

doi:10.11835/j.issn.1008-5831.2018.06.004

欢迎按以下格式引用:张宇,傅敏,罗加蓉.震灾网络舆情风险监测指标及其评估方法[J].重庆大学学报(社会科学版),2018(1):33-44.

Citation Format: ZHANG Yu, FU Min, LUO Jiarong. Risk monitoring indexes and assessment method of network public opinion of earthquake disaster [J]. Journal of Chongqing University (Social Science Edition), 2018(1):33-44.

震灾网络舆情风险监测 指标及其评估方法

张宇¹,傅敏²,罗加蓉¹

(1.西南科技大学经济管理学院,四川绵阳 621010;2.四川雪宝乳业集团有限公司,四川绵阳 621000)

摘要:为科学构建震灾网络舆情风险评估体系,在舆情监测指标构建基础上提出基于加速遗传算法的BP神经网络(AGABP)风险评估方法。依据网络舆情演化理论,围绕震灾网络舆情的物理属性和社会属性提出2个维度、4个二级指标和10个三级指标的震灾网络舆情风险监测指标体系。在评估方法上,针对常规评估系统对非线性、高维度和非正态评估问题的局限性,利用BP网络能够以任意均方差的精度逼近任意平方可积非线性连续函数的优势,将BP网络用于震灾网络舆情风险监测评估中,并用加速遗传算法(AGA)对BP网络参数进行优化,以解决常规BP网络存在训练速度慢和容易过早收敛的问题。通过随机样本数据对AGABP模型进行自学习训练,并用实际样本数据验证,结果表明:与BP神经网络、逻辑斯蒂曲线相比,本研究所构建的AGABP模型在收敛速度、评估准确度上有明显优势,能够用于震灾网络舆情风险管理实践中。

关键词:震灾网络舆情;风险评估;AGABP模型;风险监测指标

中图分类号:D632.5;O233

文献标志码:A

文章编号:1008-5831(2018)06-0033-12

一、文献回顾与问题提出

截至2017年6月,中国网民规模达到7.51亿,互联网普及率达54.3%。互联网的普及为网民表达自己对事件的态度和看法提供了便捷途径,这种对事件的态度和看法形成网络舆情,网络舆情通过网络媒体的传播和扩散,在一定程度上能够促进或阻碍事态发展。地震造成基础设施、人员和财产的巨大损失,在各种因素作用下震灾舆情增加了震灾的应对难度,甚至造成更为严重的次生或衍

修回日期:2018-06-12

基金项目:国家自然科学基金项目“多源风险下基于期权契约的供应链优化模型研究”(71702156)

作者简介:张宇(1974—),男,四川人,西南科技大学经济管理学院副教授,博士,主要从事信息经济、信息计量学研究,Email: yuzhang@swust.edu.cn。

生灾害。

现行的各种舆情管控实践证明,如果能够对震灾网络舆情风险进行科学评估,一方面能够有效减轻震灾带来的灾害损失和社会危害,另一方面能够为应急管理部门掌握灾区需求信息和制定舆情管控预案提供支持。目前相关研究主要集中于网络舆情的热度、传播规律、风险管控等方面。在网络舆情热度研究方面,王慧军等^[1]为降低舆情热度负面影响,依据优化理论,研究了政府对舆情热度监控问题;柳虹^[2]提出了舆情热点发现模型,对均值聚类和支持向量机两种算法在网络舆情发现作用方面进行了比较;方倩等^[3]认为微信公众号文章热度与主题、推送时间、标题特征相关,与推送频率弱相关;焦超等^[4]提出了独立网络突发事件热度分布的泊松分布模型;翟晓芳等^[5]通过分析微博提出了新的热度指标计算方法;曹学艳等^[6]研究了突发事件网络舆情热度等级与突发事件风险等级之间的对应关系,根据对应关系的“一致”或“不一致”提出突发事件应对策略;张一文等^[7]研究了非常规突发事件网络舆情热度评估指标和权重;魏德志等^[8]将新闻网页内容的相似度和页面链接分析作为话题热度的计算依据,提出一种基于时间序列的热点话题发现模型;Fu等分别采用不同聚类算法和仿真模型对网络热点时间进行发现和模拟^[9-12]。在网络舆情传播研究方面,王治莹等^[13]研究了政府干预下突发事件舆情传播规律与管控决策问题;Huo等^[14]建立了突发事件下官方行为与舆情传播的交互作用模型;Lebensztayn等^[15]构建了群众谣言传播的马尔科夫模型;Zhao等^[16]研究了突发事件演化与舆情传播及官方媒介的交互作用;齐佳音等^[17]研究了突发公共事件网络舆情的耦合机制,认为网络舆情的耦合协调度与社会影响力呈正比;康伟^[18]研究了网络舆情传播结构对信息的传播路径、速度和范围的影响。在网络舆情风险管控方面的研究,张玉亮^[19]依据突发事件发生周期系统提出网络舆情风险评估指标;路枝芳等^[20]以舆情热点事件为验证对象,采用因子聚类分析方法研究了2015年中国第一季度各类热点舆情事件的风险程度。

综上,现有研究主要有两方面需要继续深入:一是在指标构建上,现有研究更多考虑事件的外因,即社会属性,如舆情转载量、评论量及发文量等,较少考虑事件本身的物理属性,如事件自身导致舆情演化的动因、敏感性、促进因素等。对外因的考虑也多从定量角度提取指标,较少关注一些定性指标,尤其是面对突发灾害背景下网络舆情监测所涉及的关键不确定性指标;二是针对震灾网络舆情综合风险评估问题较少涉及,已有研究更多围绕舆情热点的分布、线性拟合及预测等方面,少量成果针对突发事件风险提出监测指标与评估方法,但监测指标更多突出舆情风险的社会属性,很少涉及导致舆情风险的物理属性,而且评估方法较为单一,往往忽略突发灾害网络舆情指标数据的非线性、高维度和非正态问题,评估结果与实际偏差较大。基于以上原因,本研究的不同之处在于:从突发灾害网络舆情的物理属性和社会属性两方面提出适应性强、能够覆盖整个舆情演化周期的风险监测指标;将BP神经网络运用于突发灾害网络舆情综合风险评估之中,并用AGA对BP神经网络进行改进,提出基于AGABP的震灾网络舆情风险评估方法。

二、震灾网络舆情风险监测指标

(一) 指标构建

网络舆情主要有潜伏、扩散和消退三个阶段,其传播曲线呈“S”型(如图1),其风险点一般处于传播中的扩散前期和扩散后期的转折处,如图1中的N点,N点也是舆情高涨转折点。

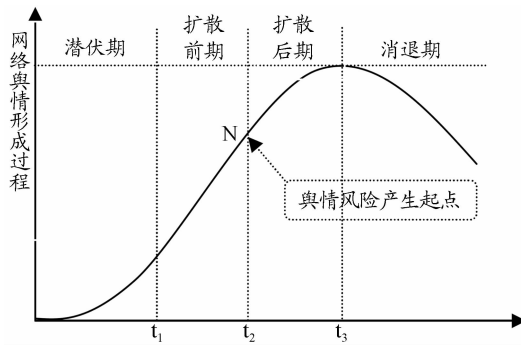


图1 网络舆情传播过程

在图1中,网络舆情传播风险的产生和高涨的出现主要由载体、主体和事件本身共同推动完成。网络媒体是震灾舆情产生的载体,属条件因素;网民是震灾舆情产生的主体,属推动因素;震灾是舆情事件的本身,属舆情产生的诱导因素。条件因素和推动因素是震灾网络舆情扩散和传播的社会属性,诱导因素是震灾网络舆情的物理属性,敏感性对舆情的产生起诱导作用。在风险监测指标设计上,本研究围绕震灾网络舆情产生的物理属性和社会属性两个维度的作用力设计风险监测指标。

其一,震灾网络舆情产生的物理属性。这由震灾本身的危害程度决定,一般来讲,震灾危害程度越大,社会影响力就越高,诱发网络舆情扩散的能力也会越强。可采用震灾危害等级、受灾程度等二级指标予以衡量,其中震灾危害等级主要包括地震震级和震中烈度2个三级指标,受灾程度主要包括伤亡人数、受灾范围、财产损失3个三级指标。

其二,震灾网络舆情产生的社会属性。可采用震灾应急能力、网络媒体作用力、舆情扩散能力3个二级指标予以衡量,其中震灾应急能力对减轻灾害损失、缓解网民负面情绪有直接影响,主要包括应急资源满足能力、政府部门反应能力、监测预警能力3个三级指标;网络媒体作用力主要包括各主要网络平台的发文量、评论数及转发量3个三级指标;舆情扩散能力主要包括点击数与回复数之比、原发文数变化率2个三级指标。

综上所述,本研究围绕物理属性和社会属性两个维度提出5个二级指标和13个三级指标的震灾网络舆情风险监测评估指标,进一步通过文献调查、专家问询等方式,在震灾危害等级中增加1个次生灾害发生概率的三级指标,在震灾应急能力中增加1个灾民综合满意度的三级指标,删除舆情扩散能力指标,形成4个二级指标和13个三级初选指标。为验证初选指标之间相对独立性、与评估目标的相关性及代表性,本研究采取专家计分的方式获取了13个初选指标的初始数据,并采用Cronbach's Alpha参数来度量问卷信度,通过做KMO和Bartlett球形检验来验证问卷表的效度。通过检验,总问卷信度值Cronbach's Alpha = 0.871,说明问卷可靠性很高;KMO = 0.743, Bartlett球形检验为0.000,说明问卷效度很好,可用来做因子分析。检验结果发现Bartlett球形检验的 $P < 0.001$,说明初选指标中有重复解释指标存在,通过分析指标两两之间的相关关系,删掉两两之间相关度 > 0.6 以上的指标1个,即删掉“受灾范围”,保证剩下12个初选指标的相对独立。为检验剩下12个初选指标与评估对象之间的相关度,进一步做因子相关性分析,删掉相关度 < 0.65 的指标1个,即删掉“震中烈度”指标,保障剩下11个指标与被评估对象高度相关。本研究采用主成分分析方法寻找剩下指标,将累积贡献率达85%的指标作为最终评估指标。最终形成2个维度4个二级指标和10个

三级指标作为震灾网络舆情风险监测指标(如表1)。

表1 震灾网络舆情风险监测指标体系

维度	二级指标	三级指标	数据获取
物理属性	震灾危害等级	地震震级	根据官方公布数据
		次生灾害发生率	根据历史数据估计具有危险性次生灾害发生概率,4级以上余震发生概率设定为:主震为六级,次生灾害发生率为30%;主震为7级,次生灾害发生率为60%;主震8级以上,次生灾害发生率为90%。
	受灾程度	伤亡人数 财产损失	以官方公布数据为准 以官方公布数据为准
社会属性	震灾应急能力	应急资源满足能力	5级量表赋值:强(5)、较强(4)、一般(3)、弱(2)、很弱(1)
		监测预警能力	5级量表赋值:强(5)、较强(4)、一般(3)、弱(2)、很弱(1)
		灾民综合满意度	5级量表赋值:强(5)、较强(4)、一般(3)、弱(2)、很弱(1)
	网络作用力	发文量	在规定期间统计主流媒介发文量
		评论数	在规定期间统计主流媒介评论数
		转发量	在规定期间统计主流媒介转发量

表1中的震灾应急能力所属的三级指标采取专家计分方式确定;网络作用力所属三级指标采用北京清博大数据科技有限公司开发的“清博大数据舆情系统”平台中的数据,通过统计微信、微博、网页、报刊、客户端和论坛6个舆情传播平台上地震网络舆情的发文量、评论数和转发量数据。

(二) 建立风险等级标准

建立震灾网络舆情风险等级标准包括两部分内容:设定风险等级数和确定各风险等级下每个指标的标准区间值。在设定震灾网络舆情风险等级数上,本研究参考国际惯例及国内相关学者研究,将震灾网络舆情风险等级划分为4级,即 $v = \{IV, III, II, I\} = \{\text{极高风险, 高风险, 一般风险, 低风险}\}$, 并赋值:“极高风险”为4、“高风险”为3、“一般风险”为2、“低风险”为1。在确定各风险等级下每个指标的标准区间上,以上海交通大学舆情研究实验室“中国公共事件数据库”为依托,从500多起震灾案例中挑选出2000年以来震级在3~9之间的150起案例作为研究样本,针对定性指标按表1赋值后采取问卷方式获取数据;针对定量指标,以特定期间客观数据为准,其中,地震震级、伤亡人数和财产损失风险标准依据《国家地震应急预案》确定,其余指标的风险等级划分借鉴蒋金才等^[21]对水旱灾害风险等级划分方法,对获取的指标数据采用聚类方法进行修正后得到震灾网络舆情风险等级标准值。由于震灾背景下的网络舆情风险指标数据往往具有不确定性,各指标的风险等级标准难以用确切数据划分,故采用模糊区间数表示各指标的风险等级更为合理。鉴于此,本研究以150起案例为依据,采用四分位风险划分方法来确定震灾网络舆情风险等级标准。具体做法是:首先,将各个指标的专家建议值 $Z(i)$ 进行升序排列,设定 $Z(i)_{\max}$ 为极高风险上限值, $Z(i)_{\min}$ 为最低风险下限值,中位数 $Z(i)_{\text{med}}$ 为一般风险值;然后,找出 $Z(i)_{\min}$ 和 $Z(i)_{\text{med}}$ 序列中的中位数 $Z^*(i)_{\text{med}}$ 、 $Z(i)_{\text{med}}$ 和 $Z(i)_{\max}$ 序列的中位数 $Z^{**}(i)_{\text{med}}$, 其风险等级划分方法如表2。

表2 震灾网络舆情风险等级标准

风险等级	I级	II级	III级	IV级
区间值	$(-\infty, z^*(i)_{med})$	$[z^*(i)_{med}, z(i)_{med})$	$[z(i)_{med}, z^{**}(i)_{med})$	$[z^{**}(i)_{med}, +\infty)$

最后,依据表2方法,计算各个风险等级下每个指标的标准区间值,如表3。

表3 震灾网络舆情风险等级标准

指标	低风险=1	一般风险=2	高风险=3	极高风险=4
	I	II	III	IV
地震震级	[1,3)	[3,5)	[5,7)	[7,+∞)
次生灾害发生率	[0,0.2)	[0.2,0.45)	[0.45,0.65)	[0.65,1]
伤亡人数	[0,10)	[10,50)	[50,300)	[300,+∞)
财产损失(亿元)	[0,0.01 * GDP)	[0.01 * GDP, 0.06 * GDP)	[0.06 * GDP, 0.1 * GDP)	[0.1 * GDP, +∞)
应急资源满足能力	[4,5]	[3,4)	[2,3)	[1,2)
监测预警能力	[4,5]	[3,4)	[2,3)	[1,2)
灾民综合满意度	[4,5]	[3,4)	[2,3)	[1,2)
发文量	$(-\infty, 40\ 000)$	[40 000, 100 000)	[100 000, 500 000)	[500 000, +∞)
评论数	$(-\infty, 240\ 000)$	[240 000, 600 000)	[600 000, 3 000 000)	[3 000 000, +∞)
转发量	$(-\infty, 120\ 000)$	[120 000, 300 000)	[300 000, 1 500 000)	[1 500 000, +∞)

三、AGABP 评估模型构建

BP神经网络是一种负梯度优化算法,具有自适应、自组织、容错性和鲁棒性强的优点,而且易于在计算机上实现编译。将BP神经网络用于震灾舆情网络风险评估点评估中,能够有效解决多指标非线性、高维度和非正态评估问题的局限性,有利于评估系统以任意均方差的精度逼近任意平方可积非线性连续函数,保证评估结果的可靠。但BP神经网络学习速度较慢、存在局部最小问题,会在很大程度上影响网络的外推能力。为解决这个问题,采用AGA对常规BP神经网络进行改进,在BP算法训练评估网络出现收敛速度缓慢时启动AGA来优化网络参数,把优化的参数作为BP算法的初始值。通过这样的改进,能够有效增强BP网络的外推能力,防止训练网络进入局部循环。

(一) 构建步骤

由于存在一个三层BP神经网络可以任意精度逼近从输入层到输出层的映射,对输出层节点 n 和隐层节点 m 的大小没有严格限制。根据经验,隐层的节点数 m 一般不宜过大, m 较大,BP网络的概括能力和训练速度都会降低,在能达到期望精度条件下, m 的取值范围一般控制在 $[n, 2n+1]$ 内且尽可能小。本研究构建一个三层BP神经网络(如图2),其拓扑结构为 $n:n:1, n$ 取3。

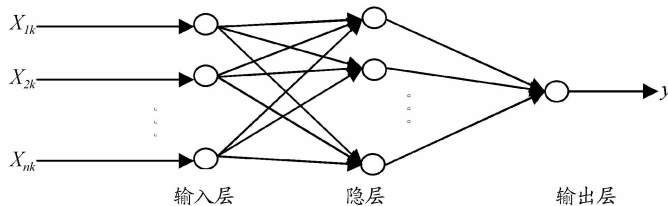


图2 三层BP神经网络拓扑结构

记输入层神经元为 h ,隐层神经元为 i ,输出神经元为 j ,隐层节点 i 和输出层节点 j 的阈值分别为 θ_i, θ_j ,输入层节点 h 与隐层节点 i 间以及输出层节点 j 与隐层节点 i 间的连线权值分别为 w_{hi}, w_{ij} ,

各点的输入、输出分别为 x, y 。基于 AGA 的 BP 网络算法设计如下。

步骤 1: 指标随机样本点获取。在后续的 BP 神经网络训练中, 首先需要随机生成无量纲的震灾网络舆情风险等级样本数据, 通过随机样本训练得到最优参数后, 再用实际样本数据进行测试。在随机样本数据获取上, 本研究参考王硕等^[22]的研究成果, 用均匀随机数在各级指标变化区间范围内随机产生 nu 个指标样本值 $x^*(k, j)$, 相应的震灾网络舆情风险标准等级值为 $y(k) = i$ 。为充分反映评估指标中各指标的边界值的信息, 取各指标边界值各 1 次, 风险等级值取与该边界值有关的两个风险等级值的算术平均值, 这样就可得震灾网络舆情风险评估标准样本系列 $\{x^*(k, j), y(k)\}, k = 1 \sim nk, j = 1 \sim nj$, 其中 nk 为样本数目。为消除指标量纲, 使本研究评估模型具有一般性, 各指标的无量纲化处理采取以下式子进行:

$$x(k, j) = x^*(k, j) / X_{\max}(j) \quad (k = 1 \sim nk, j = 1 \sim nj) \quad (1)$$

$$a(i, j) = a^*(k, j) / x_{\max}(j) \quad (i = 1 \sim ni, j = 1 \sim nj) \quad (2)$$

$$b(i, j) = b^*(k, j) / x_{\max}(j) \quad (i = 1 \sim ni, j = 1 \sim nj) \quad (3)$$

步骤 2: 初始化。设用于机器学习的已做归一化处理后的震灾舆情网络样本值的输入、输出样本对为 $\{x_{hk}, d_k | h = 1, 2, \dots, n; k = 1, 2, \dots, N\}$, 给各节点间的连接权值、阈值赋予 $(-1, 1)$ 区间上的随机值。

步骤 3: 置 $k = 1$ 。把各层输入、输出样本对 $\{x_{hk}, d_k\}$ 提供给网络, $(h = 1, 2, \dots, n; k = 1, 2, \dots, N)$ 。

步骤 4: 计算隐层各节点的输入 x_i 、输出 y_i , 输出层节点的输入 x_j 、输出 y_j , 则有:

$$x_i = \sum_{h=1}^n w_{hi} x_{hk} + \theta_i, y_i = 1 / (1 + e^{-x_i}), (i = 1, 2, \dots, n) \quad (4)$$

$$x_j = \sum_{i=1}^n w_{ij} y_i + \theta_j, y_j = 1 / (1 + e^{-x_j}) \quad (5)$$

步骤 5: 计算输出层节点所收到的总输入变化时单样本点误差 $E_k = 0.5(y_j - d_k)^2$ 的变化率以及隐层节点所收到的总输入变化时单样本点误差 E_k 的变化率。

$$\frac{\partial E_k}{\partial x_j} = y_i(1 - y_i)(y_i - d_k) \quad (6)$$

$$\frac{\partial E_k}{\partial x_i} = y_i(1 - y_i) \left(\frac{\partial E_k}{\partial x_j} w_{ij} \right), (i = 1, 2, \dots, n) \quad (7)$$

步骤 6: 修正各连接的权值和阈值。

$$w_{ij}^{m+1} = w_{ij}^m - \eta y_i \frac{\partial E_k}{\partial x_j} + a(w_{ij}^m - w_{ij}^{m-1}), \theta_j^{m+1} = \theta_j^m - \eta \frac{\partial E_k}{\partial x_j} + a(\theta_j^m - \theta_j^{m-1}) \quad (8)$$

$$w_{hi}^{m+1} = w_{hi}^m - \eta x_{hk} \frac{\partial E_k}{\partial x_i} + a(w_{hi}^m - w_{hi}^{m-1}), \theta_i^{m+1} = \theta_i^m - \eta \frac{\partial E_k}{\partial x_i} + a(\theta_i^m - \theta_i^{m-1}) \quad (9)$$

其中, m 为修正次数, η 为学习速率且 $\eta \in (0, 1)$, a 为动量因子且 $a \in (0, 1)$ 。

步骤 7: 置 $k = k + 1$, 转步骤 3, 直至全部 N 个样本点训练完毕, 转步骤 9。

步骤 8: 转步骤 2, 进行新一轮的学习, 直至网络全局误差函数小于预先设定的一个较小值或学习次数大于预先设定的值, 结束学习。

$$E = \sum_{k=1}^N E_k = \sum_{k=1}^N (y_j - d_k)^2 / 2 \quad (10)$$

为使式(10)的全局误差函数极小化,需要确定 BP 网络中 θ_i 、 θ_j 、 w_{hi} 和 w_{ij} 的最优值,以促使各层级网络连接的权值和阈值达到稳定。标准 BP 神经网络在训练参数过程中容易出现局部最小问题,且训练后的参数在学习后期收敛速度过于缓慢,在较大程度上影响评估结果的可靠性。本研究采用 AGA 对 BP 网络参数进行优化,将优化后的参数作为 BP 网络的初始值,避免标准 BP 网络的不足。其优化步骤如下^[23]。

(1)BP 网络参数变化区间构造。设 c_j 是 BP 算法训练网络出现收敛速度缓慢时网络的任一参数的值,则它的变化区间构造为 $[a_j, b_j]$,其中, $a_j = c_j - d|c_j|$, $b_j = c_j + d|c_j|$, d 为一正的常数。

(2)BP 网络参数编码。设编码长度为 e ,把区间 $[a_j, b_j]$ 等分成 $2^e - 1$ 个子区间,于是整个网络参数变化空间被离散成 $(2^e)^p$ 个网格点。其中, $p = 2n^2 + n + 1$ 。每个网格点称为个体,它对应网络 p 个参数的一种可能取值状态,并用 p 个 e 位二进制数表示。于是, p 个网络参数、网格点、个体、二进制数予以一一对应。

(3)初始父代群体的随机生成与父代个体适应度的评估。从上述 $(2^e)^p$ 个网格点中均匀随机选取 n 个点作为初始父代群体。把第 i 个个体代入式(7)中,得到相应的网络全局误差函数值 E_i , E_i 越小则个体的适应能力越强。

(4)父代个体的选择与杂交。把父代个体按优化准则函数值 E_i 从小到大排序,排序后最前面几个个体称为优秀个体。构造与 E_i 成反比的函数 p_i , $p_i > 0$, $p_1 + p_2 + \dots + p_n = 1$,从这些父代个体中以概率 p_i 选择 i 个个体。于是共选择两组各为 n 个的个体,然后将两组个体随机两两配对 n 对双亲,将双亲的二进制数组的任意一段值互换,得到两组子代个体。

(5)子代个体的变异。任取步骤父代个体杂交中的一组子代个体,将它们的二进制数组的任意两值以变异率 p_m 进行翻转,即将原值为 0 的变为 1,原值为 1 的变为 0。

(6)迭代。由步骤(6)得到的 n 个子代个体作为新的父代,算法转入父代个体适应度评估步骤,进入下一代进化过程。

(7)加速循环。用第一次、第二次进化迭代所产生的优秀个体参数的变化范围,作为参数新的初始变化区间,算法进入网络参数编码步骤,即上述步骤(2);如此循环往复,优秀个体的参数变化区间将逐步收缩,与最优的距离越来越近,直至达到给定加速次数,结束运行。

步骤 9:把待测的第 k 期对应各单项评估模型的评估值作为输入样本,输入已学习完毕的网络,其网络输出值经过与归一化对应的逆处理即为组合评估值 F_k 。

(二)控制参数配置

在 AGA-BP 算法中,根据经验,控制参数预先设定为:(1)设定 BP 网络中的学习因子 $\eta = 0.1$,动量系数 $\alpha = 0.1$;(2)设定 AGA 中的编码长度 $e = 10$,变异率 $p_m = 1.0$,父代个体数目 $q = 300$,优秀个体数目 $s = 10$ 。

四、实例应用

(一)验证数据选取

计算机运行环境为:Core(TM) 2 CPU 2.29GHZ、内存为 2.00GB,仿真工具:Matlab_R2012a。本研究从《中国震例》中选取 2005 年以来国内发生的 6 次地震作为 AGABP 评估模型的验证数据,数据获取途径包括定性指标专家评分、清博大数据舆情系统、《中国震例》及官方网站(如表 4)。其

中,网络舆情转发量、发文量及评论量的统计期间规定为震灾发生当日后30日的数据量;震级、死伤人数及财产损失以震灾发生地官方网站公布为准。评估方法应用分为两个阶段:第一阶段采用随机样本数据训练AGABP网络以获得最优参数,并与BP神经网络、逻辑斯蒂曲线对比,以观察AGABP评估模型的训练精度;第二阶段将6次地震网络舆情数据装入训练后的AGABP模型中,对6次震灾网络舆情的风险性进行评估,并提出政策建议。

表4 2005年以来6次震灾网络舆情风险监测指标数据

序号	评估指标							发文量	评论数	转发数
	地震震级	次生灾害发生率	伤亡人数	财产损失	应急资源满足能力	监测预警能力	灾民综合满意度			
1	5.7	0.30	416	20	3	3	2	5 200	110 970	35 098
2	6.4	0.25	31	18.98	3	3	2	12 390	136 708	72 390
3	8.0	0.85	443 870	8 451.4	4	3	3	160 876	3 245 626	978 690
4	7.1	0.80	9 727	8 000	4	4	4	224 560	4 321 567	1 608 900
5	7.0	0.80	11 579	422.6	4	3	3	189 678	5 789 430	997 860
6	6.5	0.55	3 760	4.6	2	3	2	6 900	21 000	35 067

注:表中各序号对应地震为:1.江西九江(2005.11.26);2.云南宁洱(2007.6.3);3.四川汶川(2008.5.12);4.青海玉树(2010.4.14);5.四川雅安地震(2013.4.20);6.云南鲁甸(2014.8.3)

(二) AGABP 网络参数训练与实例评估

1. 随机样本数据训练 AGABP 网络

根据表3的震灾网络舆情风险等级标准和步骤1,随机生成1~31组震灾网络舆情风险等级标准样本数据,如表5所示。

表5 震灾网络舆情风险评估随机样本数据及AGABP计算对比

序号	评估指标							风险等级				
	地震震级	次生灾害发生率	伤亡人数	财产损失	应急资源满足能力	监测预警能力	灾民综合满意度	发文量	评论数	转发数	标准值	AGABP计算值
1	3.0	0.12	0	0.12	5	4	5	2 354	4 106	3 298	1	1.002
2	3.0	0.16	1	0.31	5	4	5	6 549	120 001	2 109	1	1.011
3	3.0	0.13	3	0.22	5	4	5	9 102	16 090	4 190	1	1.001
4	3.0	0.17	0	0.15	5	4	4	8 100	3 768	1 789	1	1.004
5	4.0	0.28	13	0.61	3	3	3	40 100	250 019	130 290	2	1.977
6	4.0	0.34	11	0.88	3	3	3	51 089	310 988	160 901	2	2.006
7	4.5	0.41	13	1.28	3	3	3	71 000	248 907	125 890	2	2.010
.
.
.
27	6.8	0.55	67	16.69	2	2	2	118 903	600 897	439 809	3	3.031
28	6.0	0.47	52	13.37	2	2	2	158 902	310 980	389 070	3	2.891
29	7.1	0.81	7 289	1 345.67	1	1	1	789 087	5 189 089	1 790 800	4	3.899
30	8.0	0.97	32 980	7 616.24	1	1	1	1 378 190	3 190 807	2 189 070	4	4.008
31	7.7	0.88	23 109	3 178.99	1	1	1	467 190	6 678 902	2 897 602	4	4.010

对样本数据按式(1)至式(3)进行无量纲处理后输入 AGABP 网络学习训练,取训练次数 10 000 次,AGABP 加速寻优 4 次,计算结果如表 5。进一步,将随机样本数据输入 BP 神经网络、逻辑斯蒂曲线,其训练精度对比如图 3。

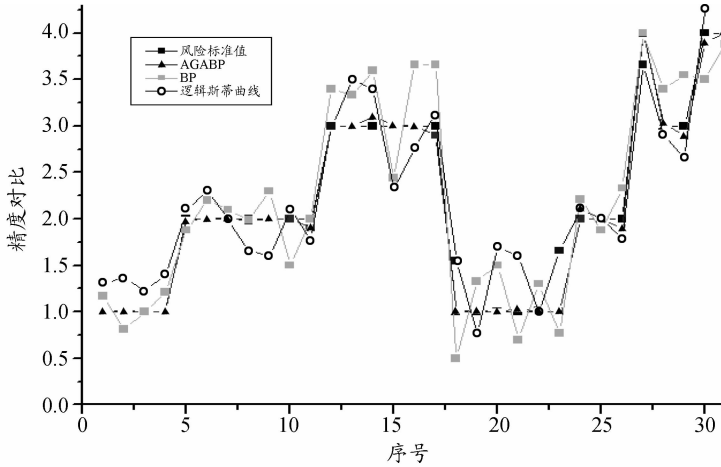


图3 AGABP 与 BP 神经网络、逻辑斯蒂曲线训练精度对比

从图 3 可知,经过 10 000 次训练后,AGABP 网络全局误差为 0.000 712,达到收敛要求;BP 网络的全局误差为 0.001 671;逻辑斯蒂曲线的拟合误差为 0.001 011。与实际风险等级值相比,AGABP 网络的精度明显高于 BP 神经网络和逻辑斯蒂曲线;此时,AGABP 网络的稳定权值和阈值见表 6。

表 6 随机样本数据训练 AGABP 网络 10 000 次后的稳定权值和阈值

参数	隐层神经元 i		
	1	2	3
w_{i1}	-0.312 0	-2.134 2	-1.675 0
w_{2i}	-3.456 1	5.130 9	1.867 4
w_{2i}	-1.375 8	2.987 9	1.232 4
θ_i	0.860 9	1.356 1	-0.204 4
w_{i1}	-2.517 6		
w_{i2}	3.325 6		
w_{i3}	-2.184 3		
θ_j	-12.312 7		

2. 实例评估

经过 10 000 次训练后,AGABP 网络的阈值和各层的权值达到稳定,训练精度符合要求,说明训练后的 AGABP 网络能够用于震灾网络舆情风险评估中。分别运用 AGABP 网络、BP 神经网络和逻辑斯蒂曲线对表 4 中 6 个案例进行评估。设定 AGABP 网络和 BP 神经网络各自训练次数 15 000、加速寻优 4 次,其他参数不变;逻辑斯蒂曲线仍采取拟合方式进行风险评估。三种方法得到结果如表 7、图 4 所示。

表7 三种方法的评估结果

	江西九江	云南普洱	四川汶川	青海玉树	雅安地震	云南鲁甸
AGABP 网络	3.000 2	1.977 0	4.001 2	3.896 4	4.011 0	3.000 0
BP 神经网络	3.000 0	2.331 0	3.097 8	4.000 4	3.997 1	2.996 3
逻辑斯蒂曲线	2.908 8	2.110 2	3.899 0	4.011 2	2.212 0	3.116 2

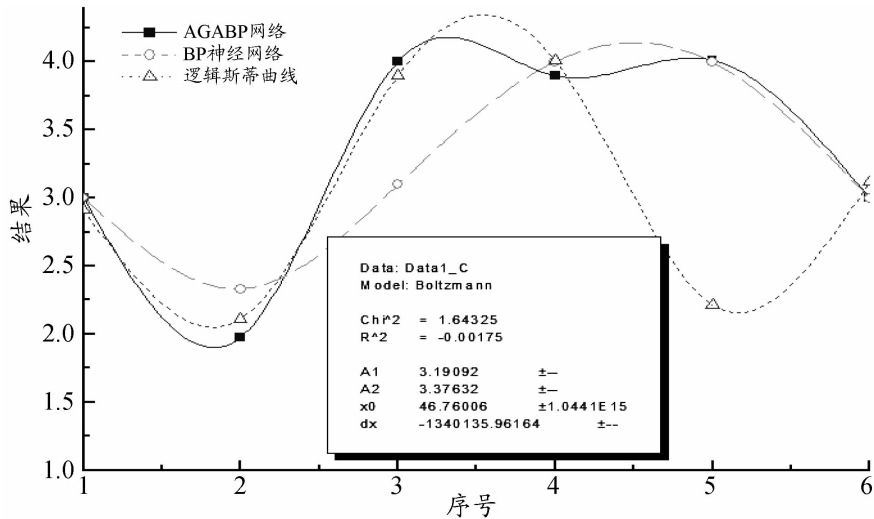


图4 AGABP、BP 和逻辑斯蒂的评估曲线

(三) 结果分析

采用 AGABP 模型对 6 次震灾网络舆情风险评估结果表明:(1)地震震级、伤亡人数、财产损失、发文量、评论数和转发量等 6 个指标是影响震灾网络舆情风险等级的关键性指标。通过观察随机数据训练结果,次生灾害发生率、应急资源满足能力、监测预警能力和灾民综合满意度这 4 个指标对震灾网络舆情风险等级的影响程度低于其余 6 个指标。(2)AGABP 模型评价结果与实际基本一致,可信度高于 BP 神经网络和逻辑斯蒂曲线。表 7 中,BP 神经网络将汶川地震网络舆情风险评估为低风险,这与实际不符,实际上汶川地震后网络舆情迅速扩散,谣言四起,经过政府、专家的正确引导和管控才杜绝众多负面舆情的膨胀。逻辑斯蒂曲线将雅安地震评估为一般风险,这也与实际不符。

五、结论

科学构建震灾网络舆情风险评估体系是制定网络舆情应对预案、提升震灾应急效率和增加灾害应急能力的基础工作。本研究得出如下结论。

第一,提出具有高度代表性特征的震灾网络舆情风险监测指标。依据网络舆情传播规律和震灾网络舆情产生特征,从事件的物理属性和社会属性两方面设置风险监测粗选指标。通过指标信度、效度检验后,采用主成分分析和累计贡献率方法对粗选指标筛选后,提出 2 个维度、4 个二级指标、10 个三级指标所构成的震灾网络舆情风险监测指标体系。模型验证结果表明,本研究所构建的指标体系能够有效对震灾网络舆情进行评估。

第二,训练后的 AGABP 模型,在收敛速度、参数优化和预防过早收敛上明显优越于 BP 神经网络。

络,能够用于震灾网络舆情风险管理实践。

第三,AGABP模型的精度高于BP神经网络和逻辑斯蒂曲线。样本训练和实例验证表明,BP神经网络容易陷入局部最优,精度低于AGABP模型,训练时间也长于AGABP模型;逻辑斯蒂曲线在众多非线性拟合中精度较高,从图3、图4也可以看出,两次拟合结果表明逻辑斯蒂曲线对非线性数据的拟合结果十分稳定,但与AGABP模型相比,逻辑斯蒂曲线的拟合精度仍然较差。

参考文献:

- [1]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.
- [2]柳虹,徐金华.网络舆情热点发现研究[J].科技通报,2011,27(3):421-425.
- [3]方倩,陆伟.微信公众号信息传播热度的影响因素实证研究[J].情报杂志,2016,35(2):157-162.
- [4]焦超,刘功申.网络突发热点事件的热度分布[J].通信技术,2012,4(4):58-60.
- [5]翟晓芳,刘全明,程耀东,等.基于转发层次分析的新浪微博热度评估研究[J].计算机工程,2015,41(7):31-35.
- [6]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.
- [7]张一文,齐佳音,方滨兴,等.非常规突发事件网络舆情热度指标评估体系构建[J].情报杂志,2010,29(11):71-75.
- [8]魏德志,陈福集,林丽娜.一种基于时间序列的热点话题发现模型和算法[J].情报科学,2017,35(10):142-146.
- [9]FU X H, LI J Q, YANG K, et al. Dynamic Online HDP model for discovering evolutionary topics from Chinese social texts[J]. Neuroncomputing, 2016, 171(1):412-424.
- [10]KHAN K, BAHARUDIN B, KHAN A, et al. Mining opinion components from unstructured reviews: A review[J]. Journal of King Saud University-Computer and Information Sciences, 2014, 26(3):258-275.
- [11]HUANG S H, LIU Y, DANG D P. Burst topic discovery and trend tracing based on Storm[J]. Physical A: Statistical Mechanics and its Applications, 2014, 416(15):331-339.
- [12]MA B J, ZHANG N, LIU G N, et al. Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach[J]. Information Processing & Management, 2016, 52(3):430-445.
- [13]王治莹,李勇健.政府干预下突发事件舆情传播规律与控制决策[J].管理科学学报,2017,20(2):43-52.
- [14]HUO L A, HUANG P Q, FANG X. Interplay model for authorities' media and rumor spreading in emergency event[J]. Physical A: Statistical Mechanics and its Applications, 2011, 390(20):3267-3274.
- [15]LEBENSZTAYN E, RODRIGUEZ P M. A connection between a system of random walks and rumor transmission[J]. Physical A: Statistical Mechanics and its Applications, 2013, 392(23):5793-5800.
- [16]ZHAO L J, WANG Q, CHENG J J, et al. The impact of authorities' media and rumor dissemination on the evolution of emergency[J]. Physical A: Statistical Mechanics and its Applications, 2012, 391(15):3978-3987.
- [17]齐佳音,刘慧丽,张一文.突发公共危机事件网络舆情耦合机制研究[J].情报科学,2017,35(9):102-108.
- [18]康伟.突发事件舆情传播的社会网络结构测度与分析——基于“11.16”校车事故的实证研究[J].中国软科学,2012,7(7):169-178.
- [19]张玉亮.基于发生周期的突发事件网络舆情风险评估指标体系[J].情报科学,2012,30(7):1034-1037.
- [20]路枝芳,薛耀文.基于因子聚类分析的舆情风险研究[J].情报科学,2016,34(7):76-80.
- [21]蒋金才,季新菊,刘良,等.河南省1950~1990年水旱灾害分析[J].灾害学,1996(4):69-73.
- [22]王硕,张礼兵,金菊良.系统预测与综合评价方法[M].合肥:合肥工业大学出版社,2006:186-191.
- [23]王思艳.自适应遗传算法的研究[D].保定:华北电力大学,2009.

Risk monitoring indexes and assessment method of network public opinion of earthquake disaster

ZHANG Yu¹, FU Min², LUO Jiarong¹

(1.School of Economics and Management, Southwest University of Science and Technology, Mianyang 621010, P. R. China; 2.Sichuan Xuebao Dairy Group Co., Ltd., Mianyang 621000, P. R. China)

Abstract: To construct the public opinion risk assessment system for earthquake disaster network, the paper puts forward the risk assessment method of the public opinion of Accelerating Genetic Algorithm BP neural network (AGABP) for earthquake disaster after the public opinion monitoring index is built. According to the evolution theory of network public opinion, the paper focuses on the physical and social attributes of the earthquake disaster network public opinion, and puts forward earthquake disaster Internet public opinion risk monitoring indexes of 2 dimensions, 4 indexes of second-level and 10 indexes of third-level. According to the disadvantage of the conventional evaluation system for nonlinear, high dimension and non-normal evaluation problem, this paper makes use of the advantage of BP network, which can approximate any nonlinear continuous function with arbitrary precision. The paper uses BP network for evaluating the risk of earthquake disaster network public opinion, and also uses Accelerating Genetic Algorithm (AGA) improving the shortcoming of BP to solve the problem of slow training and premature convergence in conventional BP networks. The paper uses random sample data to carry out self-learning training for AGABP model and verifies it with actual sample data. The research results show that: AGABP model has obvious advantages in convergence speed and accuracy to compare with BP neural network and logistics curve, and can be applied to risk management practice of earthquake network public opinion.

Key words: earthquake disaster network public opinion; risk assessment; AGABP model; risk monitoring indexes

(责任编辑 傅旭东)