

# 基于因子分析和 Elman 网络的舆情关键词热度预测<sup>①</sup>



肖光华<sup>1,2</sup>, 王清莲<sup>3</sup>

<sup>1</sup>(江苏城乡建设职业学院 设备工程学院, 常州 213147)

<sup>2</sup>(河海大学 计算机与信息工程学院, 南京 210098)

<sup>3</sup>(常州开放大学 终身教育研究中心, 常州 213001)

通讯作者: 肖光华, E-mail: smoothxiao@163.com

**摘要:** 从宏观角度研究基于关键词的网络舆情热度有助于相关机构把握目标群体的整体舆情动态, 从而实现精准施策, 提升舆论引导水平. 本文以新浪微博数据为例, 采用因子分析方法 (Factor Analysis, FA), 挖掘舆情热度内在影响因素, 并通过改进 Elman 网络结构, 利用遗传算法 (Genetic Algorithm, GA) 优化初始参数来构建模型对网络舆情关键词热度进行分析预测. 实验结果表明, 所提出的方法相较于采用原始数据集和标准 Elman 网络的预测结果, 具有更高的预测精度, 可为相关研究提供决策支持.

**关键词:** 因子分析; Elman; 遗传算法; 网络舆情; 关键词热度

引用格式: 肖光华, 王清莲. 基于因子分析和 Elman 网络的舆情关键词热度预测. 计算机系统应用, 2021, 30(3): 243-249. <http://www.c-s-a.org.cn/1003-3254/7811.html>

## Prediction on Keywords Popularity of Public Opinion Based on Factor Analysis and Elman Network

XIAO Guang-Hua<sup>1,2</sup>, WANG Qing-Lian<sup>3</sup>

<sup>1</sup>(Department of Equipment Engineering, Jiangsu Urban and Rural Construction College, Changzhou 213147, China)

<sup>2</sup>(College of Computer and Information Engineering, Hohai University, Nanjing 210098, China)

<sup>3</sup>(Lifelong Education Research Center, Changzhou Open University, Changzhou 213001, China)

**Abstract:** It is helpful for institutions to master the whole trends of target group that research on keywords popularity of network public opinions from a macroscopic perspective, precisely formulating corresponding strategies to enhance the level of opinion guidance. With Sina Weibo data set as an example, Factor Analysis (FA) is used to mine the internal factors of public opinions; a model that analyzes and predicts the keyword popularity of network public opinions is created through the initial parameters optimized by Genetic Algorithm (GA) and Elman network structure. The results show that predictions made by our method is more precise than those of original data sets and standard Elman network. Thus, it can be applied to providing reference for decision making.

**Key words:** factor analysis; Elman; genetic algorithm; network public opinion; keyword popularity

网络通信技术的进步促使社交网络成为信息传播的重要渠道. 人们通过智能终端发表观点, 表达意见, 与之相伴产生的网络舆情问题已经成为相关机构关注

的热点. 以新浪微博为例, 截至 2019 年底, 微博月活跃用户达到 5.16 亿, 其海量舆情数据已经达到了让任何部门、机构都无法忽视的地步. 当前基于舆情热度的

① 基金项目: 全国教育信息技术研究重点课题 (183220001)

Foundation item: Key Project for Educational Informatic Technology Research of China (183220001)

收稿时间: 2020-07-09; 修改时间: 2020-08-11; 采用时间: 2020-08-17; csa 在线出版时间: 2021-03-03

研究多数是针对一些具体的热门事件或热门话题,无法从宏观的角度把握目标群体的整体舆情动态,以广义的舆情“关键词”作为热度的研究对象,如“手机”、“农产品”、“房价”、“比特币”、“疫情”等,对企业产品精准营销,政府部门舆情监测将更具有宏观指导意义。目前对公众开放的基于关键词的大数据分析平台主要有百度指数、微指数、头条指数等,这类平台虽然能够提供一些关键词的分析数据,但由于侧重点不同,并不能满足对舆情热度需求的个性化分析。为此,本文以微博数据为基础,提出了一种结合因子分析、遗传算法的 Elman 网络模型对舆情关键词热度进行分析预测,以期对相关决策提供支持。

## 1 研究现状

通过相关研究文献的梳理,网络舆情热度的研究方法主要有两类:

一类是根据事物发展的规律,基于时间序列挖掘自身历史数据进行热度的分析和预测。王宁等<sup>[1]</sup>采用灰色系统理论拟合时间序列指数对诉讼案件热度进行了预测和分级管理问题。陈福集等<sup>[2]</sup>采用改进的 ABC-BP 模型对“天津港爆炸事件”的百度指数热度走势进行了预测研究。刘巧玲等<sup>[3]</sup>通过一种改进 SIR 传染病传播模型,以“魏则西”事件发生前三天的数据来预测事件后续传播趋势。高颖<sup>[4]</sup>采用 LS-SVM 算法对 2018 年哈尔滨市发生的“8.25”火灾事件进行回归建模从而预测其短期热度值。这些针对特定事件的舆情热度预测方法均能取得较好的预测效果,但对于舆情关键词这种宽泛的指标,并不符合特定事件舆情热度从产生、爆发到衰退的规律,时间序列越长,信息越分散,其波动更多呈“随机”状态。虽然林育曼等<sup>[5]</sup>构造了 ARIMA-BP 神经网络模型将预测对象视为随机序列来近似逼近,取得较好的拟合效果。但对于突发值的预测往往不是很准确。

另一类是通过挖掘自身数据特征,找到与舆情热度的关联关系,从而进行热度的分析和预测。郑志蕴等<sup>[6]</sup>提出一种基于多特征的热门微博预测算法,利用信息增益算法对微博特征热度进行度量,结合 BP 神经网络来预测是否为热门微博。谢修娟等<sup>[7]</sup>提出一种基于密度的 K-means 算法从微博内容中抓取热词,分析舆情热点。兰月新等<sup>[8]</sup>基于灰色关联度方法构建网络舆情热度模型,定性分析了网络舆情热度影响因素。这些方法

通过自定义热度计算公式,采用分类和聚类算法预测的舆情热度的发展趋势,都具备较高的准确率和稳定性,但由于没有客观指标的量化分析,对于舆情关键词的热度研究存在一定的局限性。同时各特征对热度指标的究竟有多大的贡献还需进一步研究分析。

针对以上研究存在的局限,本文以新浪微博为例,设计了一种新的组合模型可有效提高网络舆情关键词热度的预测精度。首先选取并量化影响舆情关键词热度的特征指标,采用因子分析方法提取其中贡献较大的影响因子,剔除影响预测精度的噪声冗余,然后将影响因子输入到 Elman 网络中计算预测结果。为进一步提高预测精度,一方面通过增加神经元个数来反馈调节模型预测结果,另一方面利用遗传算法优化网络权值和初始参数,最终实现舆情热度及其发展趋势的预测。

## 2 模型构建

### 2.1 特征量化

以新浪微博公开可挖掘数据为例,与舆情热度有关的特征一般包括博主特征、博文特征和传播特征。博主特征包括博主 ID(或昵称)、博主类型(即认证身份)、关注数、粉丝数和微博数等<sup>[9]</sup>,博主特征决定了它的受众范围,即热度传播的广度。博文特征包括博文的长度、博文元素(图片、视频、链接、表情符号)、博文内容(所表达的情感)以及发布时间,一般博文长度和舆情热度呈正相关态势,博文所含元素越丰富越能吸引读者的关注,内容情感的表达可以引起读者的共鸣,可以说博文特征决定了舆情热度的传播强度。传播特征包括微博的转发数、评论数、点赞数以及发布平台(PC端、手机端等)<sup>[10]</sup>,传播特征是舆情热度最直接的反映,决定的舆情热度最终传播的深度。为了能更好的分析热度影响因素,需要将特征信息中非量化信息进行量化,如表 1 所示,根据这些信息对热度的影响关系,从高到低依次赋分。

博文长度可根据字符统计个数直接赋分,博文内容的情感强度计算依据的是博文内容中所表达的正面和负面情感词汇统计,采用中国科学院计算机所软件室编写的中文分词工具 ICTCLAS<sup>[11]</sup>,对所有微博内容进行情感词汇统计。一般情感强度定义公式可由式(1)表示:

$$E = \sum \omega_i e_{\text{pos}} + \sum \omega_j e_{\text{neg}} \quad (1)$$

其中,  $E$  为情感强度得分,  $e_{\text{pos}}$  为正面得分,  $e_{\text{neg}}$  为负面得分,  $\omega$  为权重系数, 可以按照副词的强弱程度分类<sup>[12]</sup>.

表1 非量化信息赋值表

特征信息	量化值	赋分
博主类型	微博官方认证	5
	微博会员	4
	微博个人认证	3
	其它认证	2
	无认证	1
博文长度	根据字符个数赋值(含元素信息)	
博文内容	式(1)赋分	
发布时间	16:00-24:00	3
	8:00-16:00	2
	0:00-8:00	1
发布平台	PC端	3
	手机端	2
	其它(超话、第三方平台推送等)	1

## 2.2 因子分析

由于各特征信息所携带的信息量对舆情热度预测的贡献率以及关联度差异较大, 从而会影响最终的预测结果, 同时信息中的噪声和冗余也会影响预测结果的精度. 因子分析采用潜在的假想变量和随机影响变量的线性组合表示原始变量, 通过提取其中具有较高贡献率的因子来进行预测分析, 既能反映舆情热度的真实情况, 又可以过滤影响结果的冗余信息, 从而提高预测精度<sup>[13]</sup>.

首先根据所采集到的特征信息, 可视为随机特征向量  $x = (x_1, x_2, \dots, x_n)^T$ , 其均值为  $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$ , 协方差阵为  $S = \sigma_{ij}$ , 因子分析的模型如下表示:

$$\begin{cases} x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_n = \mu_n + a_{n1}f_1 + a_{n2}f_2 + \dots + a_{nm}f_m + \varepsilon_n \end{cases} \quad (2)$$

其中,  $f_1, f_2, \dots, f_n$  为公共因子,  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  为特殊因子, 彼此都不相关, 每个因子只对当前原始变量起作用, 用矩阵形式表示为:

$$x = \mu + Af + \varepsilon \quad (3)$$

其中,  $A$  为因子荷载矩阵. 然后采用主成分法进行参数估计, 通过求解其特征根  $\lambda_i (i = 1, 2, \dots, n)$  和相应的正交单位特征向量为  $\hat{t}_i (i = 1, 2, \dots, n)$ , 选取保留较多原始信息的因子数  $m$ , 计算其贡献率  $G$ .

$$G = \sum_{i=1}^m \hat{\lambda} / \sum_{i=1}^n \hat{\lambda} \quad (4)$$

当贡献率  $G$  达到一个较高的百分比, 协方差阵为  $S$  可作如下的近似分解:

$$\begin{aligned} S &= (\hat{t}_1 \dots \hat{t}_p) \begin{pmatrix} \hat{\lambda}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_n \end{pmatrix} \begin{pmatrix} \hat{t}_1^T \\ \vdots \\ \hat{t}_n^T \end{pmatrix} \\ &\approx \hat{\lambda}_1 \hat{t}_1 \hat{t}_1^T + \dots + \hat{\lambda}_m \hat{t}_m \hat{t}_m^T + \hat{D} \\ &= \hat{A} \hat{A}^T + \hat{D} \end{aligned} \quad (5)$$

其中,

$$\begin{cases} \hat{A} = (\sqrt{\hat{\lambda}_1} \hat{t}_1, \dots, \sqrt{\hat{\lambda}_m} \hat{t}_m) = (\hat{a}_{ij}) \\ \hat{D} = \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2) \\ \hat{\sigma}_i^2 = s_{ii} - \sum_{j=1}^m \hat{a}_{ij}^2, i = 1, \dots, n \end{cases} \quad (6)$$

式中,  $\hat{A}$  和  $\hat{D}$  就是因子模型的一个解.

为了能更清楚的了解各公共因子的物理解释意义, 采用最大方差法对因子荷载阵进行旋转, 使因子荷载阵的结构简化, 荷载矩阵中每列或行的元素平方趋向 0 和 1 两极分化. 令  $\Gamma$  为正交阵, 做正交变换求解:

$$B = (b_{ij})_{n \times m} = \left( \sum_{l=1}^m a_{il} \gamma_{lj} \right) \quad (7)$$

$A$  的行元素平方和为  $h_i^2 = \sum_{j=1}^m a_{ij}^2$ , 令  $d_{ij} = \frac{b_{ij}}{h_i}, i = 1, 2, \dots, n; j = 1, 2$ .

定义因子荷载矩阵  $A$  的方差为:

$$V(\theta) = \sum_{j=1}^m \frac{\sum_{i=1}^n (d_{ij}^2 - \bar{d}_j)^2}{p} \quad (8)$$

最后, 采用回归的方法将式 (3) 转化为:

$$F_j = \beta_{j1}X_1 + \dots + \beta_{jn}X_n, j = 1, \dots, m \quad (9)$$

其中,  $\beta$  为因子的得分系数,  $F$  为各因子的最终得分. 利用提取的公因子进行预测分析不仅达到了降维的目的, 而且每个因子都具备物理解释意义, 更有利于舆情热度影响因素的分析.

## 2.3 Elman 网络

### 2.3.1 Elman 网络结构设计

Elman 模型结构中不仅包含输入层、中间层 (即隐含层) 和输出层, 还包含记载中间层历史信息的承接层<sup>[14]</sup>, 相较于一般静态神经网络具有动态特性好, 逼近速度快等特点. 当然, Elman 网络自身缺少对输入输出层的反馈而忽略了其对最终结果的影响<sup>[15]</sup>, 可以通过增加神经元个数, 利用所有层的历史信息来反馈调节模型预测结果, 其网络结构如图 1 所示.

标准 Elman 网络非线性状态空间表达式为<sup>[16]</sup>:

$$x(k) = f(\omega_1 x_c(k) + \omega_2 u(k-1)) \quad (10)$$

$$y(k) = g(\omega_3 x(k)) \quad (11)$$

$$x_c(k) = x(k-1) \quad (12)$$

其中,  $\omega$ 为各层连接权值矩阵,  $f(*)$ 为中间层传递函数,  $g(*)$ 为输出层传递函数. 增加神经元后, 其改进结构状态空间表达式为:

$$x(k) = f(\omega_2 u(k-1) + \omega_3 x(k) + \omega_4 u_c(k-1) + \omega_5 y_c(k)) \quad (13)$$

$$y(k) = g(\omega_1 x(k) + \omega_6 x_c(k)) \quad (14)$$

$$x_c(k) = x(k-1) + \alpha x_c(k-1) \quad (15)$$

$$u_c(k) = u(k-1) + \beta u_c(k-1) \quad (16)$$

$$y_c(k) = m y(k-1) + \gamma y_c(k-1) \quad (17)$$

其中,  $\alpha$ 、 $\beta$ 、 $\gamma$ 为反馈增益因子.

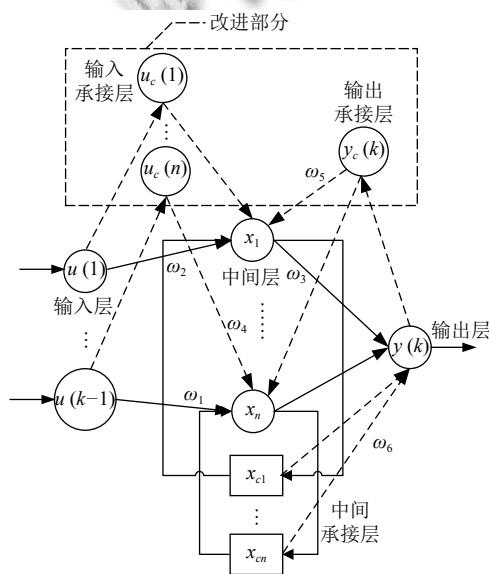


图1 Elman 网络结构图

### 2.3.2 GA\_Elman 算法设计

虽然通过改进 Elman 网络结构可以充分利用中间层信息对预测模型进行反馈调节, 但也由此增加了网络参数数量, 其预测结果更容易受到网络权值和初始参数的影响, 而且 Elman 网络采用的梯度下降算法容易陷入局部最优解<sup>[17]</sup>. 借助遗传算法理论上可以扩大到整个解空间范围来寻找最优解. 由于遗传算法中每个种群个体都包含了 Elman 网络的所有权值和参数, 个体通过适应度函数计算适应度值, 通过选择合适的

种群规模并利用迭代的方式进行选择、交叉和变异等运算就可以找到其最优适应度对应的个体, 即最优权值参数. GA\_Elman 算法的主要训练流程如图 2 所示.

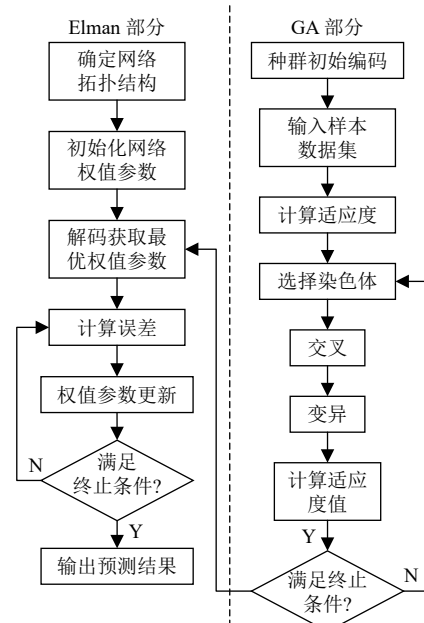


图2 GA\_Elman 算法流程

根据 2.3.1 节中式 (13)~式 (17), 可知遗传算法需要对 6 个网络权值矩阵和 3 个反馈增益因子进行训练计算, 将预测输出和期望输出误差绝对值的倒数作为适应度函数, 计算公式为:

$$fit = \frac{1}{k \left( \sum_{i=1}^n \text{abs}(E(y_i) - y_i) \right)} \quad (18)$$

其中,  $n$  为网络输出节点数,  $y_i$  为第  $i$  个节点预测输入,  $E(y_i)$  为第  $i$  个节点期望输出,  $k$  为样本系数. 式 (19) 中选择误差的绝对值而没有采用误差平方主要是为了筛除种群中适应度受到输入数据中离群点和异常值影响的个体. 此外, 染色体选择方法采用轮盘赌法, 即基于适应度比例的选择策略<sup>[18]</sup>.

## 3 实验分析

### 3.1 实验数据处理

本实验由课题小组基于 Python 语言开发的爬虫工具来获取新浪公开发布的微博数据, 以 2020 年热度始终在持续的“疫情”为关键词, 爬取从 2020 年 1 月 21 日至 2020 年 4 月 27 日的每天热门微博作为实验数据, 共采集到 67238 条数据. 按照 2.1 节特征量化的方

式对数据进行预处理,以日期为标签可汇总成 98 组数据,每组数据可提取 13 个可量化特征。

由于新浪微博仅对博主自身开放微博“浏览量”这一热度数据,无法从公开数据中采集到,因此本文分别选择和舆情热度密切相关的全网“博文数”和博主的博文“点赞数”作为目标特征,构造 2 个实验来对结果分析对比,以验证方法的客观有效性,构造方式如表 2 所示。

表 2 实验构造方式

标签	训练特征	目标特征
实验一	按天汇总,由博主的个数、认证级别、关注数、粉丝数、微博数,博文的博文长度、博文内容、发布时间以及传播平台、转发数、评论数、点赞数组成	博文数
实验二	抽取10个博主的数据分别组成实验数据集,由博文的博文长度、博文内容、发布时间以及传播平台、转发数、评论数组成	点赞数

实验采用均方误差 (MSE) 和判定系数 ( $R^2$ ) 作为评价标准,其公式如下所示:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2 \quad (19)$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (20)$$

其中,  $y_i$  为的目标值,  $p_i$  的预测值,  $SST$  为总平方和,  $SSR$  为回归平方和,  $SSE$  为残差平方和。

### 3.2 博文数预测结果分析

根据 2.2 节式 (2)–式 (4) 提取各公共因子并计算累积率,如表 3 所示。

表 3 因子成分表

因子成分	方差总计	方差百分比	累积率(%)
1	5.254	43.781	43.781
2	2.702	22.514	66.295
3	1.389	11.577	77.873
4	0.883	7.354	85.227
5	0.757	6.312	91.539
6	0.663	5.523	97.062
7	0.183	1.527	98.589
8	0.068	0.566	99.155
9	0.050	0.419	99.574
10	0.031	0.259	99.833
11	0.019	0.156	99.989
12	0.001	0.011	100.000

表 3 中显示了各主成解释原始变量总方差的情况,为了尽可能保留原始特征信息,同时又达到降维

去噪的目的,经过比较实验,选择保留前 6 个因子的实验效果最佳,因篇幅有限,分析比较的实验过程不再赘述。由式 (5)–式 (9) 得到旋转后的因子成分表,如表 4 所示。

从表 4 中可比较清晰的看出各成分的物理解释意义,成分 1 可以认为是博主自身的影响力因子,成分 2 是平台影响因子,成分 3 是微博自身热度影响因子,成分 4 为受众面影响因子,成分 5 为博文内容影响力因子,成分 6 为微博传播持久力影响因子。

表 4 旋转后的成分矩阵

特征信息	成分					
	1	2	3	4	5	6
博文长度	0.121	<b>0.951</b>	-0.008	0.018	0.203	0.002
微博内容	0.183	0.257	0.009	0.004	<b>0.948</b>	0.025
发布时间	0.206	<b>0.932</b>	0.044	0.009	0.082	0.040
发布平台	0.048	<b>0.983</b>	-0.008	-0.010	0.055	0.007
转发数	0.276	0.025	0.211	-0.008	0.025	<b>0.937</b>
点赞数	0.283	-0.042	<b>0.938</b>	-0.002	0.001	0.107
评论数	0.212	0.055	<b>0.952</b>	0.054	0.016	0.132
博主数	<b>0.907</b>	0.247	0.191	0.093	0.125	0.153
博主认证	<b>0.924</b>	0.199	0.191	0.090	0.118	0.154
关注数	<b>0.937</b>	0.110	0.197	0.175	0.085	0.098
粉丝数	0.247	0.002	0.038	<b>0.968</b>	0.004	-0.007
微博数	<b>0.942</b>	-0.017	0.155	0.086	0.023	0.076

为了更好的检验预测效果,将清洗处理后的数据随机分成两组,前 83 组为训练集,后 15 组为测试集,然后分别对原始特征数据采用 Elman 算法预测 (记为 Elman),再对提取的公因子采用 Elman 算法 (记为 FA\_Elman) 和改进后的 Elman 算法 (记为 FA\_GA\_Elman) 进行预测结果比对,实验中遗传算法的种群规模为 100,交叉概率选择为 0.3,变异概率选择为 0.2,迭代次数为 500。其结果对比如图 3 所示,评价指标对比如表 5 所示。

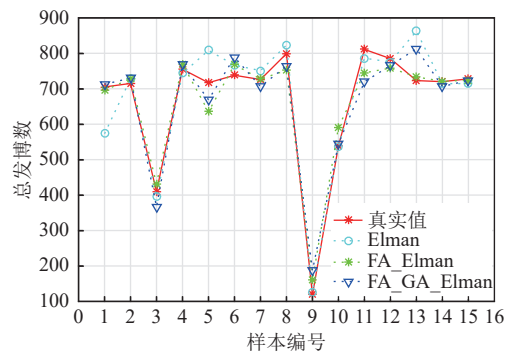


图 3 发博数预测结果与真实值的比较

表5 发博数预测评价指标比较

方法	MSE	R <sup>2</sup>
Elman	0.0190	0.9100
FA_Elman	0.0119	0.9353
FA_GA_Elman	0.0079	0.9628

从预测结果及评价指标中可以看出,采用公因子作为训练特征其预测精度要优于直接以原始数据作为训练特征,而改进后的 Elman 算法预测精度要优于标准 Elman 算法,而且经过遗传算法优化后的 Elman 网络在预测结果上要比标准 Elman 网络稳定许多。

### 3.3 点赞数预测结果分析

为更好检验本方法对于点赞数预测的有效性,在数据集中综合选择了拥有较多粉丝数量且周期内发博数量稳定的前十名热门博主,对其每天点赞数量进行预测,将 FA\_Elman 和 FA\_GA\_Elman 预测结果进行对比.由于各博主不一定每天发布相关微博,统一取 15 组数据作为测试集,其余数据作为训练集,其预测结果如图 4 所示。

从图 4 中可以看出,FA\_Elman 和 FA\_GA\_Elman 的预测结果与真实值比较均具有较好的拟合度,而且对于突发值也能很好地反映其真实趋势.表 6 给出了评价指标的对比结果以及各因子贡献率,从中可以看出 FA\_GA\_Elman 预测精度要优于 FA\_Elman,同时也表明通过因子分析,可以从数据集中提取出影响舆情热度的内在因子,消除噪声干扰,从而达到良好的预测效果.以人民日报点赞数预测为例,经过训练后,其网络参数如表 7 所示,优化后网络参数权值范数要小于优化前的网络参数权值,说明优化后的网络具有更好的泛化性能。

## 4 结论

新时期,网络舆情具有信息规模大,传播速度快,参与范围广,实时交互强等特点,把握舆情热度无论对企业的经营策略,还是社会的公共安全都极为重要.本文结合了因子分析和 GA\_Elman 网络的特点,提出了一种能有效预测舆情关键词热度的模型,首先采用因子分析方法揭示数据背后影响舆情热度的真正内因,然后通过改进 Elman 网络结构并利用遗传算法优化网络权值参数,训练输出特征目标预测值.根据不同模型的实验结果对比,本文提出的方法能有效的预测基于

关键词的舆情热度,预测精度更高,具备一定的应用价值.未来将进一步扩充数据集,以期对相关研究提供个性化的预测分析方案。

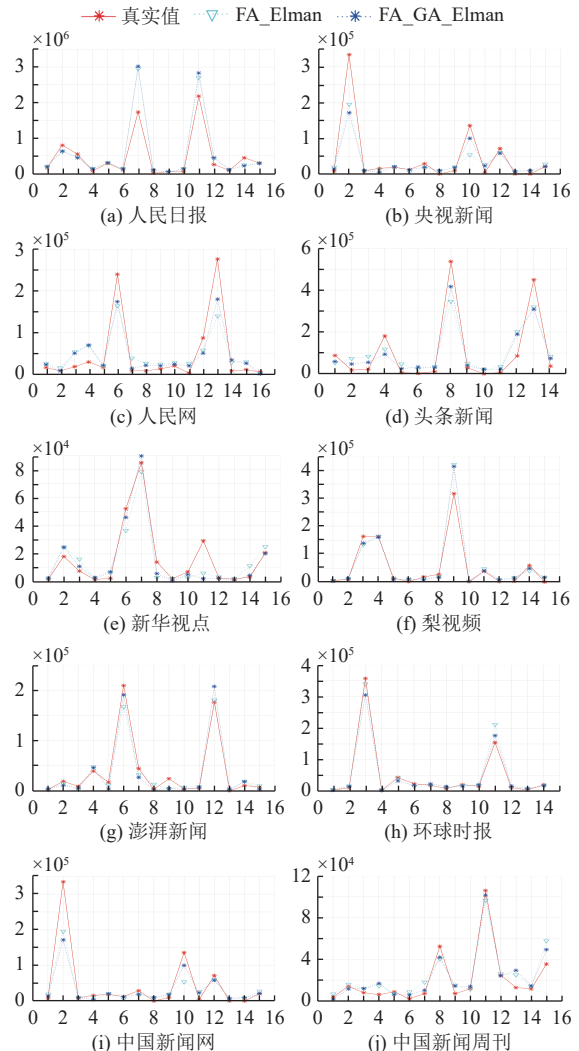


图4 点赞数预测结果与真实值的比较

表6 点赞数预测评价指标

博主ID	FA_Elman		FA_GA_Elman		因子个数	累计贡献率(%)
	MSE	R <sup>2</sup>	MSE	R <sup>2</sup>		
人民日报	0.0550	0.9295	0.0466	0.9300	3	93.02
央视新闻	0.0115	0.9097	0.0034	0.9217	4	98.91
人民网	0.0934	0.8852	0.0580	0.9288	3	92.28
头条新闻	0.0011	0.8790	0.0008	0.9122	3	97.14
新华视点	0.0430	0.8595	0.0353	0.8885	4	96.90
梨视频	0.0090	0.9537	0.0078	0.9604	3	96.86
澎湃新闻	0.0582	0.9635	0.0504	0.9647	3	93.67
环球时报	0.0116	0.9711	0.0111	0.9816	3	90.35
中国新闻网	0.0202	0.9512	0.0211	0.9646	3	95.74
中国新闻周刊	0.0024	0.9065	0.0016	0.9327	3	92.85

表7 网络参数优化前后对比

网络参数名称	优化前网络参数							优化后网络参数						
输入层权值	1.2683	0.9426	0.9598	-1.2203	-0.4741	-1.1584	-0.7339	-1.2926	0.3016	-1.4794	0.5467	0.9831	2.171	1.1053
	-0.3565	0.7758	0.6181	-0.9995	-0.1789	-1.3267	-2.4307	-1.239	-0.6646	-1.2517	0.0982	-1.1508	0.9095	-1.3338
	0.8929	-1.3012	-1.0072	-0.0198	-0.7934	1.4274	2.7018	-1.0838	-0.799	-0.3091	0.9111	-0.7468	0.6368	1.4451
	-0.5716	0.6311	-1.4451	-0.9257	0.0779	-0.2931	0.1855	0.5355	-0.7292	0.8758	-0.643	-1.0576	-0.6639	-0.1707
	1.0175	1.2444	1.2938	0.7917	1.2293	0.562	0.1624	-0.3139	-1.0515	-1.8923	1.4101	-0.0436	-0.4213	-1.0867
	-1.1207	0.4798	-0.7826	0.6243	-0.9874	-0.647	-1.188	0.4459	0.6051	-0.2508	1.6591	0.2644	0.2297	-0.7275
输入层阈值	[-0.9518 1.6933 -1.5415 0.4327 -2.2181 0.7242] <sup>T</sup>							[0.697 3.1555 0.8751 0.0057 1.3691 -0.598] <sup>T</sup>						
输出层权值	[-0.0033 1.6572 1.4671 -0.9467 -0.1558 0.2997]							[1.293 0.7081 -0.5115 1.5902 -0.5371 0.7768]						
输出层阈值	-0.6758							-0.3869						

## 参考文献

- 王宁, 赵胜洋, 单晓红. 基于灰色系统理论的网络舆情预测与分级方法研究. 情报理论与实践, 2019, 42(2): 120-126.
- 陈福集, 肖鸿雁. 基于改进 ABC-BP 模型的网络舆情热度预测研究. 图书馆学研究, 2018, (9): 84-89.
- 刘巧玲, 李劲, 肖人彬. 基于参数反演的网络舆情传播趋势预测——以新浪微博为例. 计算机应用, 2017, 37(5): 1419-1423. [doi: 10.11772/j.issn.1001-9081.2017.05.1419]
- 高颖. 基于改进混沌理论的网络舆情短期预测策略方法研究. 重庆理工大学学报(自然科学), 2019, 33(6): 171-176.
- 林育曼, 文海宁, 饶浩. 基于 ARIMA-BP 神经网络模型的微信舆情热度预测. 统计与决策, 2019, 35(14): 71-74.
- 郑志蕴, 江国林, 张行进, 等. 基于多特征的热门微博预测算法研究. 小型微型计算机系统, 2017, 38(3): 494-498.
- 谢修娟, 李香菊, 莫凌飞. 基于改进 K-means 算法的微博舆情分析研究. 计算机工程与科学, 2018, 40(1): 155-158. [doi: 10.3969/j.issn.1007-130X.2018.01.023]
- 兰月新, 刘冰月, 张鹏, 等. 面向大数据的网络舆情热度动态预测模型研究. 情报杂志, 2017, 36(6): 105-110, 147. [doi: 10.3969/j.issn.1002-1965.2017.06.019]
- 韩晴晴, 张艳梅, 牛娃. 基于半监督协同训练算法的微博水军识别. 计算机科学, 2019, 46(11): 202-208. [doi: 10.11896/j.sjcx.180901617]
- 时聪, 唐九阳, 胡艳丽. 融合用户行为与博文内容的微博用户影响力预测. 小型微型计算机系统, 2017, 38(7): 1495-1500. [doi: 10.3969/j.issn.1000-1220.2017.07.014]
- 童玉珍, 王应明. 基于犹豫模糊集的网络舆情突发事件应急群决策方法. 计算机系统应用, 2019, 28(9): 9-17. [doi: 10.15888/j.cnki.csa.007068]
- 黑富郁, 王景中, 赵林浩. 基于 CNN 和 LSTM 的异构数据舆情分类方法. 计算机系统应用, 2019, 28(6): 141-147. [doi: 10.15888/j.cnki.csa.006900]
- Zhang P, Hu Y, Pang YY, et al. Influence factors analysis of the interfacial bond behavior between GFRP plates, concrete. Structures, 2020, 26: 79-91. [doi: 10.1016/j.istruc.2020.04.005]
- Zhao XL, Shen S, Su LC, et al. Elman neural network-based identification of rate-dependent hysteresis in piezoelectric actuators. Journal of Intelligent Material Systems and Structures, 2020, 31(7): 980-989. [doi: 10.1177/1045389X20905987]
- Jia WK, Zhao DA, Ding L, et al. A reliable small sample classification algorithm by elman neural network based on PLS and GA. Journal of Classification, 2019, 36(2): 306-331. [doi: 10.1007/s00357-018-9288-4]
- 王欣欣, 刘晓彦. 基于双适应度动态遗传算法的云计算资源调度. 计算机工程与设计, 2018, 39(5): 1372-1376, 1421.
- Jia WK, Zhao DA, Zheng YJ, et al. A novel optimized GA-Elman neural network algorithm. Neural Computing and Applications, 2019, 31(2): 449-459. [doi: 10.1007/s00521-017-3076-7]
- Wang J, Wang J, Fang W, et al. Financial time series prediction using elman recurrent random neural networks. Computational Intelligence and Neuroscience, 2016, 2016: 4742515.