

一种改进的微博网络信息传播与预测模型

丁鑫,刘其成,张伟

(烟台大学计算机与控制工程学院,山东烟台 264005)

摘要:微博的快速发展使得人们进入自媒体时代,也为信息的快速传播创造了有利条件.提出一种考虑到直接免疫的改进 SIR 模型,利用该模型研究微博信息的传播规律并对微博信息进行预测.实验结果表明,与微博标准 SIR 传播模型相比,提出的改进 SIR 模型能更有效地反映微博信息传播规律,并且能更准确地对微博信息传播进行预测.

关键词:微博信息传播;直接免疫;SIR;预测

中图分类号: TP311 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2014.07.007

引用格式: Ding Xin, Liu Qicheng, Zhang Wei. An improved model for information dissemination and prediction on microblog networks[J]. Journal of University of Science and Technology of China, 2014, 44(7): 582-589, 598.

丁鑫,刘其成,张伟. 一种改进的微博网络信息传播与预测模型[J]. 中国科学技术大学学报, 2014, 44(7): 582-589, 598.

An improved model for information dissemination and prediction on microblog networks

DING Xin, LIU Qicheng, ZHANG Wei

(School of Computer and Control Engineering, Yantai University, Yantai 264005, China)

Abstract: The rapid development of microblogging not only has led people into the We-Media era, but also creates favorable conditions for the rapid dissemination of information. Taking direct immunization into consideration, an improved SIR model was presented which uses differential equations to study the rules of information dissemination and to predict information on microblogs. The experiment result shows that the improved SIR model can effectively reflect the rules of information dissemination, and can predict microblog information dissemination more precisely than the microblog standard SIR model.

Key words: microblog information dissemination; direct immunization; SIR; prediction

0 引言

微博具有个体性的特点,绝大多数用户在网上海发表的大多数微博是带有明显个人情绪的主观看

法.随着 Facebook, Twitter, Linkedin 及新浪微博等社交平台的出现,在巨大的个人信息库中也加入了用户的想法、喜好和人脉关系.文献[1]提出世界上在不同文化背景中的人每天甚至每周的心情都遵

收稿日期:2014-03-21;修回日期:2014-06-15

基金项目:国家自然科学基金(61170224),山东省科技发展计划项目(2012GGB01017),山东省自然科学基金(ZR2011FL018, ZR2012FL07),山东省计算机网络重点实验室开放课题基金(SDKLCN_2012_03)资助.

作者简介:丁鑫,男,1987年生,硕士生.研究方向:海量数据处理和多 Agent 系统. E-mail: newbruce@aliyun.com

通讯作者:刘其成,博士/教授. E-mail: ytliuqc@163.com

循着相似的模式,情绪已经可以数据化。

微博还具有及时性的特点,再加上微博的快速传播,使得人们现在进入了自媒体时代.与传统媒体相比,在突发事件的传播速度上微博是最快的,但是这也为谣言的传播创造了有利的条件.2011年全国食盐脱销,原因就是网络上传播谣言说碘能防核辐射,再加上很多不明真相的用户在短时间内大量转发传播谣言,从而使人们的日常生活也受到了影响。

在微博网络中进行信息处理,并且对微博中信息传播模型规律进行研究和预测,这有着非常重要的现实意义.研究信息在微博网络中的传播规则,尽量在短时间内发现谣言,使政府也可以有效地监管微博网络,从而构建一个和谐的网络环境。

1 相关研究

传统的传播行为都是遵循某些规则的,比如计算机网络中病毒的蔓延^[2]、人群中传染病的流行^[3]、社会中谣言的扩散^[4-5]等.在利用传统传播模型对传播行为进行的研究中,最经典并且应用最广泛的模型有两个,分别是 SIR 模型^[6]和 SIS 模型^[7].在这两个模型的基础上,很多学者提出了适用于各种情况的传播模型,并进行了大量的研究.文献[8]利用平均场近似方法提出基于交通流量 SIS 病毒修正扩散模型.文献[9]对移动环境下网络病毒的传播进行了研究,提出了 SEIRS 模型.文献[10]从数学角度出发,分析了 SIS 病毒传播过程中的稳定性和分岔行为.文献[11]研究了流行病在全接触和单点接触两个方面的传播动力学行为,得出了多关系会使流行病爆发阈值降低的结论.把免疫失败和免疫失效同时作用在均匀网络和无标度网络上的影响用于研究复杂网络中传播的动力学行为,并且提出了考虑不完全免疫的病毒传播模型^[12].文献[13]针对分布式传染流行病的影响,提出了一种改进的 SI 模型。

社会性网络服务(social networking services, SNS)的传播方法与传统的传播方法是不一样的,因为 SNS 网络中两个人之间的交互关系可能对信息与人之间的关系产生影响,进而信息与人之间的关系又会反作用于两个人之间的交互关系.怎样来描述网络中信息的传播行为并且发现它的特性,这有着非常重要的理论和应用价值.在 Sznajd-Weron 模型^[14]中,所有人都在格点上,每人都必须在两个观点中选择一个,当任意两个相邻的人都持有相同的观点时,就可以劝说他们的邻居采取相同的选择,进

而拥有相同的观点.随后,在 Deffuant 模型^[15]和 Hegselmann 模型^[16]中将观点值建立在大于 0 和小于 1 之间的连续尺度上,使之具有更加精细的量化特点.文献[17]提出基于 SNS 网络的一种观点传播模型,它研究了 SNS 网络中舆论观点扩散的形式与特征.考虑到真实信息和谣言在传播结构和时序上的不均匀性,文献[18]提出了在时变传播规则下危机信息的 ISKS-ISK 两阶段传播模型,并且对减少危机中谣言的危害传播提出了有用的建议.考虑到边权是认可概率的关键影响因子,文献[19]提出了一种短信网络的加权演化模型,对加权短信网络上的谣言传播行为进行了研究。

文献[20-21]指出社交网络中信息的传播与传染病的传播非常相似,可以认为信息的传播也符合 SIR 模型.文献[17]提出的基于在线社交网络信息传播模型把用户的状态分成健康者、传播者和免疫者三种状态.文献[22]提出了有疾病潜伏期的传染病动力学模型,根据这一模型,文献[23]中提出了 SEIR 模型,此模型考虑到潜伏态节点多次与传播节点接触而变为传播态.文献[24]提出了基于移动社交网络上的 CSR 谣言传播模型,利用多 agent 仿真平台对 CSR 模型的传播动力学进行了研究.在文献[25]中,钱颖等在标准传染病 SIR 模型的基础上对微博信息的传播进行研究,得出 SIR 模型基本符合微博信息传播规律的结论,所以 SIR 模型也被认为是一种微博信息传播模型。

此前的各种网络信息传播模型研究,较少考虑信息免疫者对信息易感者直接转变为信息免疫者的影响.而在实际微博信息传播过程中,信息易感者与信息免疫者在交互信息时,可使一部分信息易感者直接变为信息免疫者,即直接免疫.本文提出的改进 SIR 微博信息传播模型中考虑到了直接免疫的情况,利用此模型研究微博网络中信息的传播规律和预测方法,并进行了相关实验。

2 改进的微博信息传播模型建模

2.1 微博信息传播规则

在微博平台中,每个用户可以都看到他们所关注的用户发布的微博,并且会以一定的概率评论转发,评论转发的微博同样也会被其他关注的用户看到.在微博的信息传播过程中,相信信息的真实性并且转发微博的用户成为信息传播者;如果对信息的内容不感兴趣或者从一开始就不相信信息的真实性

从而不转发,这些用户成为信息免疫者。

经典传播模型(传染病模型)中一般将用户所处的状态分成三类:健康者,传播者和免疫者。在微博网络上,如果用节点表示用户,用边表示用户之间的关系,微博信息只能沿着边传播,从而把节点分为三类:未感染节点、传播节点和免疫节点。未感染节点从没有接受过来自邻居节点的信息,在以后会以一定的概率从邻居节点接触到信息。接触到信息后,未感染节点会以一定的概率成为传播节点。传播信息后,传播节点会以一定概率变为免疫节点,免疫节点不会传播信息。

2.2 相关定义

定义 2.1 微博网络

微博网络符合小世界网络的特性,由于小世界网络的度是符合幂指数分布的,所以微博网络可以被认为是所有节点的度都近似相等的均匀网络。利用统计物理学中的平均场近似方法,忽略那些局部不平等的相互作用,把微博网络中的节点信息交互作用平均化。由此微博网络可表示为一个二元组 $\{U, T\}$ 。

U 代表微博用户节点的有限集合, $u_i \in U$ 代表具体的节点,令节点总数目为 N ,则

$$U = \{u_1, u_2, \dots, u_N\};$$

T 为 $U \times U$ 的任一子集合,

$$T = \{(u_i, u_j) \mid u_i, u_j \in U\}.$$

若存在 $(u_i, u_j) \in T$,则表示 u_i 与 u_j 进行信息交互,即阅读微博、转发或者评论微博。

定义 2.2 信息传播网络

微博中的信息传播网络可表示为一个四元组 $\{S, I, R, T\}$ 。

S 代表从未听说过信息的微博网络节点,处于未感染状态,类似于传染病模型中的易感者,设总数为 l ,

$$S = \{s_1, s_2, \dots, s_l\}.$$

I 代表听说过信息并传播的微博网络节点,处于传播状态,类似于传染病模型中的感染者,设总数为 m ,

$$I = \{i_1, i_2, \dots, i_m\}.$$

R 代表听到过信息但并不传播的微博网络节点,处于免疫状态,类似于传染病模型中的免疫者,设总数为 n ,

$$R = \{r_1, r_2, \dots, r_n\}.$$

其中, $S \subseteq U, I \subseteq U, R \subseteq U$, 并且,

$$l + m + n = N.$$

T 有六种情况,代表着 S, I 以及 R 相互交互的

6 种结果,即微博网络中信息交互有 6 种情况,分别是:① s_i 与 s_j 交互,即 s_i 与 s_j 接触,双方没有变化, $(s_i, s_j) \rightarrow (s_i, s_j)$. ② s_i 与 i_j 交互,即 s_i 与 i_j 接触, s_i 会以一定概率变为 i_i , i_j 不变, $(s_i, i_j) \rightarrow (i_i, i_j)$. ③ s_i 与 r_j 交互,即 s_i 与 r_j 接触, s_i 会以一定概率变为 r_i , r_j 不变, $(s_i, r_j) \rightarrow (r_i, r_j)$. ④ i_i 与 i_j 交互,即 i_i 与 i_j 接触,双方没有变化, $(i_i, i_j) \rightarrow (i_i, i_j)$. ⑤ i_i 与 r_j 交互,即 i_i 与 r_j 接触, i_i 会以一定概率变为 r_i , r_j 不变, $(i_i, r_j) \rightarrow (r_i, r_j)$. ⑥ r_i 与 r_j 交互,即 r_i 与 r_j 接触,双方没有变化, $(r_i, r_j) \rightarrow (r_i, r_j)$.

由以上 6 种交互结果来看,产生变化的有三种,分别是第②、第③和第⑤种情况。其中第③种情况显示 s_i 没有经过 i_i 的状态转换,而是直接变为 r_i ,产生了直接免疫的情况。

引进封闭世界假设,即信息从微博网络中产生,并且信息仅在微博网络中传播。同时,假设微博网络中未感染节点和传播节点均匀混合。

根据上面的相关定义和假设,研究微博信息传播规律和预测方法。

2.3 考虑直接免疫的微博信息传播模型建模

文献[21]指出社交网络中信息的传播与传染病的传播非常相似,信息的传播符合 SIR 模型。在文献[25]中, SIR 模型也被用于研究微博信息的传播, SIR 模型也是一种微博信息传播模型。

本节考虑直接免疫的情况,针对微博信息传播提出改进的 SIR 传播模型,并与标准 SIR 微博信息传播模型进行比较,阐述直接免疫在微博网络中对于信息传播的性质和意义。

标准 SIR 微博信息传播模型微分方程:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\lambda\langle k \rangle I(t) S(t) \\ \frac{dI(t)}{dt} &= \lambda\langle k \rangle I(t) S(t) - \beta\langle k \rangle I(t) \\ \frac{dR(t)}{dt} &= \beta\langle k \rangle I(t) \end{aligned} \right\} \quad (1)$$

定义 2.3 节点密度

微博网络中, $S(t)$ 表示群体中在 t 时刻,处于未感染状态的节点占微博网络节点总数的比例,称为未感染节点密度; $I(t)$ 表示群体中在 t 时刻,处于传播状态的节点占微博网络节点总数的比例,称为传播节点密度; $R(t)$ 表示群体在 t 时刻,处于免疫状态的节点占微博网络节点总数的比例,称为免疫节点密度。

$S(t), I(t), R(t)$ 三者满足归一化条件:

$$S(t) + I(t) + R(t) = 1.$$

定义 2.4 转化概率

k 为微博网络中的平均节点度, 则:

$\lambda\langle k \rangle$: 从未听过信息的 S 遇到传播信息的 I 时 S 转化为 I 的概率;

$\beta\langle k \rangle$: 传播信息的 I 转化为不传播信息的 R 的概率;

$n\langle k \rangle$: 当 S 与 R 交互时 S 转化为 R 的概率.

定义 2.5 直接免疫

在标准 SIR 微博信息传播模型中, 易感者必须经过传播者这个状态才能转变成免疫者, 即上文中 T 的 6 种情况中第 ② 和第 ⑤ 种情况, 免疫者对易感者没有影响, 但是对于现实社会微博网络中的信息传播, 必须要考虑到信息免疫者对信息易感者的影响.

微博网络中互动性非常强, 其中信息易感者和信息免疫者双方群体人数的变化与他们相互接触是有关系的, 信息易感者在与信息免疫者互动后会以一定概率转变为信息免疫者, 一部分信息易感者在了解事实真相后转变为信息免疫者.

信息易感者与信息免疫者在对话题的讨论中可使信息易感者直接转变为信息免疫者, 即直接免疫. 在交互中表示 s_i 与 r_j 接触, s_i 会以一定概率变为 r_i, r_j 不变, 即 T 中第 ③ 种情况 $(s_i, r_j) \rightarrow (r_i, r_j)$.

在本文提出的改进 SIR 微博信息传播模型微分方程中用 $n\langle k \rangle S(t)R(t)$ 这一项来表示直接免疫. $n\langle k \rangle S(t)R(t)$ 在 微博网络中代表未感染节点与免疫节点交互后, 未感染节点转换为免疫节点的那一部分.

直接免疫性质:

性质 2.1 $n\langle k \rangle S(t)R(t)$ 会减少未感染节点的密度, 在标准 SIR 微博信息传播模型微分方程组中第一个式子中要减去这一项;

性质 2.2 $n\langle k \rangle S(t)R(t)$ 会加快免疫节点的密度, 在标准 SIR 微博信息传播模型微分方程组中第三个式子中要加上这一项.

考虑到直接免疫的情况, 本文提出的改进的 SIR 微博信息传播模型微分方程如下:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\lambda\langle k \rangle I(t)S(t) - n\langle k \rangle S(t)R(t) \\ \frac{dI(t)}{dt} &= \lambda\langle k \rangle I(t)S(t) - \beta\langle k \rangle I(t) \\ \frac{dR(t)}{dt} &= \beta\langle k \rangle I(t) + n\langle k \rangle S(t)R(t) \end{aligned} \right\} (2)$$

在式(2)微分方程组中: 第一个方程表示未感染

节点密度的变化率; 第二个方程表示传播节点密度的变化率; 第三个方程表示免疫节点密度的变化率.

微博信息传播模型微分方程中的 $\lambda\langle k \rangle, \beta\langle k \rangle$ 及 $n\langle k \rangle$ 无法直接求得, 本文根据原始数据, 利用最小二乘法, 使得拟合数据点与原始数据点的距离平方和最小, 从而计算求出微博信息传播模型微分方程的转化概率.

初始条件: 初始时刻信息在 微博网络中传播, 绝大部分节点为未感染节点, 传播节点很少, 没有免疫节点. 假设节点随机均匀分布在 整个微博网络中, 从而可得改进 SIR 微博信息传播模型初始条件:

$$\begin{aligned} S(0) &\approx 1; \\ I(0) &\approx 0; \\ R(0) &\approx 0. \end{aligned}$$

稳定状态: 随着信息在 微博网络中的传播, $S(t)$ 会随着时间的推移逐渐减少到 0; $I(t)$ 随着时间的推移会达到一个峰值, 然后再逐渐减少到 0; $R(t)$ 会随着时间的推移逐渐逼近数值 1, 从而可得改进 SIR 微博信息传播模型稳定状态:

$$\begin{aligned} \lim_{t \rightarrow \infty} S(t) &= \lim_{t \rightarrow \infty} I(t) = 0; \\ \lim_{t \rightarrow \infty} R(t) &= 1. \end{aligned}$$

局部稳定解: 对于改进的 SIR 微博信息传播模型式(2)微分方程组, $R(t) = 1 - S(t) - I(t)$, 所以考虑前两个式子, 令 $\frac{dS(t)}{dt} = 0$ 并且 $\frac{dI(t)}{dt} = 0$, 得到式(2)的局部稳定解(平衡点) $(s, i) = (1, 0), (0, 0)$ 或 $(\frac{\beta}{\lambda}, \frac{n(\lambda - \beta)}{(n - \lambda)\lambda})$. 其中, $(s, i) = (1, 0)$ 和 $(s, i) = (0, 0)$ 为改进 SIR 传播模型初始状态和稳定状态, $(\frac{\beta}{\lambda}, \frac{n(\lambda - \beta)}{(n - \lambda)\lambda})$ 为改进 SIR 微博信息传播模型的局部平衡点.

由以上初始条件和稳定状态, 依照改进 SIR 微博信息传播模型, 在 微博网站中信息的传播过程和规律如下:

- (I) 在某个话题信息开始传播的初始时刻, 一小部分节点为传播节点, 其他所有节点为未感染节点;
- (II) 在每个时刻 t 之内, 传播节点向所有节点传播信息, 未感染节点以一定概率转变成传播节点;
- (III) 在每个时刻 t 之内, 某些传播节点对信息不再感兴趣, 以一定概率转变成免疫节点;
- (IV) 考虑直接免疫的情况, 在每个时刻 t 之内, 如果未感染节点与免疫节点连接, 未感染节点会

以一定概率跳过传播状态直接转变成免疫节点；

(V) 随着时间推移,最终网络中所有的节点都转变成免疫节点.

3 实验

不仅规模庞大的微博数据不容易取得,而且真实的微博数据操纵性也比较差,文献[26]指出,在现阶段利用计算机上的网络模型产生数据并进行仿真是研究网络信息传播的重要方法.本文利用复杂网络建模工具建立了一个小世界网络来模拟某个话题信息在微博网络上的传播.

根据前文定义中微博网络信息的传播规则,设定总节点数为 N ,初始时刻设定 m 个传播节点,0 个免疫节点,则有 $N-m$ 个未感染节点.根据传播规则,每个时刻三种节点数目都会相应变化,记录每个时刻的未感染节点、传播节点和免疫节点的数目,进而可以得到各时刻三种节点的密度,即 $S(t)$, $I(t)$, $R(t)$ 的原始数据.

表 1 是时刻 t 从 1 到 16 的某个话题每个时刻 $S(t)$, $I(t)$, $R(t)$ 的数据.

表 1 原始数据
Tab. 1 Initial data

t	$S(t)$	$I(t)$	$R(t)$
1	0.900 2	0.099 8	0
2	0.794 7	0.165 0	0.040 3
3	0.645 2	0.245 7	0.109 1
4	0.484 5	0.322 3	0.193 2
5	0.336 9	0.362 0	0.301 1
6	0.218 7	0.356 1	0.425 2
7	0.151 5	0.327 4	0.521 1
8	0.090 1	0.284 9	0.625 0
9	0.072 2	0.256 7	0.671 1
10	0.040 6	0.195 9	0.763 4
11	0.039 1	0.171 3	0.789 6
12	0.017 3	0.119 5	0.863 2
13	0.005 0	0.109 5	0.885 5
14	0.002 1	0.066 2	0.931 7
15	0.005 2	0.068 5	0.926 2
16	0.007 6	0.037 2	0.955 2

图 1 是根据表 1 数据绘制的曲线,分别表示 S , I , R 的密度随时间变化的情况.

已有文献利用标准 SIR 微博信息传播模型来进行微博信息的研究^[25],并得出标准 SIR 微博信息传播模型符合微博上的信息传播规律的结论.本文将对标准 SIR 微博信息传播模型和提出的改进 SIR

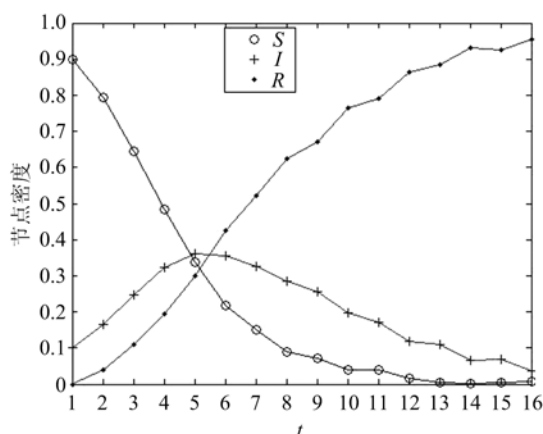


图 1 原始数据

Fig. 1 Initial data

微博信息传播模型进行实验对比.

首先进行微博信息传播规律实验,观察改进 SIR 微博信息传播模型是否比标准 SIR 微博信息传播模型更符合微博信息的传播规则;其次进行微博信息的预测实验,验证改进 SIR 微博信息传播模型是否比标准 SIR 微博信息传播模型更准确地预测微博信息的传播.

本文采用均方误差(MSE)和相关系数(R)这两个评价标准对两个微博传播模型进行对比评价.均方误差体现了原始数据与拟合数据之间数值的差异;相关系数反映了原始数据和拟合数据在变化趋势上的差异,取值范围在 $[-1, 1]$ 之间.均方误差越小,并且相关系数越大,则说明实验拟合效果越好,反之则不好.

均方误差计算公式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (u_i - v_i)^2.$$

相关系数计算公式如下:

$$R = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^n (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^n (v_i - \bar{v})^2}}.$$

其中, $U = \{u_1, u_2, \dots, u_n\}$ 为原始数据, $V = \{v_1, v_2, \dots, v_n\}$ 为拟合数据.

3.1 微博信息传播规律实验

由于标准 SIR 微博信息传播模型微分方程中的 $\lambda(k)$, $\beta(k)$ 与改进 SIR 微博信息传播模型微分方程中的 $\lambda(k)$, $\beta(k)$ 和 $n(k)$ 无法直接求得,所以根据表 1 中的数据,利用最小二乘法,使得拟合点与原始数据点的距离平方和最小,从而计算求得标准 SIR

表 2 拟合数据对比
Tab. 2 Contrast of fitting data

t	S(t)			I(t)			R(t)		
	原始	标准	改进	原始	标准	改进	原始	标准	改进
1	0.900 2	0.918 0	0.900 2	0.099 8	0.082 0	0.099 8	0	0	0
2	0.794 7	0.822 0	0.797 4	0.165 0	0.149 3	0.166 7	0.040 3	0.028 7	0.036 0
3	0.645 2	0.680 1	0.652 0	0.245 7	0.241 9	0.248 9	0.109 1	0.078 0	0.099 1
4	0.484 5	0.512 2	0.486 8	0.322 3	0.336 0	0.322 7	0.193 2	0.151 7	0.190 5
5	0.336 9	0.356 8	0.336 3	0.362 0	0.397 4	0.362 6	0.301 1	0.245 8	0.301 1
6	0.218 7	0.239 5	0.222 0	0.356 1	0.411 0	0.361 8	0.425 2	0.349 5	0.416 3
7	0.151 5	0.161 9	0.144 7	0.327 4	0.386 8	0.331 5	0.521 1	0.451 4	0.523 8
8	0.090 1	0.113 3	0.095 2	0.284 9	0.342 5	0.287 3	0.625 0	0.544 2	0.617 6
9	0.072 2	0.083 1	0.063 9	0.256 7	0.292 1	0.240 2	0.671 1	0.624 8	0.695 9
10	0.040 6	0.064 0	0.043 9	0.195 9	0.243 3	0.196 5	0.763 4	0.692 7	0.759 7
11	0.039 1	0.051 6	0.030 8	0.171 3	0.199 6	0.158 3	0.789 6	0.748 8	0.810 9
12	0.017 3	0.043 2	0.022 1	0.119 5	0.162 1	0.126 4	0.863 2	0.794 6	0.851 5
13	0.005 0	0.037 5	0.016 1	0.109 5	0.130 8	0.100 2	0.885 5	0.831 7	0.883 7
14	0.002 1	0.033 4	0.011 9	0.066 2	0.105 0	0.079 1	0.931 7	0.861 5	0.909 0
15	0.005 2	0.030 5	0.008 9	0.068 5	0.084 1	0.062 3	0.926 2	0.885 4	0.928 8
16	0.007 6	0.028 3	0.006 7	0.037 2	0.067 1	0.048 9	0.955 2	0.904 6	0.944 4

微博信息传播模型微分方程的转化概率为 $\lambda\langle k\rangle=1, \beta\langle k\rangle=0.288 2$,改进 SIR 微博信息传播模型微分方程转化概率为 $\lambda\langle k\rangle=0.892 8, \beta\langle k\rangle=0.248 7, n\langle k\rangle=0.248 3$.

表 2 是标准 SIR 微博信息传播模型微分方程和改进 SIR 微博信息传播模型微分方程的拟合数据对比。

图 2 和图 3 是根据表 2 数据绘制的曲线,分别表示标准 SIR 微博信息传播模型微分方程和改进 SIR 微博信息传播模型微分方程的 S, I, R 的密度随时间变化的情况。

表 3 和表 4 分别是根据标准 SIR 微博信息传播模型和改进的 SIR 微博信息传播模型研究微博信

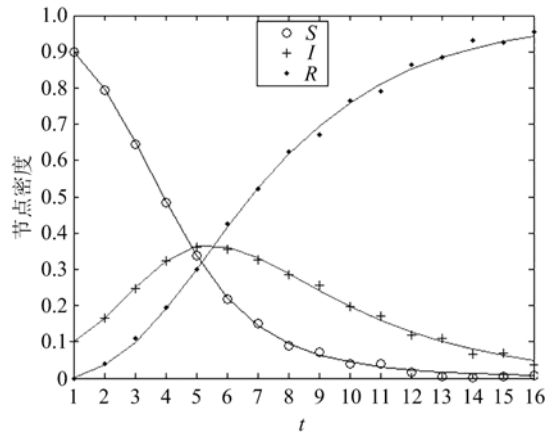


图 3 改进 SIR 节点密度

Fig. 3 Node density of the improved SIR model

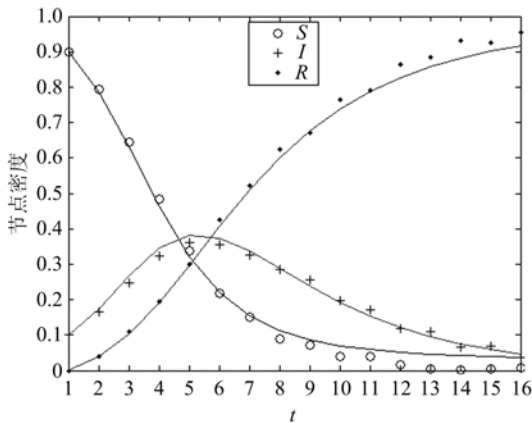


图 2 标准 SIR 节点密度

Fig. 2 Node density of of the standard SIR model

表 3 均方差对比

Tab. 3 MSE contrast of fitting data

	S(t)	I(t)	R(t)
标准 SIR	$5.832 2 \times 10^{-4}$	$1.839 0 \times 10^{-4}$	$5.208 3 \times 10^{-4}$
改进 SIR	$0.341 2 \times 10^{-4}$	$0.615 0 \times 10^{-4}$	$1.331 6 \times 10^{-4}$

表 4 相关系数对比

Tab. 4 Correlation coefficient contrast of fitting data

	S(t)	I(t)	R(t)
标准 SIR	0.998 9	0.994 7	0.999 4
改进 SIR	0.999 8	0.997 3	0.999 4

息传播规律得到数据的均方差和相关系数对比。

从表 3 中可以看到,改进 SIR 微博信息传播模

型 $S(t), I(t), R(t)$ 三条曲线的均方误差都比标准 SIR 微博信息传播模型的要小,同时从表 4 中可以看到,改进 SIR 微博信息传播模型的每条曲线相关系数比标准 SIR 微博信息传播模型的要大,说明改进 SIR 微博信息传播模型比标准 SIR 微博信息传播模型能更有效地反映微博网络中的信息传播规律.

3.2 微博信息预测实验

下面分别利用标准 SIR 微博信息传播模型和基于微博网络的改进 SIR 微博信息传播模型,使用历史数据对未来的信息传播情况进行预测.

选取表 1 中 t 为 5 之前的数据作为历史数据,进行最小二乘法拟合,得出标准 SIR 微博信息传播

模型微分方程的转化概率为 $\lambda\langle k\rangle=0.9838, \beta\langle k\rangle=0.3055$,改进 SIR 微博信息传播模型微分方程的转化概率为 $\lambda\langle k\rangle=0.9335, \beta\langle k\rangle=0.2744, n\langle k\rangle=0.1451$.

根据以上转化概率,表 5 为标准 SIR 微博信息传播模型微分方程和改进 SIR 微博信息传播模型微分方程的拟合数据对比,其中 t 为 6 之后的数据为预测数据.

图 4 和图 5 是根据表 5 数据绘制的曲线,分别表示标准 SIR 微博信息传播模型微分方程和改进 SIR 微博信息传播模型微分方程的 S, I, R 密度随时间变化的情况.

表 5 预测数据对比

Tab. 5 Contrast of prediction data

t	$S(t)$			$I(t)$			$R(t)$		
	原始	标准	改进	原始	标准	改进	原始	标准	改进
1	0.900 2	0.900 2	0.900 2	0.099 8	0.099 8	0.099 8	0	0	0
2	0.794 7	0.789 9	0.793 8	0.165 0	0.169 5	0.167 9	0.040 3	0.040 6	0.038 3
3	0.645 2	0.641 8	0.646 8	0.245 7	0.253 1	0.250 5	0.109 1	0.105 0	0.102 6
4	0.484 5	0.482 1	0.483 6	0.322 3	0.324 0	0.322 7	0.193 2	0.193 9	0.193 7
5	0.336 9	0.343 6	0.338 3	0.362 0	0.357 3	0.359 1	0.301 1	0.299 1	0.302 7
6	0.218 7	0.241 9	0.229 5	0.356 1	0.350 1	0.354 6	0.425 2	0.408 0	0.415 9
7	0.151 5	0.174 1	0.156 1	0.327 4	0.315 7	0.321 9	0.521 1	0.510 2	0.522 1
8	0.090 1	0.130 4	0.108 6	0.284 9	0.269 8	0.276 4	0.625 0	0.599 8	0.615 0
9	0.072 2	0.102 4	0.078 0	0.256 7	0.222 7	0.228 9	0.671 1	0.674 9	0.693 1
10	0.040 6	0.084 1	0.057 9	0.195 9	0.179 7	0.185 3	0.763 4	0.736 3	0.756 8
11	0.039 1	0.071 8	0.044 3	0.171 3	0.142 9	0.147 6	0.789 6	0.785 4	0.808 1
12	0.017 3	0.063 3	0.034 7	0.119 5	0.112 5	0.116 4	0.863 2	0.824 2	0.848 9
13	0.005 0	0.057 4	0.027 8	0.109 5	0.087 9	0.091 0	0.885 5	0.854 7	0.881 1
14	0.002 1	0.053 2	0.022 7	0.066 2	0.068 4	0.070 8	0.931 7	0.878 4	0.906 5
15	0.005 2	0.050 1	0.018 7	0.068 5	0.053 0	0.054 9	0.926 2	0.896 9	0.926 4
16	0.007 6	0.047 9	0.015 6	0.037 2	0.041 0	0.042 4	0.955 2	0.911 2	0.942 0

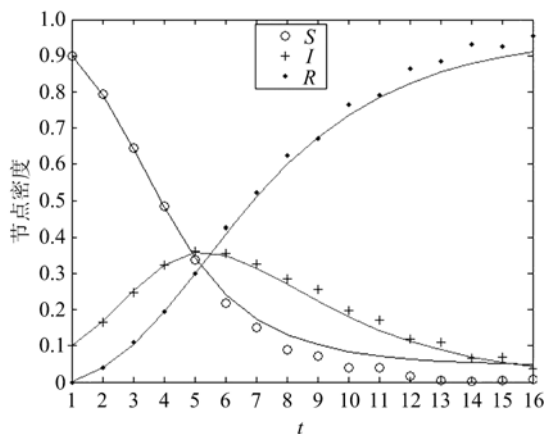


图 4 标准 SIR 预测

Fig. 4 Prediction of the standard SIR model

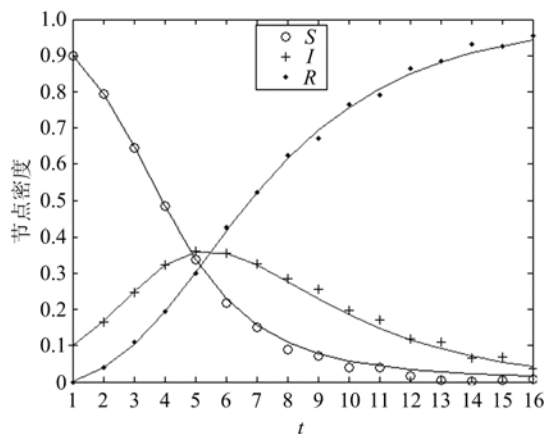


图 5 改进 SIR 预测

Fig. 5 Prediction of the improved SIR model

表 6 和表 7 分别是根据 t 为 5 之前的历史数据,进行预测得出的 t 为 6 之后的预测数据的均方误差和相关系数对比。

表 6 预测数据均方误差对比

Tab. 6 MSE contrast of prediction data

	$S(t)$	$I(t)$	$R(t)$
标准 SIR	16×10^{-4}	3.0941×10^{-4}	9.0284×10^{-4}
改进 SIR	2.1170×10^{-4}	1.9423×10^{-4}	1.8984×10^{-4}

表 7 预测数据相关系数对比

Tab. 7 Correlation coefficient contrast of prediction data

	$S(t)$	$I(t)$	$R(t)$
标准 SIR	0.996 0	0.994 7	0.996 9
改进 SIR	0.996 1	0.995 3	0.997 0

从表 6 的均方误差对比中可以看到,改进 SIR 微博信息传播模型中 $S(t)$, $I(t)$, $R(t)$ 每个都比标准 SIR 微博信息传播模型的均方误差要小,同时从表 7 的相关系数对比中可以看到,改进 SIR 微博信息传播模型 $S(t)$, $I(t)$, $R(t)$ 三条曲线的相关系数比标准 SIR 微博信息模型的要大,表明改进 SIR 微博信息传播模型比标准 SIR 微博信息传播模型能更准确地预测微博网络中信息的传播。

4 结论

本文针对微博网络信息的传播进行了研究,考虑到微博网络中信息传播直接免疫的情况,在标准 SIR 微博信息传播模型的基础上提出了改进 SIR 微博信息传播模型.基于该模型,研究了微博信息传播过程中的变化规律和预测方法。

研究表明,在微博网络中,由于考虑到直接免疫,改进 SIR 微博信息传播模型比标准 SIR 微博信息传播模型能更好地描述信息传播规律,能反映出信息在微博网络中的传播特性.同时在微博网络中,改进 SIR 微博信息传播模型比标准 SIR 微博信息传播模型能更准确地预测信息的传播趋势,可以为网络中舆情预测和分析提供参考依据。

本文提出的改进 SIR 微博信息传播模型没有考虑用户间的个体差异和微博信息的可信度,进一步研究会主要包括以下两方面:首先,在本文提出的改进 SIR 微博信息传播模型中融入微博信息可信度进行实验;其次,考虑微博用户个体差异和偏好,在改进 SIR 微博信息传播模型中加入参数进行验证。

参考文献 (References)

- [1] Miller G. Social scientists wade into the tweet stream [J]. Science, 2011, 333(6051): 1 814-1 815.
- [2] Newman M E J, Forrest S, Balthrop J. Email networks and the spread of computer viruses [J]. Physical Review E, 2002, 66(3): 035101(1-4).
- [3] 倪顺江, 翁文国, 范维澄. 具有局部结构的无标度网络中传染病传播机制研究 [J]. 物理学报, 2009, 58(6): 3 707-3 713.
- [4] Moreno Y, Nekovee M, Pacheco A F. Dynamics of rumor spreading in complex networks [J]. Physical Review E, 2004, 69(6): 066130(1-8).
- [5] Zanette D H. Dynamics of rumor propagation on small-world networks [J]. Physical Review E, 2002, 65(4): 041908(1-9).
- [6] Yan G, Fu Z Q, Ren J, et al. Collective synchronization induced by epidemic dynamics on complex networks with communities [J]. Physical Review E, 2007, 75(1): 016108(1-5).
- [7] Zhao H, Gao Z Y. Modular effects on epidemic dynamics in small-world networks [J]. EuroPhysics Letters, 2007, 79(3): 38002(1-5).
- [8] 仇慎伟, 王开, 刘茜, 等. 基于交通流量的病毒扩散动力学研究 [J]. 物理学报, 2012, 61(15): 150201(1-7).
- [9] 巩永旺, 宋玉蓉, 蒋国平. 移动环境下网络病毒传播模型及其稳定性研究 [J]. 物理学报, 2012, 61(11): 110205(1-9).
- [10] 鲁延玲, 蒋国平, 宋玉蓉. 自适应网络中病毒传播的稳定性和分岔行为研究 [J]. 物理学报, 2013, 62(13): 130202(1-9).
- [11] 李睿琪, 唐明, 许伯铭. 多关系网络上的流行病传播动力学研究 [J]. 物理学报, 2013, 62(16): 168903(1-7).
- [12] 王亚奇, 蒋国平. 复杂网络中考虑不完全免疫的病毒传播研究 [J]. 物理学报, 2010, 59(10): 6 734-6 743.
- [13] Xia Chengyi, Liu Zhongxin, Chen Zengqiang, et al. Behavior of epidemic spreading with distributed infection rate on complex networks [J]. Journal of University of Science and Technology of China, 2008, 38(12): 1 444-1 450. 夏承遗, 刘忠信, 陈增强, 等. 复杂网络上具有分布式传染速率的流行病传播行为 [J]. 中国科学技术大学学报, 2010, 38(12): 1 444-1 450.
- [14] Sznajd-Weron K, Sznajd J. Opinion evolution in closed community [J]. International Journal of Modern Physics C, 2000, 11(6): 1 157-1 165.
- [15] Deffuant G, Amblard F, Weisbuch G, et al. How can extremism prevail? A study based on the relative agreement interaction model [EB/OL]. <http://jasss.soc.surrey.ac.uk/5/4/1.html>.

(下转第 598 页)