

基于 BP 神经网络的网络舆情预测研究

张军芳¹ 杜鹏² 张利民¹ 刘旭浩¹ 程凤林¹ 刘士琴¹ 孙朝云¹

(1. 衡水学院数学与计算机学院 河北省衡水市 053000 2. 衡水学院电子信息工程学院 河北省衡水市 053000)

摘要：网络舆情是社会舆情的重要组成部分，对网络舆情进行预测研究具有重要的意义。本文分析了衡水市网络舆情现状，构建了基于 BP 神经网络的网络舆情预测模型，对该模型进行了实证分析，结果表明，该模型在进行网络舆情预测时具有较高的准确度。最后，本文提出了网络舆情预测的相关对策建议。

关键词：网络舆情；BP 神经网络；舆情预测

随着互联网的普及和发展，网络已经成为人们日常生活中必不可少的一部分。由于信息的复杂性和多样性，以及传播途径的多样化，人们很难对其进行准确地分析和判断。因此，在互联网环境下如何对网络舆情进行正确有效的引导就显得尤为重要。研究表明，网络舆情作为一种社会现象，其产生发展具有一定的规律性和特点，因此利用有效方法对网络舆情进行准确预测就显得尤为重要。本文旨在通过建立基于 BP 神经网络的网络舆情预测模型对网络舆情进行预测分析，为政府、企业、媒体等单位提供一种有效、快捷的信息分析手段。

1 网络舆情预测研究现状

近年来，随着网络技术的发展，网络舆情作为一种新的信息传播形式也日益受到重视。网络舆情是指网民在互联网上关于社会事件、政府政策、文化现象、政治问题等方面的观点、态度和情绪的总和。网络舆情预测先是通过数据采集、数据清洗和数据分析等流程得到舆情事件的数据集，然后选择合适的预测模型，通过不断地学习、验证、调优等处理后对网络舆情事件的发展趋势作出有效的预测，从而实现了对舆情事件的准确监控 [1]。网络舆情的发生与发展受众多因素的影响，具有复杂性和不确定性，因而也就带来了较难预测的问题。目前，国内外学者对网络舆情的预测研究主要集中在以下几个方面：首先是基于 ARMA 模型的网络舆情预测方法；其次是基于灰色理论的网络舆情预测方法；最后是基于神经网络理论的网络舆情预测方法。而本文将重点研究 BP 神经网络模型在网络舆情预测方面的应用。

2 BP 神经网络概述

BP 神经网络是由 Boltzmann 于 1956 年提出的一种误差反向的多层前馈神经网络，可以完成非线性映射[2]。其主要思想是：对于输入 n 个学习训练样本“ $\gamma_1, \gamma_2, \dots, \gamma_n$ ”，与其相对应的输出 m 个样本为“ p_1, p_2, \dots, p_m ”。用网络的实际输出（ t_1, t_2, \dots, t_m ）与目标矢量（ q_1, q_2, \dots, q_m ）之间的误差来调整其权值，使 $t_l (l = 1, 2, \dots, m)$ 与期望的 q_l 尽可能地趋近，使输出层的误差平方和达到最小。BP 神经网络的模型结构包括输入层、隐含层和输出层。BP 网络能够存储和学习输入-输出类型的非线性映射关系，无需事前揭示这种关系且对任意非线性函数的完成趋近。BP 神经网络在进行网络训练时，需要输入一定数量的样本数据，然后通过对训练样本数据进行反复迭代计算来完成学习过程，其网络拓扑结构如图 1 所示。

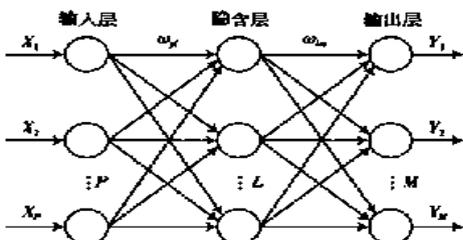


图 1 BP 神经网络拓扑

3 网络舆情预测模型设计

在进行网络舆情预测时，本文首先将影响网络舆情的年龄、性别、职业、受教育程度等因素作为输入，其次运用已知数据对这些输入变量进行训练，得到一个输出结果，最后利用训练好的 BP 神经网络对网民关注的热点问题进行分析。其中 BP 神经网络采用 3 层拓扑结构，输入层节点数设置为 5，输出层节点数设置为 5；设置最大迭代次数 5000，训练的目标误差为 $1e-5$ ，学习速率为 0.05，动量因子为 0.9；隐含层传递函数设置为“losig”，输出层传递函数设置为“losig”，训练算法设置为 traingdx 函数，隐含层节点数由公式（1）估计，再用估计值附近数值进行试错，选择使均方误差较小的值作为隐含层节点数[3]。

$$hiddennum = \sqrt{in\ dim + out\ dim} + a, \quad a = \{1, 2, \dots, 10\} \text{ 整数} \quad (1)$$

4 实证分析

论文数据来自 基于大数据和人工智能的网络舆情挖掘及预测研究课题组对衡水市居民调查所得数据，最终样本容量为 678 份。

运用神经网络模型对网络舆情进行预测，表 1 为案例处理汇总，其中共有效样本 678 份，其中 494 名被调查者被分配到训练样本，占比 72.86%，另外 184 名被调查者被分配到测试样本，占比 27.14%。

表 1 为案例处理汇总

		N	百分比
样本	训练	494	72.86%
	测试	184	27.14%
	有效	678	100%
	已排除	0	
	总计	678	

表 2 模型汇总

训练	交叉熵错误	267.871
	百分比错误预测	19.6%
	中止使用的规则	错误未减少的 1 连续步骤 a
	培训时间	0:00:00.29
测试	交叉熵错误	102.783
	百分比错误预测	20.2%
	因变量: B5	

由表 2 模型汇总表可知，训练集的错误比率为 19.6%，测试集的错误比率为 20.2%；从分类表中可以看出，训练集的正确率为 80.4%，测试集的正确率为 79.8%。因此神经网络模型对于样本的预测效果良好。

表 3 模型摘要表

样本	已观测	已预测					正确百分比
		教育领域	社会保障	医疗卫生	健康养生	社会事件	
	教育领域	450	11	11	7	15	91.1

训练	社会保障	17	420	8	20	29	85.0
	医疗卫生	12	8	443	18	13	89.7
	健康养生	10	11	18	435	20	88.1
	社会事件	11	12	10	20	441	89.3
	总计百分比	90	90.9	90.4	87	85.1	88.6
测试	教育领域	160	6	7	5	6	87.0
	社会保障	5	163	6	7	6	88.6
	医疗卫生	5	5	165	4	5	89.7
	健康养生	7	6	5	160	6	87.0
	社会事件	6	7	7	6	158	85.9
	总计百分比	87.4	87.2	86.8	87.9	87.3	87.6

由表3模型摘要表可知,首次构建的神经网络模型训练集预测正确率为88.6%,测试集预测的正确率为87.6%,进一步说明所构建的神经网络模型对衡水市网民关注的热点问题预测效果良好。

5 结论

本文以年龄、性别、职业、受教育程度等因素作为输入,网民关注的热点问题教育领域、社会保障、医疗卫生、健康养生和社会事件作为输出,建立BP神经网络模型对衡水市网民关注的热点问题实证分析,结果表明,运用BP神经网络模型进行网络舆情预测具有较高的准确度。

大数据及人工智能技术,为对海量信息和数据的处理和在此基础上

的舆情研判提供了方便条件[4]。但是进行深入、准确的舆情研判(此处的舆情研判,其实也包含了对特定舆论的状况、特征、因果、走势等的研判),仍然并非易事,正确的舆情研判是网络舆论舆情引导的重要抓手[5]。

参考文献:

- [1]张和平,陈齐海.基于灰色马尔可夫模型的网络舆情预测研究[J].情报科学,2018,36(1):75-79.
- [2]薛圆圆.基于BP神经网络的网络舆情危机预警研究[J].图书情报工作,2011(20):59-63.
- [3]游丹丹,陈福集.基于改进粒子群和BP神经网络的网络舆情预测研究[J].情报志,2016.8(35).
- [4]岳梦怡,王保华.原则与路径:融媒体视域下的舆情研判研究[J].新闻世界,2023,(1).
- [5]丁柏铨.网络舆论舆情引导刍议——关于引导策略的研究[J].西北师大学报(社会科学版),2023,60(03):86-95. DOI:10.16783/j.cnki.nwnus.2023.03.010.

基金项目:河北省教育厅资助青年基金项目自然科学类(QN2020529):基于大数据和人工智能的网络舆情挖掘及预测研究。