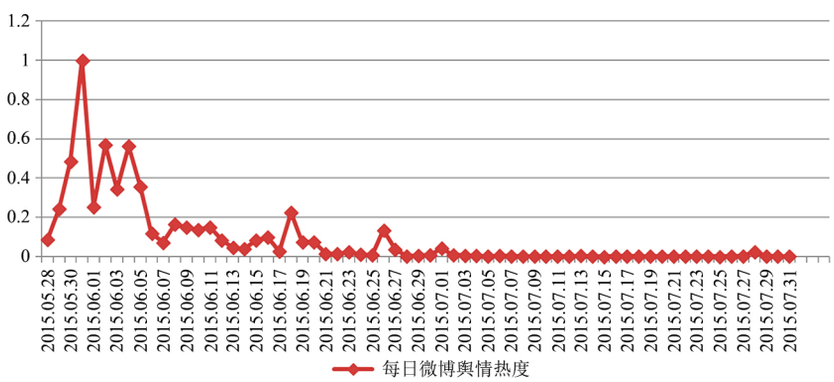

图 3 MERS 微博舆情热度曲线图

表 6 MERS 微博舆情热度趋势值

序号	时间	MERS 微博舆情热度	MERS 微博舆情热度趋势值	类别	二进制输出
01	2015. 05. 28	0. 086660857	0. 158651566	2	010
02	2015. 05. 29	0. 245312423	0. 239156547	2	010
03	2015. 05. 30	0. 48446897	0. 51558103	1	001
04	2015. 05. 31	1. 00005	-0. 745559908	6	110
05	2015. 06. 01	0. 254490092	0. 314191871	1	001
06	2015. 06. 02	0. 568681963	-0. 225038057	5	101
07	2015. 06. 03	0. 343643907	0. 220337773	2	010
08	2015. 06. 04	0. 56398168	-0. 20695974	5	101
09	2015. 06. 05	0. 357021939	-0. 23816727	5	101
10	2015. 06. 06	0. 118854669	-0. 046347513	4	100
11	2015. 06. 07	0. 072507156	0. 091870094	3	011
12	2015. 06. 08	0. 164377249	-0. 015961007	4	100
13	2015. 06. 09	0. 148416243	-0. 010835775	4	100
14	2015. 06. 10	0. 137580468	0. 013571487	3	011
15	2015. 06. 11	0. 151151954	-0. 066551899	4	100
…	…	…	…	…	…

“Tansig”，训练函数采用 Levenberg-Marquardt 反向传播算法训练函数 “Trainlm”[21]，训练目标误差为 0.005，学习率为 0.05，最大训练次数设置为1 000。各训练参数设置见表 7。

表 7 训练参数设置

N	训练参数	参数值
1	隐含层传递函数	Tansig
2	输出层传递函数	Tansig
3	训练函数	Trainlm
4	性能函数	Mse
5	显示间隔时间	net.trainParam.show = 100
6	目标误差	net.trainParam.goal = 0.005
7	学习率	net.trainParam.lr = 0.05
8	最大训练次数	net.trainParam.epochs = 1 000

4.7 突发传染病舆情热度趋势预测

本文将 2015 年 5 月 28 日至 2015 年 7 月 23 日期间的数据作为训练样本。根据隐含层节点数的公式（12），将 α 进行逐个试验，。先设定初始隐含层节点数为 3（$\alpha=1$ 时），然后训练 10 次，去掉最大和次大误差，取剩下 8 个误差的平均值并记录下来，再设置隐含层节点数为 4……一直到取隐含层节点数为 12（$\alpha=10$ 时），得到不同隐含层节点数下网络训练平均误差见图 4。

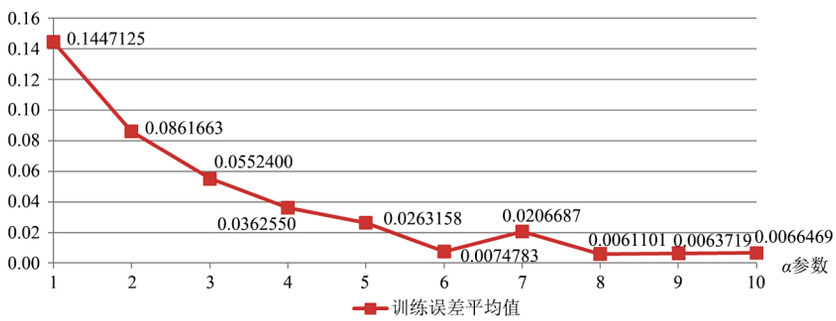

图 4 BP 神经网络训练平均误差

从图 4 可以看出，随着隐含层个数的增加，训练样本的 Mse 平均值基本呈下降趋势。当 $\alpha=8$ 时，即隐含层节点数 N = 10 时，网络误差最小。因此，本文所建 BP 神经网络拓扑结构为 “4-10-3”，将隐含层为 10 中训练误差较小、预测结果准确率较高的 BP 神经网络结构的参数保存。经过 141 次迭代，达到最小误差值 0.005848，网络训练提前停止，其训练结果混淆矩阵见图 5。

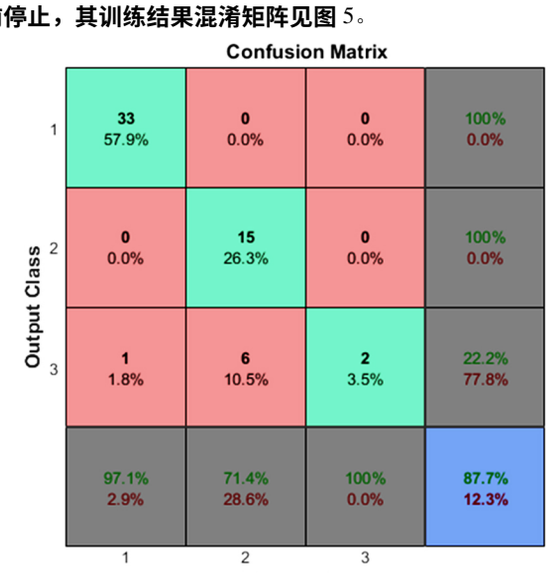

图 5 训练结果混淆矩阵

从图 5 可知，57 个训练样本的准确率为 87.7%，说明该模型的训练结果较为理想。基于此，通过训练好的网络预测 2015 年 7 月 24 日至 2015 年 7 月 31 日 MERS 舆情热度趋势值，对应的预测结果见表 8。

表 8 MERS 微博舆情热度趋势值预测结果

序号	时间	预测值	预测类别
58	2015. 07. 24	011	3
59	2015. 07. 25	011	3
60	2015. 07. 26	011	3
61	2015. 07. 27	011	3
62	2015. 07. 28	100	4
63	2015. 07. 29	100	4
64	2015. 07. 30	011	3
65	2015. 07. 31	100	4

4.8 突发传染病舆情热度趋势预测模型的评价

将预测得出的 MERS 微博舆情热度趋势值的类别与实际类别进行误差分析见表 9，模型预测结果的混淆矩阵见图 6。

表 9 微博舆情热度趋势预测模型误差分析

序号	时间	实际类别	预测类别	正确与否
58	2015. 07. 24	4	3	错误
59	2015. 07. 25	3	3	正确
60	2015. 07. 26	3	3	正确

表 9（续）

序号	时间	实际类别	预测类别	正确与否
61	2015. 07. 27	3	3	正确
62	2015. 07. 28	4	4	正确
63	2015. 07. 29	4	4	正确
64	2015. 07. 30	3	3	正确
65	2015. 07. 31	4	4	正确

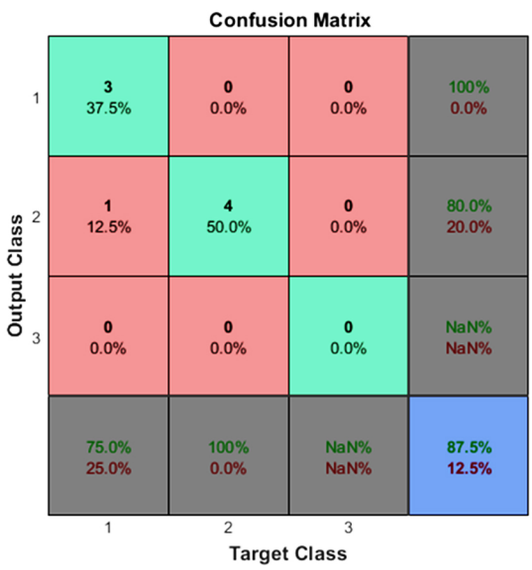

图 6 模型预测结果混淆矩阵

由表 9 的误差分析以及图 5 的混淆矩阵可以看出，基于 BP 神经网络预测的 MERS 微博舆情热度趋势的后 7 个类别与实际相符，只有第一个类别预测有偏差，准确率达到 87. 5%，预测结果较为理想。说明基于 BP 神经网络模型的突发传染病舆情热度趋势的预测是可行的。

5 结 语

突发传染病发生后，由于严重危及生命安全，事件相关信息往往迅速在网络上传播交流，从而形成网络舆情。本文将 BP 神经网络应用到突发传染病舆情热度趋势的预测中。首先在查阅文献的基础上，针对微博自身的特点，提出舆情热度的定量方法，然后利用信息熵确定各个指标的权重，最后建立基于 BP 神经网络的突发传染病舆情热度趋势预测模型，并选取新浪微博“MERS 病毒卫生突发事件”的舆情热度数据进行实例分析，预测该突发传染病事件的发展趋势，从而验证模型的可行性。研究结果可以为突发传染病微博舆情的管控提供决策支持。

然而，本文研究也存在一定的局限性：第一，本文仅针对新浪微博这一媒介平台建立指标体系，忽略了微信、论坛、新闻网站等互联网平台上的舆情信息，而且选取的微博舆情热度评价指标只包括新浪微博的原创微博发布量、转发量、评论量，点赞量四项指标，较为单一，忽略了微博内容主题特征、博主类型，博主粉丝量等信息对舆情热度的影响，该体系还有待进一步完善；第二，突发传染病舆情具有爆发性、演变不确定性等特点，在一段时间内（1 天或者数小时）可能会有大幅度的波动，本文以天为单位进行数据统计，在未来的研究中还需考虑选取更加细粒度的单位进行分析；第三，本文只选取了一个突发传染病案例进行实证分析，结果表明模型预测的准确率较高，但对于该模型是否同样适用于其他突发传染病舆情尚未进行探讨。因此，在未来的研究中还需要选取一定数量的案例来验证模型的合理性和有效性。

参 考 文 献

[1] 中国互联网络信息中心（CNNIC）. 第 40 次中国互联网络发展状况统计报告 [R]. 中国互联网络信息中心，2017.

[2] 安璐，杜廷尧，余传明，等. 突发公共卫生事件的微博主题演化模式和时序趋势 [J]. 情报资料工作，2016，(5)：44-52.

[3] 徐旖旎. 基于微博的媒体奇观网络舆情热度趋势分析 [J]. 情报科学，2017，35 (2)：92-97，125.

[4] 赵磊，王松. 基于 BP 神经网络的舆情热度趋势仿真模型研究 [J]. 情报学报，2016，35 (9)：989-999.

[5] Yu L，Li L，Tang L. What Can Mass Media do to Control Public Panic in Accidents of Hazardous Chemical Leakage into Rivers? A Multi-Agent-Based Online Opinion Dissemination Model [J]. Journal of Cleaner Production，2017，(143)：1203-1214.

[6] 张行钦，张东红，付刚厪，等.“乙肝疫苗”事件网络舆情热度演变特点及应对研究 [J]. 中国预防医学杂志，2017，18 (1)：60-62.

[7] Lax JR，Phillips JH. How Should we Estimate Public Opinion in the States? [J]. American Journal of Political Science，2009，53 (1)：107-121.

[8] 曹学艳，张仙，刘樑，等. 基于应对等级的突发事件网络舆情热度分析 [J]. 中国管理科学，2014，22 (3)：82-89.

[9] 王慧军，石岩，胡明礼，等. 舆情热度的最优监控问题研究 [J]. 情报杂志，2012，31 (1)：71-75.

[10] 袁国平，许晓兵. 基于系统动力学的关于突发事件后网络舆情热度研究 [J]. 情报科学，2015，33 (10)：52-56.

[11] 屈启兴，齐佳音. 基于微博的企业网络舆情热度趋势分析 [J]. 情报杂志，2014，33 (6)：133-137.

[12] 王新猛. 基于马尔可夫链的政府负面网络舆情热度趋势分析——以新浪微博为例 [J]. 情报杂志，2015，34 (7)：161-164.

[13] Chen XG，Duan S，Wang L. Research on Trend Prediction and Evaluation of Network Public Opinion [J]. Concurrency and Computation: Practice Practice and Experience，2017，29 (24)：e4212.

（下转第 52 页）

[16] Wrench J S，Punyanunt-Carter N M. The Relationship Between Computer-Mediated-Communication Competence，Apprehension，Self-Efficacy，Perceived Confidence，and Social Presence [J]. Southern Communication Journal，2007，72 (4)：355-378.

[17] Dholakia U M，Bagozzi R P，Pearo L K. A Social Influence Model of Consumer Participation in Network-and Small-Group-Based Virtual Communities [J]. International Journal of Research in Marketing，2004，21 (3)：241-263.

[18] Houtman E，Makos A，Meacock H L. The Intersection of Social Presence and Impression Management in Online Learning Environments [J]. E-Learning and Digital Media，2014，11 (4)：419-430.

[19] Lu Y，Zhao L，Wang B. From Virtual Community Members to C2C E-commerce Buyers：Trust in Virtual Communities and Its Effect on Consumers´ Purchase Intention [J]. Electronic Commerce Research and Applications，2010，9 (4)：346-360.

[20] Zhao L，Lu Y，Wang B，et al. Cultivating the Sense of Belonging and Motivating User Participation in Virtual Communities：A Social Capital Perspective [J]. International Journal of Information Management，2012，32 (6)：574-588.

[21] Ogonowski A，Montandon A，Botha E，et al. Should New Online Stores Invest in Social Presence Elements? The Effect of Social Presence on Initial Trust Formation [J]. Journal of Retailing and Consumer Services，2014，21 (4)：482-491.

[22] Blanchard A L，Markus M L. The Experienced Sense of a Virtual Community：Characteristics and Processes [J]. ACM Sigmis Database，2004，35 (1)：64-79.

[23] Kankanhalli A，Tan B C Y，Wei K K. Contributing Knowledge to Electronic Knowledge Repositories：An Empirical Investigation [J]. MIS Quarterly，2005，29 (1)：113-143.

[24] 金晓玲，汤振亚，周中允，等. 用户为什么在问答社区中持续贡献知识?：积分等级的调节作用 [J]. 管理评论，2013，25 (12)：138-146.

[25] 秦敏，乔晗，陈良煌. 基于 CAS 理论的企业开放式创新社区在线用户贡献行为研究：以国内知名企业社区为例 [J]. 管理评论，2015，27 (1)：126-137.

[26] Lu Y，Zhao L，Wang B. From Virtual Community Members to C2C E-commerce Buyers：Trust in Virtual Communities and Its Effect on Consumers´ Purchase Intention [J]. Electronic Commerce Research and Applications，2010，9 (4)：346-360.

[27] Gharib R K，Philpott E，Duan Y. Factors Affecting Active Participation in B2B Online Communities：An Empirical Investigation [J]. Information & Management，2017，54 (4)：516-530.

[28] 范晓屏，韩洪叶，孙佳琦. 网站生动性和互动性对消费者产品态度的影响——认知需求的调节效应研究 [J]. 管理工程学报，2013，27 (3)：196-204.

[29] 赵宏霞，王新海，周宝刚. B2C 网络购物中在线互动及临场感与消费者信任研究 [J]. 管理评论，2015，27 (2)：43-54.

[30] Gefen D，Straub D W. Consumer Trust in B2C E-Commerce and the Importance of Social Presence：Experiments in E-Products and E-Services [J]. Omega，2004，32 (6)：407-424.

[31] 黄敏学，廖俊云，周南. 社区体验能提升消费者的品牌忠诚吗——不同体验成分的作用与影响机制研究 [J]. 南开管理评论，2015，18 (3)：151-160.

[32] 王哲. 社会化问答社区知乎的用户持续使用行为影响因素研究 [J]. 情报科学，2017，35 (1)：78-83.

(责任编辑：孙国雷)

(上接第 44 页)

[14] 靳松，庄亚明. 基于 H7N9 的突发事件信息传播网络簇结构特性研究 [J]. 情报杂志，2013，32 (12)：12-17.

[15] 安璐，周思瑶，余传明，等. 突发传染病微博影响力的预测研究 [J]. 情报科学，2017，35 (4)：27-31.

[16] 杜洪涛，滕琳，赵志云. 突发性传染病舆情中的公共管理沟通效果研究——以中东呼吸综合征疫情微博社区舆情为例 [J]. 情报杂志，2017，36 (2)：108-114.

[17] 翁士洪，顾佩丽. 公共突发事件中微博谣言的机制与治理——以 H7N9 事件为例 [J]. 电子政务，2015，(10)：10-18.

[18] 魏志惠，何跃. 基于信息熵和未确知测度模型的微博意见领袖识别——以“甘肃庆阳校车突发事件”为例 [J]. 情报科学，2014，32 (10)：38-43.

[19] 朱喜安，魏国栋. 熵值法中无量纲化方法优良标准的探讨 [J]. 统计与决策，2015，(2)：12-15.

[20] 赵志勇. 简单易学的机器学习算法——神经网络之 BP 神经网络 [EB/OL]. http://blog.csdn.net/google19890102/article/details/32723459，2017-09-01.

[21] 陈明. MATLAB 神经网络原理与实例精解 [M]. 北京：清华大学出版社，2013：95.

(责任编辑：郭沫含)