

基于个体局部交互作用的舆情传播模型研究*

郭强

(上海理工大学管理学院, 上海 200093)

摘要: 考虑社会网络中的个体局部交互作用, 提出一种新的舆情传播模型, 细致地考虑了个体意见接收概率对于舆情传播的影响。假设系统中每个个体的意见由两种状态组成, 如同意或反对。系统中存在两组持有不同意见的人群, 他们分别对自己的意见深信不疑, 从不改变。第三组人群在单位时间内随机选择一个邻居进行交流, 并以某个概率接受该邻居的意见, 解析结果和数值模拟显示并且最终状态时, 持有两种观点人群的比例只与初始状态中持不同意见的两组人群的比例有关, 而与接收概率无关, 说明舆情传播中初始人群的规模非常重要。

关键词: 舆情传播; 社会网络分析; 随机网络

中图分类号: TP393.08

文献标志码: A

文章编号: 1001-3695(2012)11-4085-02

doi:10.3969/j.issn.1001-3695.2012.11.021

Research on opinion spreading model based on local individual interaction

GUO Qiang

(Business School, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: By taking into account individuals' interaction behaviors in social networks, this paper presented a new opinion spreading model based on the possibility that accepting neighbors' opinions. The model supposes there were two kinds of opinions in the systems: agree and disagree, and there were two kinds of persons hold their unchangeable opinions. The third kind of person discussed with a random selected neighbor, and accepted his/her opinion with a possibility. The simulation and theoretical results show that, in the final state, the percentage of the persons with different opinions is constant, which depends on the initial state and irrelevant with the accepted possibility. This work indicates that the initial state is very important for the opinion spreading.

Key words: opinion spreading; social network analysis; random networks

互联网的迅猛发展为人们相互交流提供了新的平台, 人们不仅可以利用网络浏览信息, 而且可以发布信息^[1], 因此信息的传播与扩散途径由经典的口口相传转变为快速的网络传播^[2]。然而互联网上用户的社会网络用户是如何交互的? 用户的交流模式对舆情传播具有什么样的影响? 这些问题都没有得到很好的回答。本文考虑用户的局部交互作用, 利用复杂网络理论提出了一种新的舆情传播模型。实际生活中有些人的意见不会轻易受到朋友或环境的影响, 而有的人的意见则很容易受到环境的影响。考虑人们对意见的坚守程度和局部的交互作用, 考察了初始人群比例和交互概率对舆情传播的影响, 解析结果和数值模拟显示这两个因素都会影响社会网络中的舆情传播。

经典的舆情传播中都把个体抽象成节点, 个体通过与其他人的交互作用构成网络。根据意见值的不同, 这些模型可以大体分为离散意见型^[3]和连续意见型^[4,5]。离散意见型模型假设系统中仅存在两种意见, 这两种意见可以用来表示支持或反对、同意或不同意, 等等。在每个时间步模型中任意选取一对节点进行交互。二维格子网络上的实验发现, 系统中所有用户的意见最终会趋于一致^[3]。然而模型并没有给出意见最终一致的原因。连续型模型假设不是每个人都清晰地表达自己

的观点, 例如有的人会认为支持或反对都可以, 于是该类模型用 $[0, 1]$ 中的随机数表示用户的意见。此类模型中最为经典的就是 Deffuant 模型^[4]和 Hegselmann-krause 模型^[5]。然而这两个工作只给出了模拟结果, 缺乏理论解析工作。Wu 和 Huberman^[6]提出了基于网络结构的舆情传播模型, 简称 WH 模型, 他们发现, 系统经过足够长的时间总会达成共识。聂哲等人^[7]通过引进接收概率提出了改进的 WH 模型, 他们假设当系统中两个节点进行交互时, 一方以一定概率接受对方的意见。潘新等人^[8]发现存在一个特定的概率可以使得系统中持不同意见人的比例达到一个稳定值, 然而该工作没有考虑初始意见人群比例对于结果的影响。本文综合考虑初始人群比例和交互概率, 提出了一种新的舆情传播模型。该模型假设存在两种人群: 一种坚持意见, 另一种可以与邻居交互。随机在坚持意见人群中赋予两种不同意见的处置, 观察系统的初始状态和交互概率对舆情传播的影响。

1 模型描述

由于社会结构的改变速度相对于舆论的传播速度要慢很多, 本文假设舆情传播在短时间内其网络结构是静态的。首先给定一特定网络结构, 其中节点表示个人, 节点之间的边表示

收稿日期: 2011-12-27; 修回日期: 2012-05-22 基金项目: 国家自然科学基金资助项目(10905052, 70901010, 71071098, 71171136); 上海市科研创新基金资助项目(11ZZ135, 11YZ110); 上海市智能信息处理重点实验室开放基金资助项目(HPL-2010-006); 上海市重点学科资助项目(S30501, S30504); 上海市青年科技启明星计划(A类)资助项目(11QA1404500); 国家教育部科研创新重点项目(211057)

作者简介: 郭强(1975-), 女, 辽宁大连人, 副教授, 博士, 主要研究方向为舆情传播、社会网络分析(liujg004@ustc.edu.cn).

人与人之间的关联关系。系统参数定义如表 1 所示。

表 1 变量定义及其意义

N	系统的节点数
k	节点度
b_k	度为 k 持观点 -1 且不变的人数比例
w_k	度为 k 持观点 +1 且不变的人数比例
n_k	度为 k 的节点数
$p_k = n_k/N$	度分布
m	系统中观点为 +1 的人数
m_k	度为 k 且观点为 +1 的人数
$q = m/N$	观点为 +1 的人占总人数比例
$q_k = m_k/N$	度 k 、观点为 +1 的人占总人数比例

2 舆情传播模型

假设系统中只存在两种观点,分别用 +1 和 -1 表示。 X 表示持有观点 +1 且其观点不会改变; Y 表示持有观点 -1 且其观点也保持不变; Z 表示观点可以根据邻居的观点以一定概率变化的人群。

模型的演化规则如下:在每个时间步内,随机选择一个节点 A ,并随机选择其一个邻居 B 。如果 A 节点属于集合 Z ,则 A 节点以概率 λ 接受其邻居 B 的观点,其中 $\lambda \in (0, 1]$ 。不失一般性,假设 A 节点持有观点 -1 且度为 k ,根据 A 节点与 B 节点观点的不同,系统的更新可以归纳为以下两种情况:a) 如果 B 节点持有观点 -1,则 A 节点的观点都不改变;b) 如果 B 节点持有观点 +1,则 A 的观点以概率 λ 由 -1 变为 +1。 A 节点由 -1 变为 +1 的概率可用式(1)表示。

$$P_{w \rightarrow b}(u) = \lambda p_k (1 - q_k - w_k) \frac{\sum_j p_j q_j}{\sum_j p_j} \quad (1)$$

其中: p_k 、 $(1 - q_k - w_k)$ 分别表示 A 属于集合 Z 、度为 k 且观点为 -1 的概率。相应地, A 节点的观点由 +1 变成 -1 的概率为

$$P_{b \rightarrow w}(u) = \lambda p_k (q_k - b_k) \frac{\sum_j p_j (1 - q_j)}{\sum_j p_j} \quad (2)$$

如果定义式(3)为 q_k 的加权平均

$$\langle q \rangle = \frac{\sum_j p_j q_j}{\sum_j p_j} \quad (3)$$

则式(1)和(2)可以写为

$$\begin{aligned} P_{w \rightarrow b}(u) &= \lambda p_k (1 - q_k - w_k) \langle q \rangle \\ P_{b \rightarrow w}(u) &= \lambda p_k (q_k - b_k) (1 - \langle q \rangle) \end{aligned} \quad (4)$$

为了考察系统最终的演化状态,考察随时间变化系统中观点为 +1 的个数 m_k 和其占所有人数的比例 q_k 。在时间间隔 $(t, t + dt)$ 内, m_k 的变化可以通过如下公式得到。

$$\Delta m_k = \begin{cases} +1 & \text{以概率 } \lambda n_k p_k (1 - q_k - w_k) \langle q \rangle \\ -1 & \text{以概率 } \lambda n_k p_k (q_k - b_k) (1 - \langle q \rangle) \\ 0 & \text{其他} \end{cases}$$

其中: $n_k = N p_k$, 则 Δm_k 的期望值可以用式(5)度量。

$$E[\Delta m_k] = \lambda n_k (1 - q_k - w_k) \langle q \rangle - \lambda n_k (q_k - b_k) (1 - \langle q \rangle) = \lambda n_k (\langle q \rangle - q_k + b_k (1 - \langle q \rangle) - w_k \langle q \rangle) \quad (5)$$

Δm_k 的二阶矩为

$$E[(\Delta m_k)^2] = \lambda n_k (1 - q_k - w_k) \langle q \rangle + \lambda n_k (q_k - b_k) (1 - \langle q \rangle)$$

因此, Δm_k 的方差为

$$\begin{aligned} \text{var}[\Delta m_k] &= E[(\Delta m_k)^2] - (E[\Delta m_k])^2 = \\ &= \lambda n_k (1 - q_k - w_k) \langle q \rangle + \\ &= \lambda n_k (q_k - b_k) (1 - \langle q \rangle) + o(\lambda n_k) \end{aligned}$$

由定义 $q_k = \frac{m_k}{n_k}$,可以得到系统中持有观点 +1 人数变化期望值和方差。

$$E[\Delta q_k] = \frac{1}{n_k} E[\Delta m_k] = \lambda (\langle q \rangle - q_k)$$

$$\text{var}[\Delta q_k] = \frac{1}{n_k^2} \text{var}[\Delta m_k] = \frac{\lambda}{n_k} \sigma_k^2$$

其中: $\sigma_k^2 = (1 - q_k - w_k) \langle q \rangle + (q_k - b_k) (1 - \langle q \rangle)$ 。注意到 Δq_k 的增加步长为 $1/n_k$,当系统规模非常大时,该变化是非常小的,因此由上面两个公式可以近似地得到 q_k 在单位时间内的变化。

$$dq_k = \lambda (\langle q \rangle - q_k + b_k (1 - \langle q \rangle) - w_k \langle q \rangle) dt + \sqrt{\frac{1}{n_k}} \sigma_k dB_t^{(k)}$$

其中: $B_t^{(k)}$ 表示 k 个独立的布朗运动粒子。公式两侧分别进行加权平均可以得到

$$d\langle q \rangle = \lambda [b_k (1 - \langle q \rangle) - w_k \langle q \rangle] dt + \langle \sqrt{\frac{\lambda}{n_k}} \sigma_k dB_t^{(k)} \rangle$$

由于布朗运动的平均近似可以忽略不计^[8],因此有

$$\frac{d\langle q \rangle}{dt} = \lambda [b_k (1 - \langle q \rangle) - w_k \langle q \rangle]$$

在系统的平衡状态,方程的右侧等于 0,因此有

$$\langle b \rangle (1 - \langle q_\infty \rangle) - \langle w \rangle \langle q_\infty \rangle = 0$$

则有

$$\langle q_\infty \rangle = \frac{\langle b \rangle}{\langle b \rangle + \langle w \rangle}$$

即系统的平衡状态中,两种观点比例关系由初始状态属于 X 和 Y 集合的人数比例决定。

下面考察什么状态下系统中 $E[\Delta q_k]$ 等于零。在具有周期边界的格子网络中,所有节点具有相同的度,因此 $\langle q \rangle = q_k$,则 $E[\Delta q_k] = 0$ 。对于非格子系统,需要考察参数 λ 的影响。假设参数 λ 依赖于节点的度信息,即 $\lambda = \lambda(k)$,则式(5)可写为

$$E[\Delta q_k] = \frac{1}{n_k} E[\Delta m_k] =$$

$$\lambda(k) (\langle q \rangle - q_k + b_k (1 - \langle q \rangle) - w_k \langle q \rangle) \quad (6)$$

如果 $E[\Delta q_k] = 0$,那么有

$$\langle q \rangle = \frac{\sum_k \lambda(k) (q_k - b_k)}{\sum_k \lambda(k) (1 - b_k - w_k)} \quad (7)$$

由式(3),取 $\hat{p}_i = 1 - b_i - w_i, \hat{q}_j = \frac{(q_j - b_j)}{(1 - b_j - w_j)}$,可以发现

$$\lambda(k) = c k \hat{p}_k \quad (8)$$

是式(6)的一个特解。 \hat{p}_i 对应着 Z 集合中度为 i 的人群规模, \hat{q}_j 对应着度为 j 且不属于 X 集合的人群规模。因此,当 λ 满足式(8),且 \hat{p}_i, \hat{q}_j 满足式(7)条件的系统中的人数比例都将保持在一个稳定水平。对于度分布比较复杂的系统, $\lambda(k)$ 非常难以确定,因此本文只在二维格子网络上进行数值模拟,数值模拟的结果与解析结果基本吻合。

3 数值实验

生成 10×10 的具有周期边界的格子网络(格子节点的边界节点与另一侧边界的节点也进行连接),节 (下转第 4112 页)

模型的问题,采用变量选择方法在模型确定之前进行变量选择,选择适当的输入进行模型确定,可以有效地改善模型精度。鉴于传统变量选择不能有效选择得到对复杂非线性化工过程模型的输出变量有很好解释能力的输入变量,本文分别从变量选择的搜索策略和变量选择准则两个方面进行考虑,提出基于虚假最近邻点 Gamma 检验准则的变量选择方法。该方法借鉴 FNN 思想,能够快速全面地对所有变量进行遍历搜索。通过分析可知,采用 FNN 进行搜索,在搜索计算上较其他搜索策略有更快的搜索能力。在 FNN 对变量进行快速、全面搜索的基础上,采用有输出变量监督的 GT 准则来确定最终的模型输入变量,可以选择得到对模型输出有很好解释作用的输入变量,而没有输出变量监督的相似度选择准则所选择输入变量对输出变量的解释作用有限。通过对构造模型和实际模型的仿真研究,结果表明该方法对复杂非线性化工过程模型进行变量选择取得了很好的效果。本文变量选择方法对输入/输出变量是连续数据的模型均适用,下一步可以对其他系统或领域的模型进行变量选择的研究。但是,通过对算法进行时间复杂度分析可知,该算法的时间复杂度取决于 GT 准则,采用 GT 准则对大样本数据进行变量选择会对算法的时间复杂度造成不良影响,因此对大样本进行变量选择,更加有效的有监督变量选择的准则应被采用。

参考文献:

- [1] 王心哲,韩敏. 基于变量选择的转炉炼钢终点预报模型[J]. 控制与决策,2010,25(10):1589-1592.
- [2] PACHECO J, CASADO S, NUNEZ L. A variable selection method based on Tabu search for logistic regression models[J]. *European Journal of Operational Research*,2009,199(2):506-511.
- [3] 王艳玲,李龙澍,胡哲. 群体智能优化算法[J]. 计算机技术与发展,2008,18(8):114-117.

- [4] BLANCHET F G, LEGENDRE P, BORCARD D. Forward selection of explanatory variables[J]. *Ecology*,2008,89(9):2623-2632.
- [5] CHAN K Y, KWONG C H, DILLON T S, *et al.* Reducing overfitting in manufacturing process modeling using backward elimination based genetic programming[J]. *Applied Soft Computing*,2011,11(2):2623-2632.
- [6] 梁朝林,沈本贤,刘纪昌,等. 用延迟焦化逐步回归法模型预测焦化产物的分布[J]. 东华大学学报:自然科学版,2009,35(2):185-191.
- [7] ABARBANEL H D I, KENNEL M B. Local false nearest neighbors and dynamical dimensions from observed chaotic data[J]. *Physical Review E*,1993,47(5):3057-3068.
- [8] 李太福,易军,苏盈盈,等. 基于 KCCA 虚假邻点判别的非线性变量选择[J]. 仪器仪表学报,2012,33(1):213-22.
- [9] 李太福,易军,苏盈盈,等. 基于特征子空间虚假邻点判别的软传感器模型变量选择[J]. 机械工程学报,2011,47(12):7-12.
- [10] STEFANSSON A, KONCAR N, JONES A J. A note on the Gamma test[J]. *Neural Computing and Applications*,1997,5(3):131-133.
- [11] NOORI R, KARBASSI A R, MOGHADDAMNIA A, *et al.* Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction[J]. *Journal of Hydrology*,2011,401(3-4):177-189.
- [12] JAAFAR W Z W, HAN D. Variable selection using the Gamma test forward and backward selections[J]. *Journal of Hydrologic Engineering*,2012,17(1):182-191.
- [13] 李德详. 基于 Gamma test 的最小二乘支持向量机参数在线优化方法[D]. 大连:大连理工大学,2010.
- [14] 任启伟,陈洋波. 基于 Gamma test 的非线性降雨径流回归模型研究[J]. 水文,2010,30(1):39-43.

(上接第 4086 页)点度为 8。设 10% 的节点属于 X 集合,20% 的节点属于 Y 集合,在 Z 集合中随机选取 20% 的节点,设其初始观点为 +1,其余节点设为 0。每次实验最大时间步设置为 10^6 次。表 2 给出了不同接收概率情况下系统的最终状态分布,同时显示了不同参数情况下系统最终状态中观点为 +1 所占的人数比例,所有数据经过了 10 000 次平均得到。

表 2 不同接收概率情况下系统的最终状态分布

λ	0.1	0.3	0.5	0.7	1.0
q	0.501 3	0.501 9	0.495 1	0.505 4	0.498 7

4 结束语

考虑到用户的初始意见和交互概率,本文提出了一种新的舆情传播模型。考虑到实际生活中有些人的意见非常坚定,而有些人则是不会轻易受到邻居的意见左右的,本文假定模型中有些用户的意见不随环境变化,而有些则可以和邻居进行交互从而改变自己的意见。通过对模型进行解析分析和数值模拟,发现格子网络中持有各种观点的人数比例与初始状态人群所占的比例有关,而与交流概率无关;对特定的网络存在特殊的参数也可以使得系统达到平衡状态。下一步工作将着重解决系统的收敛时间以及与收敛时间相关因素等问题。

参考文献:

- [1] CASTELLANO C, FORTUNATO S, LORETO V. Statistical physics of social dynamics [J]. *Reviews of Modern Physics*,2009,81(12):591-646.
- [2] 潘新. 基于复杂网络的舆情传播研究[D]. 大连:大连理工大学,2010.
- [3] SZNAID-WERON K, WERON R. A simple model of price formation [J]. *International Journal of Modern Physics C*,2002,13(1):115-123.
- [4] DEFFUANT G, NEAU D, AMELAND F, *et al.* Mixing beliefs among interacting agents [J]. *Advances in Complex Systems*,2000,3(1-4):87-98.
- [5] HEGSELMANN R, KRAUSE U, COMPUT E. Opinion dynamics driven by various ways of averaging[J]. *Computational Economics*,2005,25(4):381-405.
- [6] WU Fang, HUBERMAN B A. Social structure and opinion formation [R]. Palo Alto, CA:HP Labs,2006.
- [7] 聂哲,李粤平,温晓军,等. 个体相互影响的网络舆情演变模型[J]. 计算机工程与应用,2009,45(14):220-222.
- [8] 潘新,邓贵仕,佟斌. 基于社会网络的舆情传播模型构建与分析[J]. 运筹与管理,2011,20(2):176-179.
- [9] KARATZAS I, SHREVE S E. Brownian motion and stochastic calculus [M]. 2nd ed. [S. l.]:Springer,1997.