



基于观点动力学的在线点评研究

林自展¹, 肖井华¹, 周金连², 吴 晔^{2*}

(1. 北京邮电大学理学院 北京 海淀区 100876; 2. 北京师范大学新闻传播学院 北京 海淀区 100875)

【摘要】互联网时代的快速发展使得消费者能够在网络平台上通过浏览他人的在线评论收集产品的相关信息以及消费体验, 从而对消费者的观点产生重要影响。消费者群体表现出大众化、个性化、多元化和异质性等特点, 进一步促成了网络群体意见的多样性。为了探究在线平台群体意见的演化过程, 文章以“大众点评”为例, 在传统动力学模型的基础上对网络结构、观点阈值等进行优化改进, 构建出适用于在线点评的观点动力学模型并通过仿真模拟与实际数据相结合的方式进行分析。该方法的提出有助于提高对在线点评观点演化内在机理的深入认识。

关键词 Hegselmann-Krause 模型; 在线点评; 观点动力学; 观点演化

中图分类号 O41; TP391 **文献标志码** A **doi**:10.12178/1001-0548.2018320

An Online Review Research Based on Opinion Dynamics Model

LIN Zi-zhan¹, XIAO Jing-hua¹, ZHOU Jin-lian², and WU Ye^{2*}

(1. School of Science, Beijing University of Posts and Telecommunications Haidian Beijing 100876;

2. School of Journalism and Communication, Beijing Normal University Haidian Beijing 100875)

Abstract Due to the rapid growth of Internet, consumers obtain information about products and experiences from online rating platforms, which further influences their purchase decisions. The popularity, individuation, diversification, and heterogeneity of consumers contribute to the diversity of online review. This study examines the evolution process of opinions on the online review platform. Specifically, the network structure and confidence thresholds of traditional models are optimized based on the ground truth of Dianping via the combination of simulation and actual data. The proposed method can deepen the understanding for the evolution process of online opinion.

Key words Hegselmann-Krause model; online review; opinion dynamics; opinion evolution

在互联网时代, 每个人不仅是社会信息的消费者, 同时也是信息的生产者和观点的传播者^[1]。在线点评作为一种个体观点表达的方式, 指用户或者消费者通过线上平台直接对商品或服务评分并发布评论。这种直接的观点表达, 将作为一种反馈信息, 为下一个潜在消费者的行为决策提供参考。因此研究点评平台中群体意见分布的观点动力学机制, 既可以进一步了解个体线上点评的行为规律, 也可以为商家保持良好口碑提供参考。

目前, 关于个体在线点评观点动力学的研究主要采用多主体建模的方法对点评意见分布进行解释或预测。文献 [2] 认为个体在观影后发表影评时容易受到电影质量、其他用户评分以及自身评分习惯的影响, 基于此建立了关于电影评分的理论模型, 并使用豆瓣影片在线点评数据进行验证, 该模型可

以较好地预测影片评分的意见分布。与预测评分意见不同的是, 文献 [3] 则通过建立连续观点的动力学模型, 通过判断个体是否会选择发表评分, 从而预测在线点评的人数发展趋势。文献 [4] 则是针对电子商务环境下, 基于个体时间与精力的有限性、个人对他人观点的信任程度、发表评论的概率及现有的历史评论等影响因素, 构建出在线点评意见演化动力学模型。通过仿真模拟实验的方式, 分别讨论了上述的各个因素对最终意见群体演化结果的影响。

目前主流的观点动力学模型归为离散模型和连续观点模型两个大类。在离散模型中, 个体的观点只有赞成和反对两种情况。Sznajd 模型^[5]、Majority rule 模型^[6]和 Voter 模型^[7]是离散模型的典型代表。实际上个体的观点并不能简单地归结于 0 和 1 两种情况, 在连续模型中, 个体的观点值是一定范围内

收稿日期: 2018-11-28; 修回日期: 2019-06-04

基金项目: 国家自然科学基金(11875005, 61672108); 中央高校基本科研业务费专项资金(310422107)

作者简介: 林自展(1993-), 男, 主要从事机器学习, 大数据处理等方面的研究。

通信作者: 吴晔, 教授. E-mail: wuye@bnu.edu.cn

的连续数值, 基于有界信任的 Deffuant 模型^[8]和 Hegselmann-Krause 模型^[9]是最具代表性的两个模型。这两个模型用 $[0,1]$ 区间内的连续数值来表示个体的观点值, 只有在观点差值小于阈值的情况下, 两个个体之间才会发生意见交流, 并互相产生影响。在 Deffuant 模型中每次选择两个个体进行观点交互, 与 Deffuant 模型的不同点在于, HK 模型中的个体会受到群体中所有满足观点差值小于阈值的其他个体的观点的影响, 并以此来更新自己的观点值。目前, 关于舆论动力学的研究主要是在传统的 Deffuant 模型和 HK 模型的基础上, 从社会网络结构^[10-15]、评论文本挖掘^[16-17]、个体特征^[18]等角度进行研究。

伴随着在线点评平台的快速发展, 越来越多的学者基于观点动力学在社交网络的基础上研究用户行为特征以及舆论传播的演化过程。然而, 在线点评行为与传统的舆情演化相比存在明显的差异, 主要包括以下几个方面: 1) 在传统的舆论研究中, 人群中任意两个人都有可能发生观点交互并互相影响对方的观点。但在线点评中, 通常只能看到前人的评论, 而前人并不会受到即将发表的评论的影响。因此, 在线点评下个体间的影响是基于时间顺序的单方面影响。2) 基于社交网络的舆论演化研究, 是在复杂网络的基础上考虑节点多样性、连接多样性、网络结构复杂性等。而在在线点评模式下是多个前人对后续评论者的单向影响, 即单向网络结构。3) 社交网络下, 个体间的交流可以发生在有人际关

系连边的好友, 也可能是发生在有共同好友的两个个体之间。而在在线点评平台上, 通常是按照评论发表时间的先后顺序进行展示, 只有近期的少部分评论会对用户产生影响。

本研究基于上述在线点评与传统舆论传播的差异性, 对传统的 HK 模型进行部分针对性的修改, 并探究在线点评中所蕴含的观点动力学机制。

1 数据来源

在线点评行为常见于淘宝、京东、当当等各大电商平台, 其中大众点评是中国最大的在线点评平台之一。用户可以在大众点评平台上发布自己对商家的评论, 同时他们也能够浏览其他用户的点评辅助消费决策。

本研究选取了大众点评平台上 2012 年 12 月至 2015 年 1 月用户的历史评论数据, 共包括 3.8 万个商家, 共 330 万条历史评分记录。从中筛选出评分数量超过 500 条的店铺, 剩余 1 610 个商家, 205 万条历史评分数据。每一条评分数据由星级、口味、环境、服务组成, 均是 $[10, 20, 30, 40, 50]$ 分, 以这 4 种评分的均值作为店铺的整体评分。极端的评论会影响实际数据的真实性和准确性, 因此 50 分的极端评分在这里并不考虑, 原因是: 1) 雇佣水军模拟正常消费者的评论, 提升整体平均分, 吸引更多的人到店消费是商家常用的手段^[19]; 2) 未发表评论系统默认打分为满分 50; 这两种情况不能反映真实的用户体验。

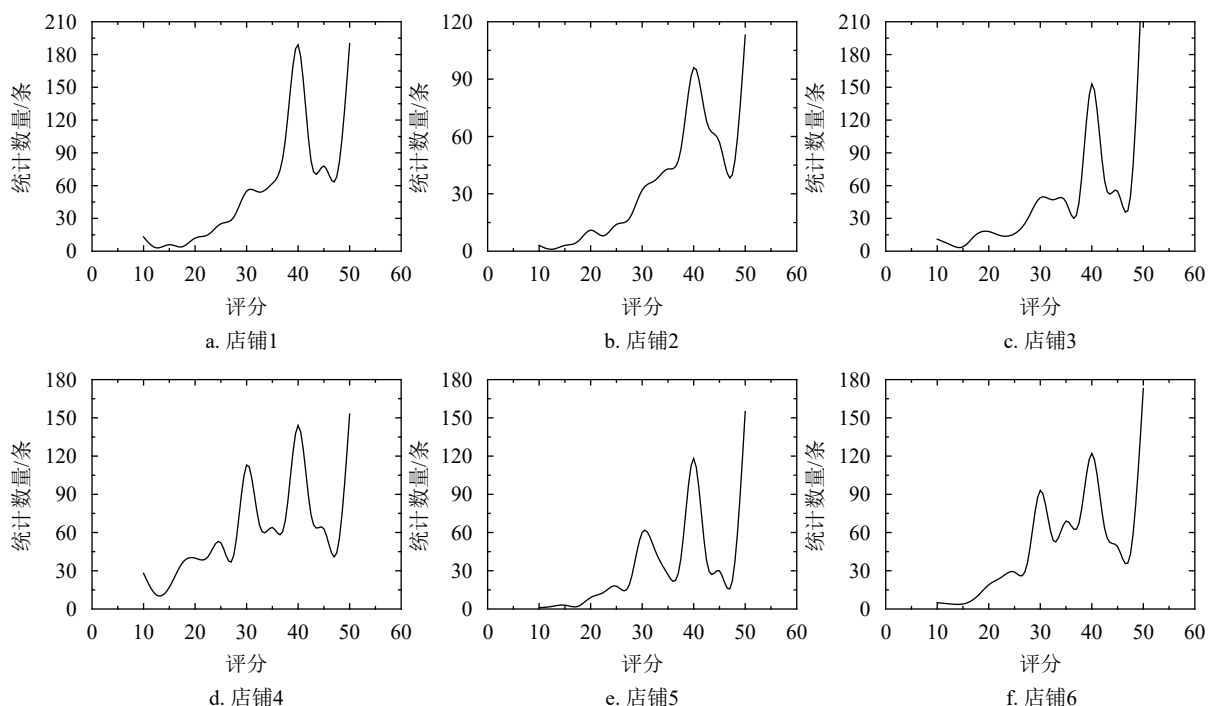


图1 所选店铺实际评分分布

通过观察 500 条以上店铺的评分分布情况, 可以将其归类为单峰模式和双峰模式两大类。单峰模式即店铺的评分呈现出单一意见群体, 双峰模式则存在两个对立的意见群体。本文分别从这 1 610 家店铺中挑选了两种模式下具有代表性的 3 个不同店铺, 实际的评分分布如图 1 所示。

2 模型构建

2.1 Hegselmann-Krause 模型

传统 HK 模型是在 Deffuant 模型的基础上演化出的一种基于有界信任的连续动力学模型, 目的是研究群体中个体观点的演化过程。假设在一个群体中, 存在 N 个个体, 用 $x_i^t \in [0, 1]$ 表示个体 i 在 t 时刻自身的观点值, 同时在 t 时刻, 个体 i 会在群体中与其他个体进行交流, 发生观点间的交互, 如果个体 j 与个体 i 的观点差值小于观点阈值 d , 即 $|x_i^t - x_j^t| < d$, 那么个体 j 的观点会对个体 i 的观点产

生影响, 反之则不会有影响。个体 i 在与其他个体发生观点交互后, 其观点值的更新如下:

$$I(i, x(t)) = \left\{ 1 \leq j \leq N \mid |x_i^t - x_j^t| \leq d \right\}$$

$$x_i^{t+1} = (1 - \mu_i) \times x_i^t + \mu_i \sum_{j \in I(i, x(t))} \omega_{ij}^t x_j^t \quad (1)$$

式中, 收敛参数 μ_i 表示个体对其他个体观点值的信任程度; 观点阈值 d 表示只有在观点阈值内的个体观点才会对个体 i 产生影响; 权重 $\omega_{ij}^t \in [0, 1]$ 表示个体 j 对个体 i 的影响程度, 且 $\sum_{j=1}^N \omega_{ij}^t = 1$ 。

2.2 模型修正

在线点评观点传播的特点与 HK 模型并不完全相同, 其传播过程如图 2 所示。基于在线点评模式观点传播的特点, 本文在 HK 原始模型的理论基础上针对以下三个部分进行了调整, 使得新模型能够适用于在线点评模式下的观点演化模式。

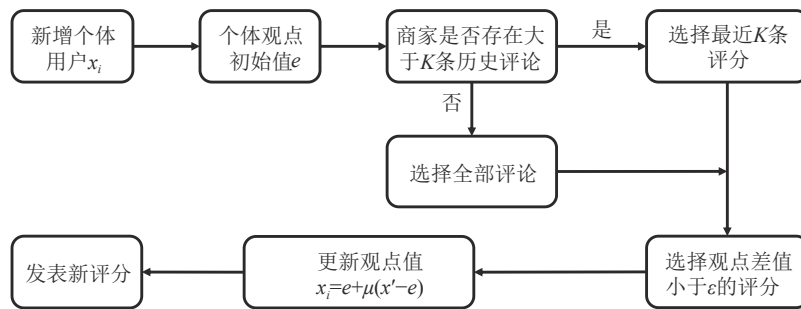


图 2 模型基本流程图

1) 在 HK 模型中, 基于社会人际关系复杂网络, 个体之间的观点交互只在有关系的个体间发生, 即只有两个节点之间存在关系连边, 且观点差值小于阈值才会发生两者间的观点交互。然而在在线点评中, 用户之间的社会网络关系并不明显, 个体间的观点交互是基于评论时间先后顺序的, 只有后发表评论的人才能看到已有的评论。因此, 本文将社会人际关系网络结构转变为基于时间先后顺序的单向网络结构。

2) 在线点评店铺中存在大量的历史评论, 且网页上每次仅只会显示几十条评论, 由于决策的时间有限, 不能看完所有的历史评论, 只有近期的评论数据会对用户观点产生直接的影响。假设最近的 K 条评论会对用户的观点有影响。

3) 不同个体之间的信任程度不一致。传统的 HK 模型通过产生一个随机数作为两个个体间的信任值。大众点评中, 用户更倾向于相信近期的评论。在观点差小于阈值 d 的前提下, 发表时间越近的评论对当前用户的评分所造成的影响程度越大。这里

用 $\omega_{ij}^t(j, K)$ 表示第 j 条评论对第 i 条评论的影响程度。

综上所述, 个体观点值的更新如式 2 所示, 其中 i 表示当前新增的评论为评论序列的第 i 条, $j \in [i - K, i - 1]$, e 表示第 i 条评论的初始观点, 服从 $[0, 1]$ 之间的均匀分布, μ 表示收敛参数。

$$x' = \sum_{j=i-K}^{i-1} \omega_{ij}(j, K) \times x_j \Big/ \sum_j \omega_{ij}(j, K)$$

$$x_i = e + \mu(x' - e) \quad (2)$$

2.3 模型仿真

根据大众点评的实际情况, 本文基于传统 HK 模型进行修正, 提出了适用于在线点评模式中的新模型, 并对分析每一个参数对最终仿真结果所造成的影响。

意见群体的形成与 K 值的大小紧密相关, 如图 3 所示。当 K 值较小时, 也就是用户阅读的评论数较少, 即只有少部分评论会影响到用户的评分, K 值越小说明用户受到其他用户的影响越小, 那么用户之间就很难形成意见群体。随着 K 值的

增大,越来越多的评论会影响到用户,那么用户就容易与周围其他用户的意见达成一致,最终形成了意见群体。从图中可以看到,当 K 值大于10时,最终的结果并不会随着影响用户的评论数量的增加而产生明显的差异。

观点差阈值 d 表示的是只有两个个体间的观点差值小于 d ,才会彼此互相影响。如图4所示,如果任意两个用户之间都无法对对方产生影响,即

当 $d=0$ 时,随着时间的推移,最终是评分区间内均匀分布的结果。当阈值 d 逐渐增大时,部分观点差较小的用户之间开始相互影响,且阈值 d 越大,两个用户之间越有可能会相互影响,向对方的意见靠近趋于一个中间值。此时会形成两个不同的意见群体,形成两种截然不同的对立意见。当 d 值足够大即 $d>0.5$ 时,观点差异较大的用户之间也会相互影响,大量用户互相交流影响,最终形成一个统一的意见。

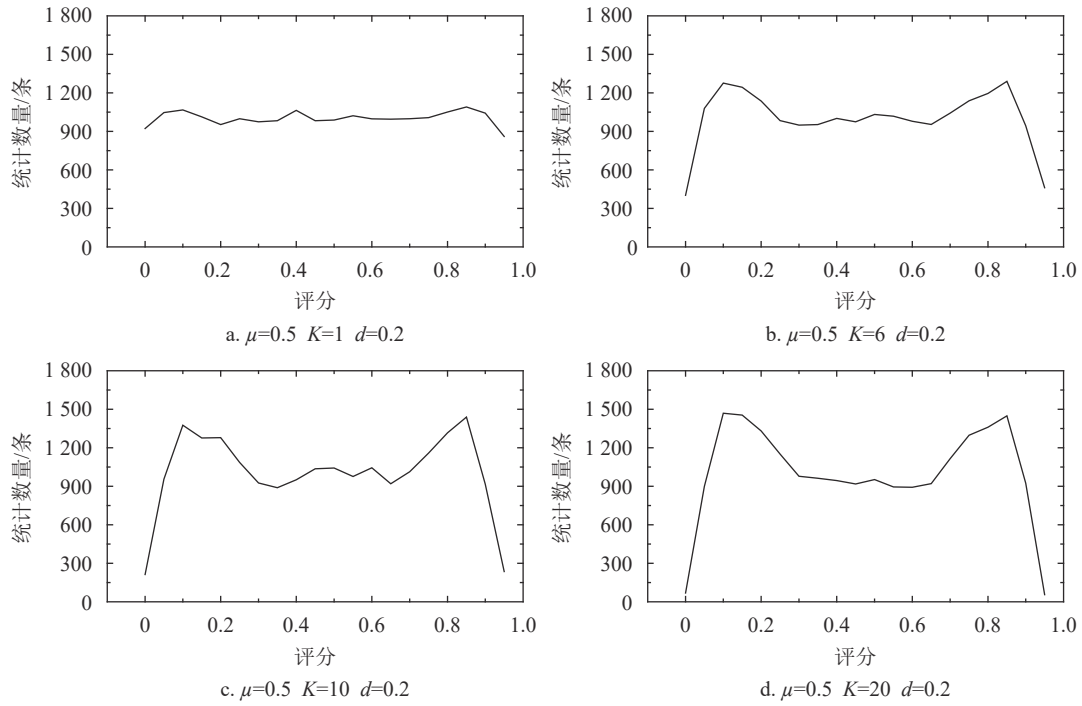


图3 K 值对意见群体形成的影响

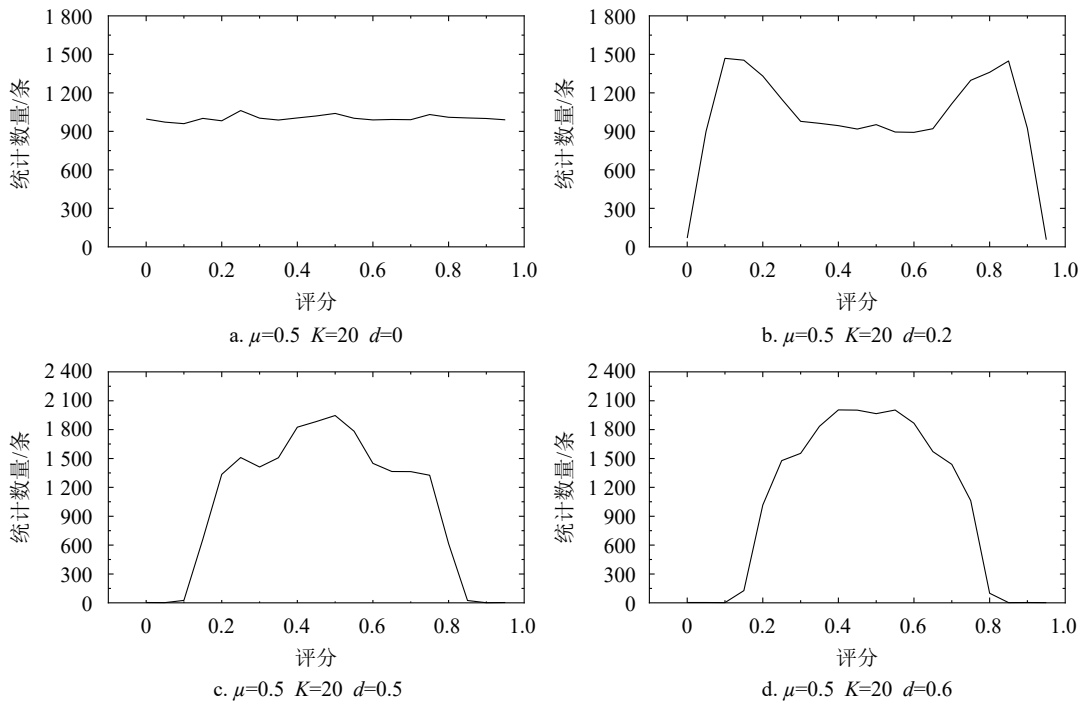


图4 观点阈值 d 对意见群体形成的影响

收敛参数 μ 表示个体受到他人观点的影响程度。如图 5 所示, μ 值较小时用户坚信自己的观点是正确的, 无论他人的观点与自身差异的大小, 都不会轻易采纳他人的观点。随着 μ 值的增大, 用户逐渐开始接受他人的观点。在受到周围其他用户的影响后, 观点差异较小的用户之间达成统一意见, 最终形成两种对立的意见群体。

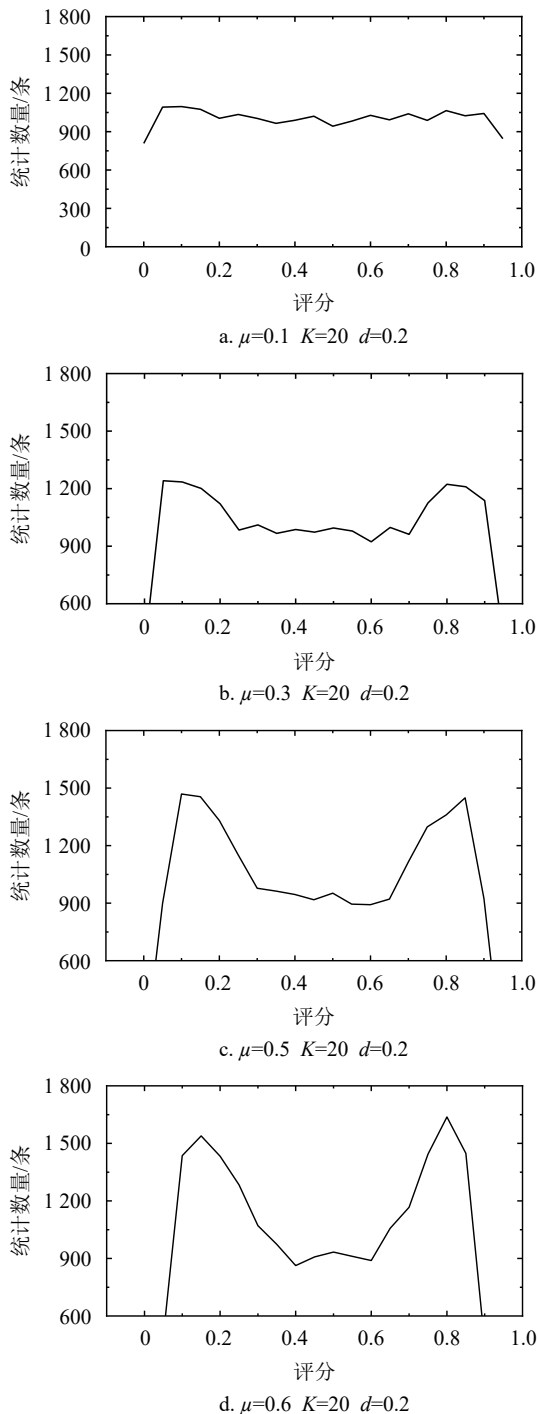


图 5 收敛参数 μ 对意见群体形成的影响

以上 3 个仿真结果表明, 个体参考评论数 K 、观点阈值 d 和收敛参数 μ 均会对在线点评的演化结

果产生重要的影响。其中, 最终形成意见群体个数是由观点阈值 d 的大小所决定的。本文在大众点评实际评分数据分布中观察到的单峰、双峰两种模式与最终的仿真结果一致。

3 结束语

以大众点评、淘宝、京东、豆瓣等为代表在线点评平台的兴起, 为消费者提供了发表个人消费体验的平台, 同时也使得其他后续的消费者能够参考他人的评论便于自身更好地做出消费决策。传统的舆论动力学模型在社会人际关系网络的基础上探究观点演化的内在机制。然而, 在线点评平台上用户之间的社交关系十分单薄, 用户之间通常不存在任何社交关系, 无法构建出有效的社会人际关系网络, 因此传统舆论动力学模型对于在线点评上的观点演化存在明显的缺陷。本文从以下两个方便针对 HK 模型进行修正: 1) 将社交关系网络替换为以时间先后为顺序的多对一的中心网络结构; 2) 将个体间的信任程度与发表评论的时间间隔因素结合。仿真结果表明, 参考评论数量 K 、观点差阈值 d 以及收敛参数 μ 三者是影响最终意见群体演化结果的主要因素。后续的研究可以针对不同点评店铺的实际情况, 更深入分析各个参数, 进一步丰富模型。

参考文献

- [1] 彭兰. 网络传播学 [M]. 北京: 中国人民大学出版社, 2009. PENG Lan. Network communication [M]. Beijing: Chinese People's Publishing House, 2009.
- [2] 由明阳, 黄飞虎. 基于观点动力学的电影在线点评预测 [J]. 现代计算机, 2016(2): 7-11. YOU Ming-yang, HUANG Fei-hu. Online rating score prediction for movie based on opinion dynamics [J]. Modern Computer, 2016(2): 7-11.
- [3] 苏炯铭, 刘宝宏, 李琦, 等. 基于观点动力学的在线点评人数预测 [J]. 计算机工程, 2014, 40(10): 155-160, 167. SU Jiong-ming, LIU Bao-hong, LI Qi, et al. Number prediction for online rating based on opinion dynamics [J]. Computer Engineering, 2014, 40(10): 155-160, 167.
- [4] WAN Y, MA B, PAN Y. Opinion evolution of online consumer reviews in the e-commerce environment [J]. Electronic Commerce Research, 2018, 18(2): 291-311.
- [5] SZNAJD-WERON K, SZNAJD J. Opinion evolution in closed community [J]. International Journal of Modern Physics C, 2000, 11(6): 1157-1165.
- [6] HOWE D A, PERCIVAL D B. Wavelet variance, Allan variance, and leakage [J]. IEEE Transactions on Instrumentation and Measurement, 1995, 44(2): 94-97.
- [7] HOLLEY R A, LIGGETT T M. Ergodic theorems for weakly interacting infinite systems and the voter model [J]. The Annals of Probability, 1975: 643-663.

- [8] DEFFUANT G, NEAU D, AMBLARD F, et al. Mixing beliefs among interacting agents[J]. *Advances in Complex Systems*, 2000, 3(1-4): 87-98.
- [9] HEGSELMANN R, KRAUSE U. Opinion dynamics and bounded confidence models, analysis, and simulation[J]. *Journal of Artificial Societies and Social Simulation*, 2002, 5(3): 1-33.
- [10] PROSKURNIKOV A V, TEMPO R. A tutorial on modeling and analysis of dynamic social networks. Part II[J]. *Annual Reviews in Control*, 2018, 45: 166-190.
- [11] MENG X F, VAN GORDER R A, PORTER M A. Opinion formation and distribution in a bounded-confidence model on various networks[J]. *Physical Review E*, 2018, 97(2): 022312.
- [12] LI D, HAN D, MA J, et al. Opinion dynamics in activity-driven networks[J]. *EPL(Europhysics Letters)*, 2018, 120(2): 28002.
- [13] PEREZ I J, CABRERIZO F J, ALONSO S, et al. On dynamic consensus processes in group decision making problems[J]. *Information Sciences*, 2018, 459: 20-35.
- [14] 苟智坚, 范明钰, 王光卫. 复杂网络中无信任边界限制的连续观点演化研究[J]. *电子科技大学学报*, 2015, 44(5): 749-756.
GOU Zhi-jian, FAN Ming-yu, WANG Guang-wei. Research of evolution of continuous opinions without bounded confidence in complex network[J]. *Journal of University of Electronic Science and Technology of China*, 2015, 44(5): 749-756.
- [15] 阚佳倩, 谢家荣, 张海峰. 社会强化效应及连边权重对网络信息传播的影响分析[J]. *电子科技大学学报*, 2014, 43(1): 21-25.
KAN Jia-qian, XIE Jia-rong, ZHANG Hai-feng. Impacts of social reinforcement and edge weight on the spreading of information in networks[J]. *Journal of University of Electronic Science and Technology of China*, 2014, 43(1): 21-25.
- [16] LEE J, PARK D H, HAN I. The effect of negative online consumer reviews on product attitude: An information processing view[J]. *Electronic Commerce Research and Applications*, 2008, 7(3): 341-352.
- [17] GHOSE A, IPEIROTIS P G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(10): 1498-1512.
- [18] 张亮, 杨闪, 李霞. 具有个体信息差异的 Deffuant-I 舆论传播模型[J]. *系统工程*, 2017(6): 82-88.
ZHANG Liang, YANG Shan, LI Xia. The Deffuant-I opinion diffusion model with differences in individual information levels[J]. *Systems Engineering*, 2017(6): 82-88.
- [19] JINDAL N, LIU B. Opinion spam and analysis[C]// *Proceedings of the 2008 International Conference on Web Search and Data Mining*. [S. l.]: ACM, 2008: 219-230.

编辑 蒋晓