

基于参数反演的网络舆情传播趋势预测——以新浪微博为例

刘巧玲, 李 劲*, 肖人彬

(华中科技大学 自动化学院, 武汉 430074)

(* 通信作者电子邮箱 ljhust@hust.edu.cn)

摘 要: 针对现有的舆情传播模型研究与实际舆情数据结合较少以及难以从舆情大数据中挖掘舆情传播内在规律的问题, 提出一种基于实际网络舆情大数据采用神经网络的舆情传播模型参数反演算法。改进经典 SIR 传染病传播模型, 构建一种网络舆情传播模型, 基于该模型对实际案例进行参数反演, 预测网络舆情的后续传播趋势, 并与马尔可夫预测模型对比, 所提算法可以精确预测舆情的具体热度值。实验结果表明, 所提算法在预测性能上具有一定的优越性, 可以用于网络突发事件传播的数据拟合、过程模拟和趋势预测。

关键词: 新浪微博; SIR 模型; 反向传播神经网络; 参数反演; 舆情传播

中图分类号: TP393.09 **文献标志码:** A

Trend prediction of public opinion propagation based on parameter inversion — an empirical study on Sina micro-blog

LIU Qiaoling, LI Jin*, XIAO Renbin

(School of Automation, Huazhong University of Science and Technology, Wuhan Hubei 430074, China)

Abstract: Concerning that the existing researches on public opinion propagation model are seldom combined with the practical opinion data and digging out the inherent law of public opinion propagation from the opinion big data is becoming an urgent problem, a parameter inversion algorithm of public opinion propagation model using neural network was proposed based on the practical opinion big data. A network opinion propagation model was constructed by improving the classical disease spreading Susceptible-Infective-Recovered (SIR) model. Based on this model, the parameter inversion algorithm was used to predict the network public opinion's trend of actual cases. The proposed algorithm could accurately predict the specific heat value of public opinion compared with Markov prediction model. The experimental results show that the proposed algorithm has certain superiority in prediction and can be used for data fitting, process simulation and trend prediction of network emergency spreading.

Key words: Sina micro-blog; Susceptible-Infective-Recovered (SIR) model; Back-Propagation (BP) neural network; parameter inversion; public opinion propagation

0 引言

互联网及社交网络的发展如新浪微博等, 在便利信息传播的同时导致网络舆情传播研究越来越困难。舆情传播模型能对舆情传播的过程和趋势进行描述, 因此, 基于舆情传播模型研究社交网络的舆情传播具有极其重要的意义。

对社交网络舆情的传播国内外都有较多研究。早期的舆情传播研究主要沿用 SIS(Susceptible-Infected-Susceptible)、SIR(Susceptible-Infective-Recovered) 传染病模型^[1-2]。后来很多研究者为了更好地揭示社交网络的舆情传播规律改进这些传统模型: 2012 年, 朱恒民等^[3]以无标度网络为载体提出了舆情传播的 SIRS(Susceptible-Infective-Recovered-Susceptible) 模型; 2013 年, 张彦超等^[4]结合传染病模型构造了一个基于社交网络的信息传播模型; Borge-Holthoefer 等^[5]将社交网络中用户的活跃度和对信息的感知度加入到信息传播模型中; 2014 年, 蒙在桥等^[6]考虑信息的传播延迟, 提出了动态消息传播模型; 张晓伟^[7]借鉴日常生活中人与人之间的信任原

理, 提出了一种基于信任度的消息传播模型; 2015 年, 陈骁等^[8]为探究微博转发网络的构建机制提出了一种有向加权网络模型; 2016 年, 黄宏程等^[9]考虑了网络中用户的不同感染状态, 引入感染用户的衰减函数, 提出了适合社交网络的信息传播模型。

现有的社交网络舆情传播研究多局限于舆情传播模型本身, 研究所提模型的有效性及其参数对舆情传播的影响, 与实际舆情数据结合较少。结合实际的传染病疫情数据对传染病传播的参数反演已有相关研究。韩卫国等人以香港和北京非典 SARS(Severe Acute Respiratory Syndrome) 疫情数据为实例, 采用 SIR 模型对 SARS 传播的时间过程参数反演^[10]。熊焱等^[11]提出一种基于 SIR 模型参数反演的新方法, 以北京 SARS 疫情数据为例进行验证。王香阁等^[12]基于 SIR 模型预测病毒传播的速度。但对于舆情传播的参数反演则不多见。

针对上述问题, 本文基于 SIR 模型构建一种网络舆情传播模型, 并提出了一种基于实际网络舆情大数据采用神经网络的舆情传播模型参数反演算法, 可实现从日益增长的舆情

收稿日期: 2016-11-14; 修回日期: 2016-12-14。 基金项目: 国家自然科学基金资助项目(61540032)。

作者简介: 刘巧玲(1993—), 女, 湖北当阳人, 硕士研究生, 主要研究方向: 舆情传播、大数据; 李劲(1980—), 男, 湖北武汉人, 博士, 主要研究方向: 舆情大数据、神经网络; 肖人彬(1965—), 男, 湖北武汉人, 教授, 博士, 主要研究方向: 复杂系统、复杂社会管理。

大数据中挖掘出舆情传播的内在规律,可用于网络突发事件传播的数据拟合、过程模拟和趋势预测,可为政府部门对网络舆情事件的控制决策提供依据。

1 基本原理

1.1 SIR 模型

Kermack 与 McKendrick 于 1927 年建立了经典的传染病 SIR 模型,将人群划分为易感染者 S (Susceptible)、感染者 I (Infected) 和移出者 R (Recovered) 三类。 S 以一定的概率 β 被 I 感染, I 以一定的概率 γ 被治愈后变成 R ,传播过程如图 1^[13]。

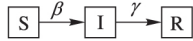


图 1 经典 SIR 模型

Fig. 1 Classical SIR model

经典 SIR 模型具体形式如下:

$$\frac{dS}{dt} = -\beta S(t) I(t) \quad (1)$$

$$\frac{dI}{dt} = \beta S(t) I(t) - \gamma I(t) \quad (2)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (3)$$

$$S(t) + I(t) + R(t) = N(t) \quad (4)$$

式中: $S(t)$ 、 $I(t)$ 、 $R(t)$ 和 $N(t)$ 分别表示 t 时刻易感人数、感染人数、移出人数和总人数, β 和 γ 分别表示感染者的感染率和移出率。

1.2 网络舆情传播模型

近年来学者们纷纷将 SIR 模型应用到复杂网络中^[14-16]。本文在经典 SIR 模型的基础上,综合考虑部分未接收到舆情信息的网民在接收信息后不参与传播和部分接收信息但不传播的网民在舆情衍生话题的影响下变为传播者这两种情况,建立突发事件网络舆情演化传播的 SIR 模型。

网络舆情发生时,基于经典 SIR 模型, S 也可能以概率 d 直接变为 R ; 当原始网络舆情衍生新话题时, R 会因为衍生话题舆情信息选择传播话题或制造新话题变为 I 的概率为 c 。易感染者 S 、感染者 I 和移出者 R 的转换关系如图 2。

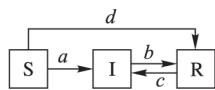


图 2 改进 SIR 模型

Fig. 2 Improved SIR model

改进经典 SIR 模型后,网络舆情传播模型具体形式如下:

$$\frac{dS}{dt} = -aS(t)I(t) - dI(t) \quad (5)$$

$$\frac{dI}{dt} = aS(t)I(t) - bI(t) + cI(t)R(t) \quad (6)$$

$$\frac{dR}{dt} = bI(t) - cI(t)R(t) + dI(t) \quad (7)$$

$$S(t) + I(t) + R(t) = N(t) \quad (8)$$

其中: a 为感染率, b 为衰退率, c 为衍生影响率, d 为转化率。本文舆情传播模型的易感染者 S 为未接收到舆情信息的用户,感染者 I 为接收到舆情信息并传播的用户,移出者 R 为接收到舆情但不传播或未接收到舆情且不传播的用户。

2 相关方法

2.1 Runge-Kutta 方法

龙格-库塔 (Runge-Kutta) 方法利用级数法的思想,间接使

用泰勒展开式和斜率表达微分,应用广泛,精度高,可保证求解过程的稳定性和解的收敛性。

微分方程组 (5) ~ (7) 可归结为一阶常微分方程组 (9):

$$\begin{cases} \frac{dS}{dt} = f_1(t, S, I, R) \\ \frac{dI}{dt} = f_2(t, S, I, R) \\ \frac{dR}{dt} = f_3(t, S, I, R) \end{cases} \quad (9)$$

我们采用经典 Runge-Kutta 四阶方法求解此微分方程模型的数值解,求解公式如下:

$$\begin{cases} S_{n+1} = S_n + \frac{h}{6}(k_{11} + 2k_{21} + 2k_{31} + k_{41}) \\ I_{n+1} = I_n + \frac{h}{6}(k_{12} + 2k_{22} + 2k_{32} + k_{42}) \\ R_{n+1} = R_n + \frac{h}{6}(k_{13} + 2k_{23} + 2k_{33} + k_{43}) \end{cases} \quad (10)$$

$$\begin{cases} k_{1i} = f_i(t, S, I, R) \\ k_{2i} = f_i\left(t_n + \frac{h}{2}, S_n + \frac{h}{2}k_{11}, I_n + \frac{h}{2}k_{12}, R_n + \frac{h}{2}k_{13}\right) \\ k_{3i} = f_i\left(t_n + \frac{h}{2}, S_n + \frac{h}{2}k_{21}, I_n + \frac{h}{2}k_{22}, R_n + \frac{h}{2}k_{23}\right) \\ k_{4i} = f_i\left(t_n + h, S_n + h k_{31}, I_n + h k_{32}, R_n + h k_{33}\right) \end{cases} \quad (11)$$

其中 $i = 1, 2, 3$; h 为步长。

2.2 舆情预测方法

常见的预测模型有两大类: 基于传统统计学的预测方法,如时间序列模型、参数回归模型等,它们更适用于线性关系的预测; 另一类基于智能机器学习算法的预测方法,如支持向量机模型、马尔可夫模型、神经网络模型等^[17]。支持向量机模型对缺失数据敏感,核函数及参数的选择没有统一的模式。马尔可夫模型通过划分状态空间构造状态转移矩阵,预测精确性取决于未来一期的预测。而 BP (Back-Propagation) 神经网络模型构造简单,无需建立解析数学模型,计算效率高,逼近效果好,非线性拟合能力强。因此,本文采用贝叶斯正则化 BP 神经网络进行群体性突发事件舆情传播的参数反演和趋势预测,网络结构如图 3 所示。

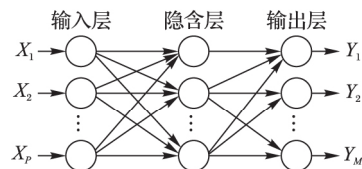


图 3 BP 神经网络结构

Fig. 3 BP neural network structure

本文 BP 网络的输入为改进 SIR 模型 $I(t)$ 的数值解,对应的参数值 (a, b, c, d, N) 作为网络的输出。

3 参数反演算法

本文结合 Runge-Kutta 微分方程数值解法和 BP 神经网络技术,构造一种适合于网络舆情传播模型的参数反演算法,进行群体性突发事件舆情传播的参数反演和趋势预测。

参数反演算法的求解步骤如下:

1) 确定改进 SIR 模型的 a, b, c, d 和 N 的取值范围 $I_a, I_b,$

I_c, I_d 和 I_N 并分别在这些参数的取值区间内选取 n 组参数值 $(a_i, b_i, c_i, d_i, N_i)$, 其中: $a_i \in I_a, b_i \in I_b, c_i \in I_c, d_i \in I_d, N_i \in I_N (i = 1, 2, \dots, n)$ 。

2) 对于每组参数值 $(a_i, b_i, c_i, d_i, N_i)$, 利用 Runge-Kutta 方法求解微分方程组 (5) ~ (8) 求出一组 $I(t)$ 的数值解 $I_{i,1}, I_{i,2}, \dots, I_{i,M}$, 将 n 组 $I(t)$ 数值解作为网络的输入, 对应的参数值 $(a_i, b_i, c_i, d_i, N_i)$ 作为网络的输出, 构造 n 组训练样本, 同时用相同方法构造 m 组测试样本。

3) 根据得到的训练样本使用贝叶斯正则化方法训练 BP 神经网络, 并用测试样本测试训练效果。

4) 将实际突发事件舆情传播的相关数据输入训练好的神经网络, 由神经网络的输出得到改进 SIR 舆情传播模型的反演参数值 a, b, c, d 和 N 。

5) 利用反演参数值 a, b, c, d 和 N 和改进 SIR 舆情传播模型预测后 L 天的舆情传播人数 $I(M+1), I(M+2), \dots, I(M+L)$ 。

4 案例分析

4.1 案例选择和数据获取

“医疗安全”问题历来是社会关注的热点话题, 2016 年 4 月 28 日针对自媒体曝出“魏则西”之死事件存在的涉事医院外包诊所给民营机构、百度竞价排名等问题, 引起了网民的热议。为验证上述算法的可行性, 选取目前国内最热门的社交网络——新浪微博平台“魏则西事件”作为本次研究实例。

采用爬虫技术抓取新浪微博相关的数据。以“魏则西”为搜索关键词, 以每两个小时为搜索区间, 抓取新浪微博上 2016-04-28—2016-05-28 一个月的微博, 针对每条微博爬取微博发布者的昵称、发布时间、转发数、评论数、赞数等。每个时段的微博数量可反映舆情传播中传播者的舆情传播水平。

4.2 数据整理和分析

针对以上数据, 提取以每两个小时为区间的微博数量。因为新浪微博的博文数据和时间息息相关, 通常每天 2:00 ~ 6:00 博文发布量偏少, 9:00 ~ 10:00 及 22:00 ~ 23:00 是博文发布的高峰期, 因此将每天的博文量以 4:00 ~ 16:00 (第一时间段) 和 16:00 ~ 4:00 (次日) (第二时间段) 时间段进行分段统计, 得到该事件新浪微博平台舆情传播的真实数据, 时间长度为 31 d, 数据个数为 62。部分数据如表 1 所示。

表 1 “魏则西”事件相关数据

Tab. 1 Related data of “WEI Zexi” event

| 日期 | 第一时间段博文量 | 第二时间段博文量 | 日期 | 第一时间段博文量 | 第二时间段博文量 |
|-------------|----------|----------|-------------|----------|----------|
| 04-29—04-30 | 5 | 4 | 04-30—05-01 | 6 | 26 |
| 05-01—05-02 | 115 | 810 | 05-02—05-03 | 1 478 | 1 442 |
| 05-03—05-04 | 1 619 | 1 289 | 05-04—05-05 | 1 181 | 775 |
| 05-05—05-06 | 723 | 482 | 05-06—05-07 | 676 | 481 |
| 05-07—05-08 | 344 | 772 | 05-08—05-09 | 517 | 284 |
| 05-09—05-10 | 193 | 345 | 05-10—05-11 | 346 | 189 |
| 05-11—05-12 | 142 | 96 | 05-12—05-13 | 61 | 64 |
| 05-13—05-14 | 56 | 55 | 05-14—05-15 | 46 | 41 |
| 05-15—05-16 | 28 | 36 | 05-16—05-17 | 35 | 34 |

4.3 网络舆情的预测

以往研究大多以日微博总量衡量舆情传播者的传播水平, 但群体性突发事件一旦发生如同火山爆发般迅速传播, 到

达高峰期时间极短, 而利用事件增长长期数据预测高峰期、衰退期等更有现实意义。因此, 本文采用以上分段处理的博文量作为模型仿真的数据集反演 SIR 模型的参数, 进而预测网络舆情的传播。由于博文量与一天中的不同时间段息息相关, 因此本文将日博文量以 4:00 ~ 16:00 和 16:00 ~ 4:00 (次日) 时间段进行统计, 用 04-29—05-01 前三天的 6 组数据作为输入数据进行参数反演获取感染率、衰退率等参数, 利用传播模型预测该事件后续舆情传播的趋势走向。该算法的步骤如下:

1) 舆情传播模型参数的取值范围, 感染率 $a \in [1.6 \times 10^{-5}, 2.6 \times 10^{-5}]$, 衰退率 $b \in [0.004, 0.014]$, 衍生影响率 $c \in [1.0 \times 10^{-7}, 2.0 \times 10^{-6}]$, 转化率 $d \in [1.0 \times 10^{-7}, 2.0 \times 10^{-6}]$, $N \in [2 000, 3 500]$;

2) 在上述参数的取值区间内随机选取 10 000 组参数值, 利用 Runge-Kutta 方法得到 10 000 组 $I(t)$ 的数值解, 将每个 $I(t)$ 数值解的前 6 个数值作为网络的输入, 对应的参数值 a, b, c, d, N 作为网络的输出, 构造 10 000 组训练样本;

3) 利用 BP 神经网络技术训练, 输入层节点数 6, 输出层节点数 5, 隐含层节点数 8, 最大迭代次数 1 000, 目标误差 0.01, 学习速率 0.05, 训练算法 trainbr, 训练函数 tansig;

4) 网络训练好后, 以前三天此事件的 6 组实际微博数作为网络的输入, 反演求得 SIR 模型的参数 a, b, c, d 和 N ;

5) 以上过程重复 100 次, 求得模型参数 a, b, c, d 和 N 的平均值, 代入 SIR 模型, 可获得该事件后续舆情传播的趋势走向及高峰期、高峰期传播人数等信息。

4.4 预测结果分析

4.4.1 “魏则西”事件预测结果分析

利用该事件前三天的微博数据参数反演 100 次, 求得 SIR 模型参数 a, b, c, d 和 N 的平均值, 结果如下: $a = 2.110 341 45 \times 10^{-5}$, $b = 0.009 987 93$, $c = 1.000 341 25 \times 10^{-7}$, $d = 1.010 321 45 \times 10^{-7}$, $N = 2 989$, 代入改进 SIR 模型, 部分预测结果如表 2 所示。表 2 列出了利用该事件前三天的微博数据参数反演预测后 16 天(05-02—05-17)的舆情传播过程。

对比参数反演的预测结果和实测数据, 如图 4 所示。

由上述仿真结果可以看出, 结合实际社交网络的实际情况网络舆情大数据, 利用该事件前三天的微博数据, 基于舆情传播模型的参数反演算法, 可挖掘感染率、衰退率等舆情传播模型参数, 预测该事件后续的舆情传播过程。仿真预测结果与实际数据变化趋势基本吻合, 拟合效果较好, 说明可以用基于网络舆情传播模型的参数反演算法来仿真模拟该类群体性突发事件的演化传播, 并进行相应的分析和预测。

1) 该事件爆发于 4 月底, 04-29—04-30 为缓慢增长期, 从 05-01 开始, 该事件如同火山爆发般迅速传播, 进入爆发阶段, 05-03 达到最高峰, 预测结果和实际数据吻合; 预测高峰期的微博量为 1 631, 与实际数据 1 619 相比, 误差为 12。

2) 05-05—05-12 是该事件的衰退期。此阶段传播者比例呈现递减的态势, 是事件衍生话题的潜伏期和关键期。此阶段实际数据存在很大的波动, 因为该阶段存在一定的随机因素的影响, 如“魏则西事件”涉及百度、央视、莆田系, 导致网民热度上升, 使网络舆情进入二度爆发

阶段。

表 2 “魏则西”事件预测结果
Tab. 2 Prediction results of “WEI Zexi” event

| 日期 | 实测值 | 预测值 | 误差率/% | 日期 | 实测值 | 预测值 | 误差率/% |
|-------|------|------|-------|-------|-----|-----|-------|
| 05-02 | 1442 | 1547 | 7.3 | 05-10 | 189 | 148 | 21.7 |
| 05-03 | 1619 | 1631 | 0.7 | 05-11 | 96 | 82 | 14.4 |
| 05-04 | 1181 | 1334 | 12.9 | 05-12 | 61 | 67 | 11.1 |
| 05-05 | 723 | 935 | 29.3 | 05-13 | 56 | 48 | 14.3 |
| 05-06 | 676 | 665 | 1.7 | 05-14 | 46 | 39 | 15.2 |
| 05-07 | 344 | 423 | 22.9 | 05-15 | 28 | 23 | 17.8 |
| 05-08 | 284 | 259 | 9.0 | 05-16 | 35 | 20 | 42.8 |
| 05-09 | 193 | 213 | 10.6 | 05-17 | 25 | 15 | 40.0 |

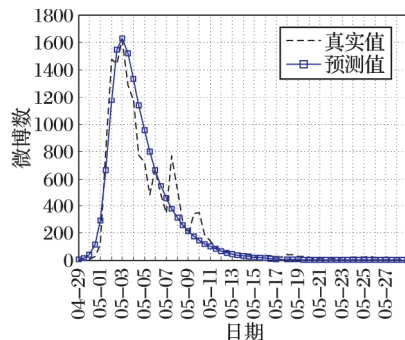


图 4 “魏则西”事件模型预测结果与实际数据比较

Fig. 4 Comparison of the predicted results and the actual data of the “WEI Zexi” event

3) 05-13 开始是该事件的平息期。经过衰退期后,网民对此事件的关注热度出现疲态,逐渐平息直至淡出网民视野。

4) 整个舆情传播过程,与实际数据相比,预测微博量的平均绝对误差为 11.8,平均相对误差为 14.2%。

4.4.2 算法有效性分析

为了更好地验证所提算法的有效性以及更好地比较预测

结果,抓取另一个案例“人机大战”2016-03-06—2016-03-21 的相关新浪微博量,类似方法进行处理,分别建立基于马尔可夫模型的网络舆情预测模型和基于 BP 神经网络的参数反演网络舆情预测模型,分别利用该事件发展期的前三天 6 个数据预测后续舆情传播趋势。

基于马尔可夫模型的网络舆情预测中,经计算获得的状态空间:急速上升 $Z_1 = [17.98, 35.97]$,缓慢上升 $Z_2 = [0, 17.98]$,缓慢下降 $Z_3 = [-15.16, 0]$,极速下降 $Z_4 = [-30.32,$

$$-15.16]; \text{状态转移矩阵 } P = \begin{bmatrix} 0.67 & 0 & 0.33 & 0 \\ 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0.33 & 0.6 & 0.07 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}。$$

基于 BP 神经网络的参数反演网络舆情预测中,参数反演求得的 SIR 模型参数 $a = 2.8986754 \times 10^{-5}$, $b = 0.0219742$, $c = 3.9836456 \times 10^{-6}$, $d = 1.2753234 \times 10^{-6}$, $N = 2686$ 。

利用 03-06—03-08 三天的 6 个数据预测后续舆情传播趋势(03-09—03-21),两者的预测结果对照如表 3 图 5 所示。

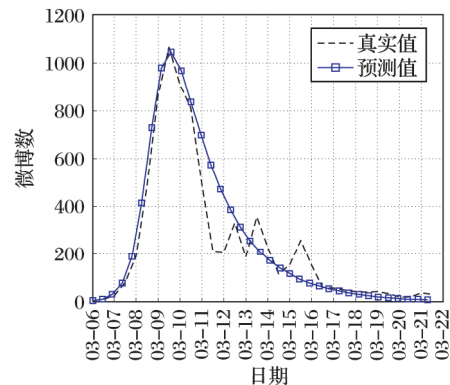


图 5 BP 神经网络参数反演预测结果

Fig. 5 Prediction results of parameter inversion based on BP neural network

表 3 本文方法与马尔可夫模型预测结果对比

Tab. 3 Comparison of the predicted results of the proposed model and Markov model

| 编号 | 真实值 | 马尔可夫模型 真实状态 | 预测状态 | 参数反演 预测值 | 编号 | 真实值 | 马尔可夫模型 真实状态 | 预测状态 | 参数反演 预测值 |
|----|------|----------------|-------|-------------|----|-----|----------------|-------|-------------|
| 1 | 4 | Z_2 | — | 4 | 9 | 228 | Z_3 | Z_3 | 178 |
| | 11 | Z_2 | — | 12 | | 115 | Z_2 | Z_3 | 143 |
| 2 | 25 | Z_2 | — | 35 | 10 | 155 | Z_2 | Z_3 | 115 |
| | 80 | Z_2 | — | 103 | | 257 | Z_3 | Z_3 | 92 |
| 3 | 189 | Z_1 | — | 271 | 11 | 152 | Z_3 | Z_3 | 74 |
| | 482 | Z_1 | — | 584 | | 60 | Z_3 | Z_3 | 59 |
| 4 | 864 | Z_1 | Z_1 | 915 | 12 | 55 | Z_3 | Z_3 | 48 |
| | 1066 | Z_3 | Z_1 | 1047 | | 48 | Z_3 | Z_3 | 39 |
| 5 | 912 | Z_3 | Z_3 | 981 | 13 | 41 | Z_3 | Z_3 | 31 |
| | 810 | Z_4 | Z_3 | 835 | | 39 | Z_2 | Z_3 | 25 |
| 6 | 488 | Z_4 | Z_3 | 683 | 14 | 43 | Z_3 | Z_3 | 20 |
| | 212 | Z_3 | Z_3 | 549 | | 34 | Z_3 | Z_3 | 16 |
| 7 | 210 | Z_2 | Z_3 | 438 | 15 | 20 | Z_3 | Z_3 | 13 |
| | 330 | Z_3 | Z_3 | 350 | | 19 | Z_2 | Z_3 | 11 |
| 8 | 189 | Z_2 | Z_3 | 279 | 16 | 37 | Z_3 | Z_3 | 9 |
| | 356 | Z_3 | Z_3 | 223 | | 29 | — | — | 7 |

注:—表示没有进行预测。

由上述对同一事件,采用两种不同方法进行预测的结果可见:1) 马尔可夫模型舆情预测通过已知状态预测未来的状

态,只能预测未来所处的状态空间,不能得到舆情的具体热度值,而本文算法可以精确预测舆情的具体热度值,更有实际意

义; 2) 马尔可夫模型通过舆情热度计算、状态空间划分、状态转移矩阵构造进行舆情热度趋势预测, 状态转移矩阵的构造需要大量的数据, 而网络舆情突发事件一旦发生往往进展很快, 到达高潮的时间极短, 数据量不足, 这限制了马尔可夫模型的使用, 而本文算法利用事件发展期的数据即可以进行参数反演, 进而预测高潮期时间、高潮峰值等数据, 更有现实意义。从图5可见实际数据的衰退期有局部点数据出现了跳跃现象, 经查发现, 其原因是3月13日“围棋人机大战第四局较量中李世石接连三局负于“阿尔法围棋”后迎来了首次胜利, 致使网民的热度上升。可见, 实际舆情数据受各种随机因素的影响可能在局部点发生一定偏离, 但模型对事件发展的整体趋势的预测是正确的, 这表明通过模型参数反演获取到了舆情传播事件发展的内在规律, 能够正确描述舆情传播进程在其内在机制驱动下的演化和发展规律。

5 结语

本文在传统传染病模型的基础上构建了一种网络舆情传播模型, 并提出了一种基于实际网络舆情大数据采用神经网络完成舆情传播模型参数反演的算法, 实现了基于实际舆情传播数据从日益增长的舆情大数据中挖掘出舆情传播的内在规律。利用该参数反演算法预测网络舆情传播的趋势, 并以新浪微博突发事件实际抓取数据为例进行了验证, 并与马尔可夫预测模型进行了对比。结果表明, 基于实际网络舆情大数据采用神经网络完成舆情传播模型参数反演的算法可以预测网络舆情的传播趋势且性能更优, 这对指导政府部门制定网络舆情的控制策略有着重要的参考价值。

在下一步的研究中, 拟构建一种考虑空间传播过程的网络结构动态演化的舆情传播模型。基于本文的参数反演算法进行实时参数反演, 实现舆情传播过程中模型的参数动态演化, 同时结合舆情大数据的地理位置信息对舆情的传播过程开展更进一步的研究。

参考文献 (References)

- [1] 张发, 李璐, 宣慧玉. 传染病传播模型综述[J]. 系统工程理论与实践, 2011, 31(9): 1736 - 1744. (ZHANG F, LI L, XUAN H Y. Survey of transmission models of infectious diseases[J]. Systems Engineering - Theory & Practice, 2011, 31(9): 1736 - 1744.)
- [2] 刘丹, 殷亚文, 宋明. 基于 SIR 模型的微博信息扩散规律仿真分析[J]. 北京邮电大学学报, 2014, 16(3): 28 - 33. (LIU D, YIN Y W, SONG M. Simulation analysis of micro-blog information diffusion based on SIR model[J]. Journal of Beijing University of Posts and Telecommunications, 2014, 16(3): 28 - 33.)
- [3] 朱恒民, 李青. 面向话题衍生性的微博网络舆情传播模型研究[J]. 现代图书情报技术, 2012, 28(5): 60 - 64. (ZHU H M, LI Q. Public opinion propagation model with topic derivatives in the micro-blog network[J]. New Technology of Library and Information Service, 2012, 28(5): 60 - 64.)
- [4] 张彦超, 刘云, 张海峰, 等. 基于在线社交网络的信息传播模型[J]. 物理学报, 2011, 60(5): 50501. (ZHANG Y C, LIU Y, ZHANG H F, et al. The research of information dissemination model on online social network[J]. Acta Physica Sinica, 2011, 60(5): 50501.)
- [5] BERGE-HOLTHOEFER J, MELONI S, GONCALVES B, et al. Emergence of influential spreaders in modified rumor models[J]. Journal of Statistical Physics, 2013, 151(1/2): 383 - 393.
- [6] 蒙在桥, 傅秀芬. 基于社交网络的动态消息传播模型[J]. 计算机应用, 2014, 34(7): 1960 - 1963. (MENG Z Q, FU X F. Dynamic information spreading model based on online social network[J]. Journal of Computer Applications, 2014, 34(7): 1960 - 1963.)
- [7] 张晓伟. 在线社交网络上基于信任度的消息传播模型[J]. 计算机应用, 2014, 34(2): 411 - 416. (ZHANG X W. Trust-based information propagation model in online social networks[J]. Journal of Computer Applications, 2014, 34(2): 411 - 416.)
- [8] 陈骁, 黄曙光, 秦李. 基于微博转发的社交网络模型[J]. 计算机应用, 2015, 35(3): 638 - 642. (CHEN X, HUANG S G, QIN L. Social network model based on micro-blog transmission[J]. Journal of Computer Applications, 2015, 35(3): 638 - 642.)
- [9] 黄宏程, 蒋艾玲, 胡敏. 基于社交网络的信息传播模型分析[J]. 计算机应用研究, 2016, 33(7): 2738 - 2742. (HUANG H C, JIANG A L, HU M. Analysis of information diffusion model on social network[J]. Application Research of Computers, 2016, 33(7): 2738 - 2742.)
- [10] 韩卫东, 王劲峰, 刘旭华. SARS 传播时间过程的参数反演和趋势预测[J]. 地球科学进展, 2004, 19(6): 925 - 930. (HAN W D, WANG J F, LIU X H. Parameter inversion and trend prediction of SARS propagation process[J]. Advances in Earth Science, 2004, 19(6): 925 - 930.)
- [11] 熊焱, 王莉, 李大卫, 等. SIR 传染病模型的参数反演[J]. 生物数学学报, 2009, 24(1): 129 - 135. (XIONG Y, WANG L, LI D W, et al. Parameter inversion of SIR epidemic model[J]. Journal of Biomathematics, 2009, 24(1): 129 - 135.)
- [12] 王香阁, 于红斌, 王飘飘, 等. 基于 SIR 模型对埃博拉病毒蔓延趋势的预测[J]. 福建电脑, 2015, 14(2): 97, 115. (WANG X G, YU H B, WANG P P, et al. The prediction of the spread of Ebola virus based on SIR model[J]. Fujian Computer, 2015, 14(2): 97, 115.)
- [13] KERMAK W O, MCKENDRICK W G. A contribution to the mathematical theory of epidemics[J]. Proceedings of the Royal Society of London: Series A, 1927, 115(1): 700 - 721.
- [14] YUAN X P, XUE Y K, LIU M X. Global stability of an SIR model with two susceptible groups on complex networks[J]. Chaos, Solitons & Fractals, 2014, 59(2): 42 - 50.
- [15] BALL F, SIRL D, TRAPMAN P. Analysis of a stochastic SIR epidemic on a random network incorporating household structure[J]. Mathematical Biosciences, 2010, 224(2): 53 - 73.
- [16] 李可嘉, 王义康. 改进 SIR 模型在社交网络信息传播中的应用[J]. 电子科技, 2013, 26(8): 168 - 171. (LI K J, WANG Y K. Application of improved SIR model in information transmission in social networks[J]. Electronic Science and Technology, 2013, 26(8): 168 - 171.)
- [17] 游丹丹, 陈福集. 基于改进粒子群和 BP 神经网络舆情预测研究[J]. 情报杂志, 2016, 35(8): 156 - 161. (YOU D D, CHEN F J. Research on the prediction of network public opinion based on improved PSO and BP neural network[J]. Journal of Intelligence, 2016, 35(8): 156 - 161.)

This work is partially supported by the National Natural Science Foundation of China (61540032).

LIU Qiaoling, born in 1993, M. S. candidate. Her research interests include public opinion propagation, big data.

LI Jin, born in 1980, Ph. D. His research interests include public opinion big data, neural network.

XIAO Renbin, born in 1965, Ph. D., professor. His research interests include complex system, complex social management.