

分级意见领袖下的微博信息传播与数学建模

黄贤英 杨林枫 刘小洋

(重庆理工大学计算机科学与工程学院 重庆 400054)

摘 要 为了有效地分析意见领袖在微博在线社交网络传播中所起的作用及微博信息传播的生命周期,提出了一种 OLL 分级意见领袖模型。首先,爬取微博数据,并进行数据统计分析;其次,将传播力构建为转发数、活跃度与粉丝量 3 个因素相关的函数,并建立一种基于层次分析的权重计算方法;最后,利用计算的传播力并结合 OLL 模型,对意见领袖传播作用及微博生命周期进行了实验分析。仿真结果表明:意见领袖在微博信息传播中有很强的传播效应,通过计算得出 OLL 模型仿真与 3 组真实数据的误差分别为 9.6%,13.4%,4.5%,从而证明了所提 OLL 模型对于分析意见领袖在微博信息传播中的生命周期是合理且有效的。

关键词 意见领袖,微博传播,传播力,微博生命周期

中图分类号 TP393 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.11.041

Information Dissemination and Mathematical Modeling of Microblog under Graded Opinion Leader

HUANG Xian-ying YANG Lin-feng LIU Xiao-yang

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract In order to analyze the role of opinion leaders in microblog online social network dissemination and the life cycle of microblog information dissemination, an OLL graded opinion leader model was proposed. Firstly, the microblog data are crawled, and statistical analysis is carried out. Secondly, the dissemination power is constructed as a function related to three factors such as the number of forwarding, the degree of activity and the number of fans, and then a weight calculation method based on hierarchy analysis is established. Finally, the computed dissemination power and OLL model are used to analyze the dissemination function of opinion leader and the life cycle of microblog. The simulation results show that the dissemination effect of opinion leader in the microblog information dissemination is very strong. The error is calculated between the OLL model and three real data sets of errors are 9.6%, 13.4% and 4.5% respectively. It is proved that the proposed OLL model is reasonable and effective for analyzing the life cycle of opinion leader in microblog information dissemination.

Keywords Opinion leader, Microblog communication, Communication power, Microblogging life cycle

1 引言

意见领袖于 20 世纪 40 年代由传播学者拉扎斯菲尔德提出^[1]。意见领袖是对他人施加影响的“活跃分子”,在大众传播过程中起到媒介的作用,属于二级传播。意见领袖在获取信息后,将信息扩散给大众,使信息传播形成二级传播。近年来,微博的出现改变了人们获取最新信息的途径,同时微博以超乎人们想象的速度快速发展。相比于传统媒介信息传播,社会网络中的信息传播呈现出许多新的特点,比如用户创造内容、“多对多”传播模式、传播实时性和快速性等。在微博中,对意见领袖的分析受很多因素影响,如被转发数、被评论

数、活跃天数、粉丝数等,利用这些因素可以研究意见领袖的传播力,进而进行建模研究。

国内外研究者在此问题上研究出了许多成果。文献[1]在 Leader Rank 算法的基础上,加入用户之间的情感倾向、用户活跃程度,提出了改进的 Leader Rank 算法;文献[2]基于用户之间的关系特征,给出了相邻节点和非相邻节点的亲密度评估算法,意见领袖节点被更精确和有效地识别;文献[3]提出了一种基于用户属性与覆盖范围的意见领袖研究方法;文献[4]提出了一个选择网络社区意见领袖的新框架,该框架通过评估用户之间的总信任值(TTV)来挖掘意见领袖;文献[5]提出了一种基于改进拓扑势的意见领袖挖掘算法,以

到稿日期:2017-08-28 返修日期:2017-10-18 本文受国家自然科学基金项目(17XXW004),教育部人文社会科学研究青年基金项目(16YJC860010),重庆市教育委员会人文社会科学研究一般项目(17SKG144),2017 年度重庆市高校网络舆情与思想动态研究咨政中心开放课题(KFJJ2017024)资助。

黄贤英(1967—),女,硕士,教授,CCF 会员,主要研究领域为计算机应用、传播模型等;杨林枫(1993—),男,硕士生,CCF 会员,主要研究领域为社交网络信息传播模型,E-mail:lfyang@2016.cqut.edu.cn(通信作者);刘小洋(1980—),男,博士后,副教授,硕士生导师,主要研究领域为信息传播模型、无线通信与计算机应用等。

挖掘更准确的意见领袖;文献[6]提出一种针对特定话题传播的意见领袖检测方法。文献[1-6]主要利用各种算法进行意见领袖的识别,研究初期的意见领袖识别工作是非常重要的。识别结束之后,需要对意见领袖的传播作用进行分析。文献[7]提出了一种基于消息传播的微博意见领袖影响力建模与测量分析方法;文献[8]从多个维度获得主要特征,并提出了一个级联扩散模型来预测传播过程;文献[9]利用水动力学模型来描述社会网络中信息的传播过程,这有助于传播特征的提取;文献[10]提出了一种综合模式,并归纳出意见领袖与电子烟的使用存在负相关关系;文献[11]通过分析得到信息通常与主题密不可分,通过主题模型分析使得信息传播中的意见领袖更加准确;文献[12]分析了意见领袖的行为,并收集了信息,根据结构方程模型得出在线意见寻求行为对客户满意度水平以及意见提供和意见传递有积极影响。文献[7-12]是对意见领袖在传播过程中所起的作用进行分析,利用数据或者算法来进行研究。在生活实际中,也应结合意见领袖进行分析研究。文献[13]提出了一个新的框架来选择在线社区的意见领袖,利用用户之间的信任关系评估其他用户之间的主要意见领袖的总信任值;文献[14]提出了一种新颖的加权方法,以根据中心性措施寻求舆论意见领袖;文献[15]调查了 Facebook 上的数据,增加了社交网站的意义,同时增加了意见领袖在社会政治交流中的作用;文献[16]提出了“潜在意见领袖”的概念,设计了分析算法来识别“潜在意见领袖”,结果说明使用“潜在意见领袖”的识别方法更为简单;文献[17]结合传统 SIR 模型与意见领袖进行信息传播分析,研究了意见领袖如何影响微博信息的传播;文献[18]按社会网络中的结构洞位置、中心度位置和边缘位置对意见领袖角色进行识别。文献[13-18]主要研究意见领袖在日常生活中的应用以及相关扩展内容,可见意见领袖在信息传播中发挥着重要作用。

近年来,通过对在线社交网络传播中意见领袖的研究发现,在线网络传播中的意见领袖具有重要意义。因微博及各类社交网络平台的信息传播量及传播力度大,故意见领袖在此类平台中对传播起到极大的推动作用,但少有通过模型来预测微博累积转发量以及微博生命周期的研究。本文将对基于分级意见领袖的微博信息传播进行建模分析研究。

2 数据采集

随着 Web2.0 的快速发展,国内外各大社交软件都存在非常大的信息流动。作为国内最大的微型博客网站——新浪微博,供人们每天进行交流。微博信息传播主要通过意见领袖及普通大众的转发来进行。本文将对新浪微博的热搜微博进行研究,并将意见领袖传播力及其转发微博后的信息传播量特征作为主要研究点,从而得到微博生命周期,并将其用于度量微博的传播力度。

微博作为社交软件,存在私密性,利用网络爬虫的方法采集微博热搜中的数据时,需要破除其限制。若在一个时间点某一账号访问的页面量爆炸,则新浪将对其封号,故选取 3 个账号完成爬取过程,爬取时间从 2017 年 3 月 20 日到 2017 年 5 月 30 日每天任选两个时间点,每个时间点取 8 个热门话题,去重后共选取 408 个热门话题。每个话题涉及的微博数

目参差不齐,其中也有普通大众为了蹭热度而发布的与话题毫无关系的微博,故每个话题只取评论转发较高的 1~3 篇进行分析,最后共选取到 772 篇微博,其中涉及的意见领袖共 1170 人。下面将对采集到的数据进行建模和分析。

3 意见领袖影响因素分析

意见领袖是在线社交网络传播中为他人提供信息的人,是传播过程中的媒介,相当于口语中的知名人士。近年来,微博已经成为社交网络中最重要的传播途径,大众通过微博可以第一时间获取最新新闻,而各路人士也通过微博向大众公布相关信息。通过研究微博可以进行舆情分析,进一步对微博中的意见领袖进行研究可有效分析意见领袖对信息传播的影响。由于微博发展迅速,意见领袖群体中人数众多,其对微博有自己的见解,且在转发微博之后的传播力度也不一样,有时会出现极端现象,因此要对意见领袖进行分级,本文将采用基于层次分析的权重计算方法。

3.1 意见领袖数据分析

将新浪微博中拥有 300 万以上粉丝的用户定为意见领袖,忽略某些特定条件下的传播;忽略粉丝数量低于 300 万的用户传播力。首先从微博风云榜中获取粉丝数量高于 300 万的 1870 名用户进行分析,并根据这些用户的粉丝量对其进行排序和分类,如图 1 所示。其次分析部分意见领袖的关注列表,在关注列表中读出意见领袖的数量(见图 2),从而分析出在意见领袖发布信息后,有多少意见领袖能看到,看到的人越多,形成多级传播的概率就越大。以上两个因素对意见领袖的多级传播都有影响。

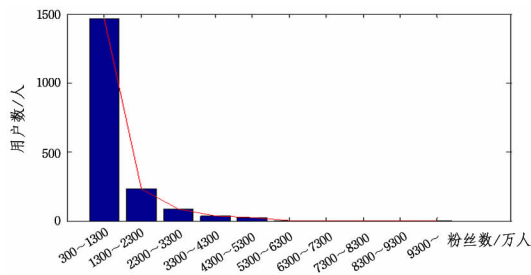


图1 微博名人粉丝分布图

Fig. 1 Distribution map of microblog celebrity fans

通过图 1 可得,意见领袖的粉丝数量大多集中在 300 万~3300 万。下一步,从图 1 所示的 10 类意见领袖中分别抽取 10%的意见领袖用户,并对其关注的意见领袖数量进行分析(人数的 10%不足 1 的取 1)。

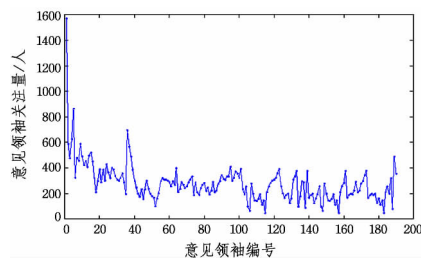


图2 关注列表中意见领袖的数量

Fig. 2 Number of opinion leaders in list of concerns

从图 2 可以看出,这些意见领袖关注的列表中是意见领

袖用户的人数为 45~1572 不等,其中最大值与最小值是两个极端。综合 190 名意见领袖,得到其关注的平均意见领袖数为 277。可以说明,信息在传播过程中被多个意见领袖转发,或经意见领袖转发后有很大的概率形成多级转发,从而使得信息传播达到最大化。

3.2 意见领袖分级

信息发布后,通常会有一级意见领袖转发,进而由二级或三级等进行二级转发,但不乏出现二级或三级意见领袖直接转发,三级或四级意见领袖二级转发,甚至转发之后没有意见领袖进行二级转发的情况。影响意见领袖传播力的指标主要有活跃度和影响力,其中活跃度是指发微博或转发微博的频率以及活跃天数;影响力是指被转发或者评论的数目及粉丝数量。本文将用权重计算方法得到的结果对意见领袖进行分级,选取的因素为平均转发量、粉丝数、意见领袖活跃度。

意见领袖用户传播力的计算如式(1)所示:

$$F_{\text{propagation}} = \omega_1 V_{\text{fans}} + \omega_2 V_{\text{forward}} + \omega_3 V_{\text{activity}} \quad (1)$$

其中, ω 代表权重值, V 代表 3 个因素的值。由于一级与四级意见领袖之间的转发量、粉丝量差值较大,为了便于比较,需要对各个影响因素数据进行归一化,使用式(2)把数据转换至 $[0,1]$ 区间:

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

下一步将确定各因素在影响力计算中的权值。归一化处理后,各属性值在评价中的重要性不能确定,因此需要确定属性权重。采用 Saaty 的 10 级重要性等级表和两两比较的判断矩阵专家法计算各个属性指标的相对重要性。比如,若评价人认为意见领袖粉丝数量 x_1 与微博转发数量 x_2 相比意见领袖活跃度 x_3 相当重要,则可取 $b_{12} = b_{13} = 5$;若明显重要则可以取 7;若介于相当重要和明显重要之间,则可以取 6;若介于略微重要与相当重要之间,则可以取 4。依次对各个指标进行比较后运用层次分析法,则得到属性相对重要性判断矩阵。最后通过一致性检验来确定各因素权重数据表(见表 1)。

表 1 意见领袖传播力影响因素权重

Table 1 Weight of influential factors on dissemination power of opinion leaders

因素	粉丝数量	平均转发数	活跃度
权重	0.2	0.6	0.2

注:一致性检验:CR=0.008<0.1

根据权重计算的结果可以对意见领袖进行分级,从而更好地分析微博的生命周期。表 2 为意见领袖分级表(此处说明微博转发过程中若无特殊情况,转发量高于 10 万的微博都默认有水军刷转发量,故本文制定微博实际转发量超过 10 万的都将其定为 10 万)。

表 2 意见领袖分级表

Table 2 Grading of opinion leaders

因素	粉丝数量/万	平均转发数/万	活跃度
一级意见领袖	3000~9000	4~10	10~20
二级意见领袖	1000~3000	2~4	5~10
三级意见领袖	500~1000	0.5~2	3~5
四级意见领袖	300~400	0.1~0.5	0~3

意见领袖权重范围情况如表 3 所列。

表 3 意见领袖权重范围表

Table 3 Weight range of opinion leader weight

等级	一级	二级	三级	四级
范围	0.398~1	0.181~0.398	0.049~0.181	0~0.049

综合表 2 和表 3 得到分级后的意见领袖,下一步将对基于分级意见领袖的微博生命周期预测模型进行研究与分析。

4 意见领袖信息传播 OLL 数学模型

4.1 微博生命周期分析

首先对一条微博的生命周期所经历的阶段进行描述与分析,如图 3 所示。

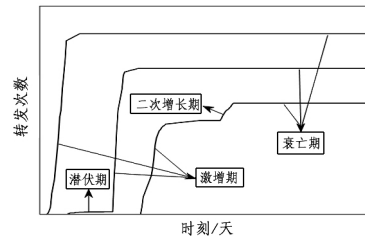


图 3 微博的生命周期

Fig. 3 Lifecycle of microblog

从图 3 中可以看出,随着时间的增加,总转发次数在慢慢趋于平稳。图中展示出了一条微博所经历的 4 个阶段。激增期是传播快速增长的阶段,此阶段转发数会大幅增加;衰亡期信息不再发生传播,转发数基本保持不变;潜伏期是指某些微博在发送之后被人们转发的数量增长很缓慢的时段,过了该时期转发量开始激增;而二次增长期就是在一条微博已经开始衰亡或进入激增之后的潜伏期后又开始增长的时段。通过对意见领袖进行建模可以有效地分析一条微博的生命周期。

4.2 基于贝塞尔函数的 OLL 模型的建立

贝塞尔函数(Bessel Functions)是贝塞尔方程的解,是数学上的一类特殊函数的总称。这类方程的解无法用初等函数表示,只能运用自动控制理论中的相平面法进行定性分析。本文基于贝塞尔函数进行模型建立,将模型命名为意见领袖等级(Opinion Leadership Level)模型,简称为 OLL 模型。

一般贝塞尔函数是常微分方程的标准解函数,如式(3)所示:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - a^2)y = 0 \quad (3)$$

由于式(3)无法用初等函数表示,需用两个独立的函数来表示,因此引出贝塞尔函数,并使用第一类和第二类贝塞尔函数来表示贝塞尔的标准解函数,如式(4)所示:

$$y(x) = C_1 J_a(x) + C_2 Y_a(x) \quad (4)$$

第一类贝塞尔函数用 $J_a(x)$ 表示,如式(5)所示:

$$J_a(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m! \Gamma(m+a+1)} \left(\frac{x}{2}\right)^{2m+a} \quad (5)$$

若 a 是整数,则式(4)和式(5)之间满足如下关系:

$$J_{-a}(x) = (-1)^a J_a(x) \quad (6)$$

其中, a 在贝塞尔函数中代表方程的阶数,在本文数学模型中用意见领袖级数 l 及转发时间与原微博发送时间的差值 t_{forward} 来表示,结果如式(7)所示:

$$\alpha = \frac{\sqrt{l} + t_{\text{forward}}}{2} \quad (7)$$

在这种情况下,两函数已不满足线性无关条件。为寻找微分方程与 $J_a(x)$ 线性无关的另一解,需要定义第二类贝塞尔函数。

第二类贝塞尔函数,又称诺依曼函数(Neumann Function),通常用 $Y_a(x)$ 表示,它与第一类贝塞尔函数 $J_a(x)$ 存在如下关系:

$$Y_a(x) = \frac{J_a(x) \cos(\alpha\pi) - J_{-a}(x)}{\sin(\alpha\pi)} \quad (8)$$

与 J 函数类似, Y 函数也存在如下关系:

$$Y_{-a}(x) = (-1)^a Y_a(x) \quad (9)$$

第三类贝塞尔函数,又称汉克尔函数(Hankel Function),其作为贝塞尔方程的另外一对重要的线性无关解 $H_a^{(1)}(x)$ 和 $H_a^{(2)}(x)$,分别定义为:

$$H_a^{(1)}(x) = J_a(x) + iY_a(x) \quad (10)$$

$$H_a^{(2)}(x) = J_a(x) - iY_a(x) \quad (11)$$

由上述推导可得第一类和第二类修正贝塞尔函数,分别用 $I_a(x)$ 和 $K_a(x)$ 来表示,其定义为:

$$I_a(x) = i^{-a} J_a(ix) \quad (12)$$

$$K_a(x) = \frac{\pi}{2} \frac{I_{-a}(x) - I_a(x)}{\sin(\alpha\pi)} = \frac{\pi}{2} i^{a+1} H_a^{(1)}(ix) \quad (13)$$

采用修正贝塞尔函数来衡量意见领袖对一条微博生命周期的影响,其由第一类、第二类和第三类贝塞尔函数及第一类修正贝塞尔函数表示。此模型的函数表达式的自变量为时刻,用 t 表示;因变量用转发次数 N_f 表示。结合分析得预测的表达式如下:

$$N_f = -\frac{\pi}{2} i^{a+1} H_a^{(1)}(it) + 10F_{\text{propagation}} \quad (14)$$

其中, $H_a^{(1)}(it)$ 由以上推导的第三类贝塞尔函数给出,主要用来控制激增期的趋势; i 为虚数单位即 $\sqrt{-1}$, α 由式(7)得到, $F_{\text{propagation}}$ 为意见领袖分级时计算权重得到的传播力值,其系数作为放大系数,由于准备工作把数据进行了归一化,为使预测数量合理,加入放大系数。该模型函数所计算出的值被用来预测转发量大小,整条曲线可以有效地对不同级别的意见领袖转发微博之后的生命周期进行预测。

对模型分析后需进行误差计算,以判断模型的有效性,如式(15)所示:

$$\Delta = \frac{|f_{\text{model}} - f_{\text{real}}|}{f_{\text{real}}} \quad (15)$$

根据式(15)的计算误差,在仿真之后,可以用其对真实数据与 OLL 模型进行误差分析,并基于误差值的大小来判断模型是否具有实际意义。

5 仿真分析

模型建立完成后,需进行数值仿真。从采集的数据中选出 3 个不同类型且均对相应受众造成较大传播影响的话题。首先对意见领袖粉丝数及活跃度进行分析;其次从转发者中寻找各级意见领袖代表,并计算其传播力 $F_{\text{propagation}}$,在计算时各因素均采集近一年内的数据;最后利用 Matlab 工具进行真实数据与 OLL 模型的仿真对比分析。

5.1 仿真参数的设置

从热门话题中抽取 3 个话题,抽取原则是话题关注人多、

影响的民众广且要选择不同类别的话题,用于综合分析模型的适用性。从 3 个话题中再选取相应代表性高的微博,所选参数如表 4—表 6 所列。

表 4 蔡康永电影的微博参数

Table 4 Microblog parameters of Cai Kangyong's film

意见领袖	杨幂	王嘉尔	范冰冰	张大大
传播力 $F_{\text{propagation}}$	0.959	0.376	0.208	0.049
粉丝量/万	7203	496	5033	420
活跃度	20	15	4	19
实际转发量/万	12.35	3.66	1.70	0.54

表 5 “乐天萨德”事件的微博参数

Table 5 Microblog parameters of "Lotte Sade" event

意见领袖	最神奇的视频	我的损友是个极品	Happy 张江
传播力 $F_{\text{propagation}}$	0.199	0.099	0.037
粉丝量/万	754	2407	911
活跃度	6	3	1
实际转发量/万	2.17	0.88	0.30

表 6 “比特币勒索病毒”的微博参数

Table 6 Microblog parameters of "bitcoin extortion virus"

意见领袖	黑客凯文	夏目家的小诗哥	游研社
传播力 $F_{\text{propagation}}$	0.260	0.053	0.007
粉丝量/万	37	170	32
活跃度	11	2	14
实际转发量/万	2.52	0.59	0.07

表 4 和表 5 中的转发人都为意见领袖,但表 6 中的不属于前四级意见领袖,属于低级意见领袖。分析可知,话题三是 2017 年 5 月 12 日爆发的“比特币病毒勒索”事件,因为事件爆发突然,所以大多数热门微博都是由计算机类的微博号发出,这类意见领袖平时的传播力较低,但在与“勒索病毒”相关的爆炸消息上的传播力较大。

5.2 真实数据与 OLL 模型的对比分析

根据上述 3 组参数,对每个意见领袖的粉丝数及活跃度进行分析。粉丝数量直接影响着能看到信息的人数;而活跃度越高,说明意见领袖在消息传播中越积极,并且会导致更多的人持续关注此意见领袖。

表 7 意见领袖传播力的排序

Table 7 Ranking of opinion leaders' dissemination power

意见领袖	排名	意见领袖	排名
杨幂	1	我的损友是个极品	6
王嘉尔	2	夏目家的小诗哥	7
黑客凯文	3	张大大	8
范冰冰	4	Happy 张江	9
最神奇的视频	5	游研社	10

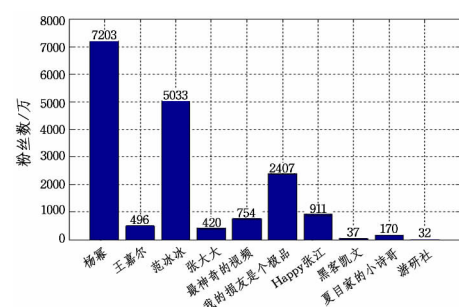


图 4 意见领袖的粉丝量

Fig. 4 Number of fans of opinion leader

根据选取的 3 个话题中涉及的 10 个意见领袖,将每个意

见领袖的粉丝量相加,得到的总人数为 17463 万,可见单取出的 10 名意见领袖就关系到 1.7 亿多人,意见领袖在信息传播过程中的力度很强。下面继续对这 10 个意见领袖的活跃度进行分析。

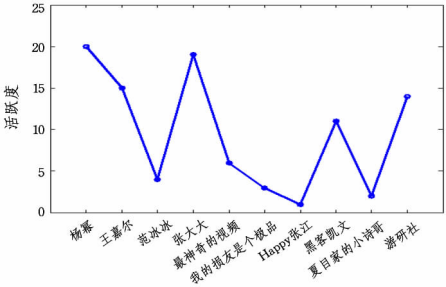


图 5 意见领袖的活跃度

Fig. 5 Activity of opinion leader

从表 7、图 4 和图 5 可以进一步看出,在微博传播过程中,并不是粉丝越多传播力 $F_{\text{propagation}}$ 就越强,也不是活跃度越高传播力就越强,但是传播力越强,粉丝数和活跃度都不会太低。从上文权重计算中可以看出,相比粉丝数和活跃度,转发数对传播力的影响最大。

从抽取的 3 个话题中选择相关微博,并对其进行真实数据与 OLL 模型的仿真。对仿真得到的结果进行分析,可以研究意见领袖在信息传播中所起作用的大小、传播力大小以及微博的生命周期,并计算其误差值以判断所提模型的合理性。图 6—图 8 这 3 个仿真图代表所取的 3 个话题中的典型微博,而图中不同的曲线代表不同的意见领袖转发微博后的生命周期,曲线的形态代表真实数据仿真与 OLL 模型仿真。

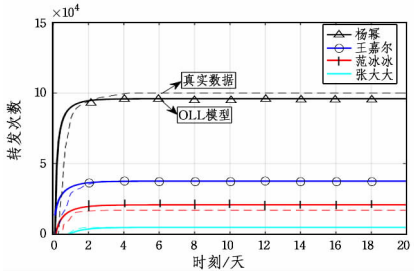


图 6 微博一的生命周期

Fig. 6 Lifecycle of first microblog

图 6 中,一级意见领袖杨幂对此微博的真实转发量大于 10 万,将其定为 10 万进行仿真。由图 6 可以分析 OLL 模型与真实数据的误差值,得出 4 个意见领袖的误差均值为 9.6%。因为微博的随机性,其误差主要由三级意见领袖引起。原微博共有 32 万转发量,单抽取的 4 个意见领袖就有 17.65 万的转发量,在总转发量中占有很大一部分比例。

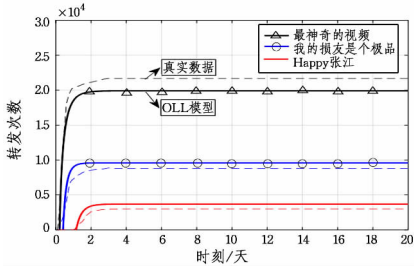


图 7 微博二的生命周期

Fig. 7 Lifecycle of second microblog

图 7 为第二个话题所抽取的微博,其发布形态不是转发,而是以原创微博的方式直接引用别人的微博,形态相当于转发。在此没有找到相关的一级意见领袖,故在二、三、四级中抽取意见领袖进行分析,最后计算得到微博二中模型仿真与真实数据的误差均值为 13.4%。造成误差的主要原因同样是微博的随机性。由此得出,在忽略随机性的前提下,意见领袖造成的转发量越极端,相对误差就越高。从图 7 中发现,原微博被引用的共有 6 万的转发量,而抽取的 3 个意见领袖就有 3.35 万的转发量,超过转发总量的一半。

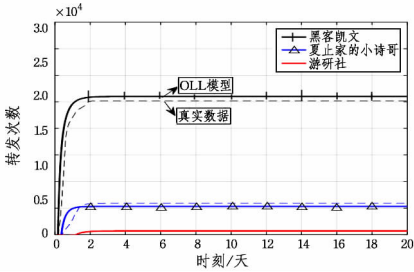


图 8 微博三的生命周期

Fig. 8 Lifecycle of third microblog

对于抽取的微博三的数据,图 8 所示黑客凯文是转发自己的原始微博并造成传播。经过计算得到其平均误差为 4.5%。此条微博的转发量主要来源于普通大众,存在的意见领袖不多。原微博的转发量有 4.5 万左右,而抽取的 3 个人所造成的转发量为 3.18 万。同样地,意见领袖在此造成了很大的传播量。

5.3 OLL 模型与传统模型比较

对微博转发量的传统预测主要是分析单位时间的转发量,本文用累积转发量进行预测,给出的仿真图更加直观,并且可以清楚呈现各个生命周期。对比当前较主流的预测方法^[7],本文在确定各因素的值时给出了明确的计算方法,然而文献^[7]在通过三因素对单位时间转发量建模之后,没有明确指出各因素值的获得方式;对于覆盖面来说,本文模型将影响力构建在传播力中,在对意见领袖分级之后覆盖面更大,而文献^[7]中的模型对于影响力较高的意见领袖的准确率较高。最后对用本文模型与当前较主流的文献^[7]中的意见领袖影响力模型进行误差对比分析。

表 8 模型的误差分析结果

Table 8 Error analysis results of model

(单位:%)

对比模型	参数 1	参数 2	参数 3
OLL 模型	9.6	13.4	4.5
意见领袖影响力模型 ^[7]	11.2	12.8	14.3

对以上 3 个话题所抽取出来的微博进行仿真分析可以得到,从话题在微博中产生到被人们所接受主要是通过意见领袖进行传播,意见领袖起到多级传播的作用。多个意见领袖在进行两次或三次转发的过程中会引入误差,因为这种情况会消耗一部分第一次就看到此消息的重复民众,所以针对式(14)会出现预测误差。通过分析分级意见领袖的微博消息传播模型可以得到,本文提出的 OLL 模型与真实数据相比,

分析微博生命周期过程中存在较高的相似度,模型误差也较低,从而证明了本文模型的有效性。

结束语 在对意见领袖进行权重计算得到其传播力值之后,根据传播力数值大小对意见领袖进行分类,并分析各级意见领袖在微博的社交网络传播中所起到的传播作用。利用贝塞尔函数对分级意见领袖传播过程进行数学建模,根据 OLL 模型与真实数据的对比,分析模型所存在的误差,以及意见领袖在传播过程中所起的作用。仿真结果证明,OLL 模型对分析微博的生命周期有较高的准确率,误差较小。下一步将对 OLL 模型进行深入研究,以减小其误差值,并降低微博的随机;同时,在分析微博之前,先对微博内容进行突发事件、日常微博的分类,然后再进行分析。

参 考 文 献

- [1] XU J M, ZHU F X, LIU S C, et al. Identifying opinion leaders by improved algorithm based on LeaderRank[J]. Computer Engineering and Applications, 2015, 51(1): 110-114. (in Chinese)
徐郡明, 朱福喜, 刘世超, 等. 改进 LeaderRank 算法的意见领袖挖掘[J]. 计算机工程与应用, 2015, 51(1): 110-114.
- [2] YANG L, QIAO Y, LIU Z, et al. Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm[J]. Soft Computing, 2016, 13(2): 1-12.
- [3] LI Y X, WANG Z K, LIU L J, et al. Micro-blog opinion leader mining method based on user attributes and coverage[J]. Application Research of Computers, 2017, 43(12): 1-7. (in Chinese)
李亚星, 王兆凯, 刘利军, 等. 基于用户属性与覆盖范围的意见领袖挖掘研究[J]. 计算机应用研究, 2017, 43(12): 1-7.
- [4] AGHDAM S M, NAVIMIPOUR N J. Opinion leaders selection in the social networks based on trust relationships propagation [J]. Karbala International Journal of Modern Science, 2016, 2(2): 88-97.
- [5] MAO T M, GUAN P, PID C. Improved Topology-potential-based Opinion Leader Mining Algorithm [J]. Computer Science, 2016, 43(6): 194-198. (in Chinese)
毛天铭, 关鹏, 皮德常. 一种改进拓扑势的意见领袖挖掘算法 [J]. 计算机科学, 2016, 43(6): 194-198.
- [6] LAN T, GUO G D. Opinion Leader Detection Method in the Communication Network of the Specific Topic [J]. Computer System Applications, 2016, 25(12): 9-15. (in Chinese)
兰天, 郭躬德. 特定话题传播网络中的意见领袖检测方法 [J]. 计算机系统应用, 2016, 25(12): 9-15.
- [7] WANG C X, GUAN X H, QIN T, et al. Modeling on Opinion Leader's Influence in Microblog Message Propagation and Its Application [J]. Journal of Software, 2015, 26(6): 1473-1485. (in Chinese)
王晨旭, 管晓宏, 秦涛, 等. 微博消息传播中意见领袖影响力建模研究 [J]. 软件学报, 2015, 26(6): 1473-1485.
- [8] WANG Y, ZHANG Z M, PENG Z S, et al. A Cascading Diffusion Prediction Model in Micro-blog Based on Multi-dimensional Features [C] // International Conference on Emerging Internet-working, Data & Web Technologies. Cham: Springer, 2017: 734-746.
- [9] HU Y, SONG J, CHEN M. Modeling for Information Diffusion in Online Social Networks via Hydrodynamics [J]. IEEE Access, 2017, PP(99): 1.
- [10] GENTINA E, KILIC D, DANCOINE P F. Distinctive role of opinion leaders in the social networks of school adolescents: an investigation of cigarette use [J]. Public Health, 2017, 144(1): 109-116.
- [11] YANG L, TIAN Y, LI J, et al. Identifying opinion leaders in social networks with topic limitation [J]. Cluster Computing, 2017, 15(3): 1-11.
- [12] NAGY A, KEMÉNY I, SZÜCS K, et al. Are Opinion Leaders More Satisfied? Results of a SEM Model about the Relationship between Opinion Leadership and Online Customer Satisfaction [J]. Society & Economy, 2017, 39(1): 39-40.
- [13] AGHDAM S M, NAVIMIPOUR N J. Opinion leaders selection in the social networks based on trust relationships propagation [J]. Karbala International Journal of Modern Science, 2016, 2(9): 88-97.
- [14] KHAN N S, ATA M, RAJPUT Q. Identification of opinion leaders in social network [C] // International Conference on Information and Communication Technologies. IEEE, 2015: 1-6.
- [15] KARLSEN R. Followers are opinion leaders: The role of people in the flow of political communication on and beyond social networking sites [J]. European Journal of Communication, 2015, 303(7): 263-264.
- [16] QU D C, LIAN H X, XIE H, et al. Research Methods on Recognizing Potential Opinion Leaders of Hot Event Based on Social Network [J]. Atlantis Press, 2015, 122(5): 65-72.
- [17] LIU H T, CHEN H, FANG C, et al. Research on Impacts of Opinion Leaders on Information Diffusion in Micro-blog [J]. Information Science, 2015, 65(7): 111-119.
- [18] CHEN Y, LIU X Y. Research on Opinion Leaders Recognition Based on Social Network [J]. Information Science, 2015, 94(10): 23-31.
- [19] (上接第 237 页)
- [20] WANG R, ZHOU Y, ZHAO C, et al. A hybrid flower pollination algorithm based modified randomized location for multi-threshold medical image segmentation [J]. Bio-medical materials and engineering, 2015, 26 Suppl 1(s1): S1345.
- [21] WANG H, OUYANG H B, GAO L Q. An Improved Global Particle Swarm Optimization Algorithm [J]. Control and Decision, 2016, 31(7): 1161-1168. (in Chinese)
王皓, 欧阳海滨, 高立群. 一种改进的全局粒子群优化算法 [J]. 控制与决策, 2016, 31(7): 1161-1168.
- [22] MENG X B, GAO X Z, LIU Y, et al. A novel bat algorithm with habitat selection and Doppler effect in echoes for optimization [J]. Expert Systems with Applications, 2015, 42(17/18): 6350-6364.