

基于隐含狄利克雷分配模型的消费者在线评论 复杂网络构建及其应用

刘晓君, 那日萨, 崔雪莲

(大连理工大学系统工程研究所, 辽宁 大连, 116024)

摘要: 为研究消费者在线评论的相互关系及整体演化发展, 以隐含狄利克雷分配模型对消费者在线评论进行话题挖掘为基础, 通过 Pearson 相似度确定评论间话题关系, 构建了以评论为节点的复杂网络模型. 以京东等平台上的手机评论语料库为例, 对相应复杂网络的拓扑特性进行分析. 研究表明, 相同产品相同规模的评论网络统计特性相近; 而相同平台上同一产品不同规模的评论网络性质差异明显, 即小规模评论(产品销售初期)构成的网络是一个无标度网络, 大规模评论(产品销售中后期)构成的网络演化为一个度分布为钟形分布的小世界网络. 揭示了消费者发表在线评论行为之间相互作用, 会影响整体评论网络的拓扑性质和演化规律, 并为市场营销和管理实践提出了指导性建议.

关键词: 在线评论; 隐含狄利克雷分配模型; 复杂网络; 皮尔森相似度; 度分布

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-5781(2017)03-0305-08

doi: 10.13383/j.cnki.jse.2017.03.002

Construction and application of the complex network about consumer online reviews based on latent Dirichlet allocation model

Liu Xiaojun, Narisa Zhao, Cui Xuelian

(Systems Engineering Institute, Dalian University of Technology, Dalian 116024, China)

Abstract: In order to study the relationship and the evolution of the consumer online reviews, this paper proposes a complex network model with reviews as nodes by calculating reviews' topics with latent Dirichlet allocation model and topic similarities among reviews with Pearson similarity. Taking the mobile phone reviews as an example, analysis is undertaken and the results indicate that the statistical characteristics of the review network with the same scale of the same product are similar; as for different scales, the characteristics vary significantly. Small-scale review(the initial stage of product sales) network is a scale-free network. Large-scale review (the middle and late stage of product sales) network evolves into a small world network with bell shaped degree distribution. It reveals the interaction of online review behaviors among consumers which influences the topological properties and the evolution rules of the review network. Instructional suggestions are put forward for the marketing and management practices.

Key words: online reviews; latent Dirichlet allocation model; complex network; Pearson similarity; degree distribution

收稿日期: 2016-07-07; 修订日期: 2017-02-09.

基金项目: 国家自然科学基金资助项目(61471083); 国家自然科学基金重点资助项目(71431002); 教育部人文社科研究规划基金资助项目(14YJA630044).

1 引言

随着信息技术和服务经济的发展,网络购物已经普及,网络消费者以非结构化或半结构化的评论文本形式相互交流沟通,表达对产品和服务的观点、态度、感觉及情绪,其中蕴含着大量的营销信息和产品改进,创新的建议^[1],同时也影响着潜在消费者的心理情感和购买行为.如何有效利用这些评论信息成为当前的迫切需要,这一现实需要引发了学术界对在线评论研究的强烈关注.

目前,关于在线评论的研究主要集中于在线评论的理论与实证研究以及利用文本自动分析技术对在线评论的挖掘研究.关于在线评论的理论与实证研究分别从在线评论的诸多特征如何影响消费者^[2],消费者为何参与在线评论交流以及企业如何利用在线评论制定营销策略等方面展开^[3,4].近年来大量研究表明,消费者进行产品在线评分时倾向于先观察已有消费者评分,然后根据已有消费者评分来调整自己的反馈^[5-9].如 Moe 等^[6]从动态角度分析影响评分的因素,发现不经常参与评分的消费者评分倾向大众化,经常参与评分的消费者评分倾向于个性化,易影响产品评分演化. Lee 等^[7]把已有的评分区分为消费者朋友和陌生人发表的,研究表明,朋友的观点经常会导致评分的群集效应. Tsekouras^[10]通过实验研究发现,关注评论中情感标签的用户受评论规模影响,随评论规模增大受极端评论的影响减小.以上研究表明,消费者在线评论之间具有相互联系,已有文献主要运用理论模型或者实验研究方法而没有充分挖掘实际数据中包含的信息,有其局限性,并且都只关注于在线评论的数量特征,而在线评论的文本内容是消费者决策的重要信息来源.因此,对文本在线评论相互关系及整体演化发展做分析具有重要意义.

互联网上日益增多的用户生成内容促进了文本自动分析技术的发展,这为研究文本在线评论提供了方法.关于利用文本自动分析技术研究在线评论的文献主要集中于在线评论的情感分类^[11,12],产品的特征提取以及在线评论有用性评价^[13,14].近年来,复杂网络方法作为认识和描述复杂现实系统的有效方法,在很多领域取得成果^[15-17].复杂网络为更好地认识和描述消费者在线评论系统提供了新途径.如索琪等^[18]基于用户在社交网络上的评分数据,构建超网络理论模型,对度分布等网络拓扑结构分析发现,该网络具有“富者愈富”现象.同时,结合文本自动分析技术与网络分析方法对在线评论进行研究取得了一定进展.如姜巍等^[11]利用复杂网络思想进行在线评论有用性分析,其网络构造方法是将评论段落作为节点,对段落单元词汇采用 ESA 相关度方法确定节点关联关系,但该方法割裂了每条评论的整体性,不利于对消费者评论语义话题的挖掘和全面理解.王伟等^[19]结合文本挖掘和情感分析技术,构建了产品特征比较网络,实验发现,特征比较网络与销售排名显著相关,但此研究仅关注于评论中的比较信息有其局限性.

常用的文本自动分析方法有文本表示模型和机器学习方法.而文本表示模型更适合在线评论网络的构建.作为一种全概率主题生成模型,隐含狄利克雷分配(latent Dirichlet allocation, LDA)模型^[20]尤其适用于自动处理大规模语料库,已经成为最受欢迎的文本表示模型之一.例如针对话题相似度计算问题, Schaal 等^[21]和 Rus 等^[22]讨论基于 LDA 模型的文本相似度的计算,解决了传统文本相似度计算中维数太高且稀疏的问题.这为基于 LDA 模型以每条在线评论为节点构建复杂网络提供了启发和可能.

针对现有研究不足,本文结合复杂网络方法和 LDA 模型的优点,基于 LDA 模型对消费者在线评论建立复杂网络,以每条文本评论为节点,根据评论间的话题关系确定连边,对网络的拓扑特性进行分析,并由此探讨了在线评论的相互关系及整体演化发展.研究发现,在产品销售初期消费者发表评论产生相互影响作用,而产品销售到中后期,消费者之间的相互影响作用变小.本文不仅是对现有文献的一个重要拓展,更为消费者在线评论提供了新的研究视角和理论解释.

2 在线评论复杂网络的构建

2.1 LDA 模型及消费者在线评论的话题计算

网络消费者通过在线评论参与互动,形成了一个以评论为节点的复杂网络.首先给出评论网络的定义.

定义 1 评论节点. 假设评论文档 D 共有 M 条评论, 记 $V = (v_1, v_2, \dots, v_M)$, 每条评论文本 v_i 称为评论节点.

定义 2 评论网络. 假设评论文档集有 M 个评论节点, 构成无向图 $G = (V, E)$, 称为评论网络. 其中 $E \subseteq V \times V$ 表示边集, 根据评论节点 $v_i, v_j \in V$ 的话题相似度大于一定阈值时连边, 即 $(v_i, v_j) \in E$.

其次, 为了进一步介绍基于 LDA 模型的评论相似度计算方法, 给出利用 LDA 模型计算评论语料集的话题及其分布的方法. 所谓话题(topic) 是 LDA 模型中的核心概念, 也称为主题. 因此, 有必要对 LDA 模型做一简单介绍.

LDA 模型是 Blei 等^[20]提出的一种文档生成概率模型, 具有明显的分层结构, 包括文档集层, 文档层和词层, 广泛用于文本建模, 文本分类, 相似度的判别等领域^[23]. 其不考虑词汇顺序, 假设每个文档都是由多个隐含话题构成, 而每一个话题又由文档中多个特定词汇构成. 这些隐含话题被文档集中所有文档共享, 不同文档的话题构成比例不同. 图 1 展示了 LDA 模型.

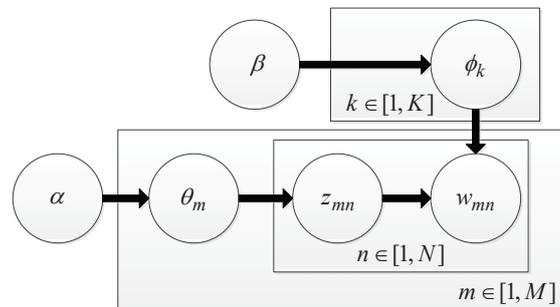


图 1 LDA 模型

Fig. 1 LDA Model

文档集合 D 包含 M 个文档和 V 个不同的词, 有 K 个潜在话题. 每个文档中有 N 个词, w_{mn} 是第 m 个文档的第 n 个词, $m = 1, 2, \dots, M; n = 1, 2, \dots, N$. α 为文档集中话题分布的 K 维参数, β 为潜在话题下的词分布参数, 是 $K \times V$ 的矩阵 ($\beta_{ij} = p(w_j = 1 | z_i)$). θ_m 表示第 m 个文档的话题的概率分布, 服从 Dirichlet 分布 $\text{Dri}(\alpha)$, ϕ_k 表示第 k 个话题下的词分布, $k = 1, 2, \dots, K$. z_{mn} 是第 m 个文档中第 n 个生成词的话题, $m = 1, 2, \dots, M; n = 1, 2, \dots, N$, $z_{mn} \sim \text{Multinomial}(\theta_m)$. LDA 模型生成文档的过程如下:

- 1) 对于话题 z_{mn} , 根据 Dirichlet 分布 $\text{Dri}(\beta)$ 得到该主题上的一个单词多项式分布向量 ϕ_k .
- 2) 根据 Dirichlet 分布 $\text{Dri}(\alpha)$ 产生话题分布 $\theta_m \sim \text{Dri}(\alpha)$;
- 3) 对于文档中的每个词 w_{mn} : 从 θ_m 的多项式分布生成话题 $z_{mn} \sim \text{Multinomial}(\theta_m)$, 从话题 z_{mn} 的多项式条件概率分布生成词 $w_{mn} \sim \text{Multinomial}(\theta_{z_{mn}})$.

建立评论语料库的 LDA 模型, 其参数估计采用 Gibbs 抽样, 具体迭代过程参见文献[21].

下面以京东商城 iPhone 6 plus 手机评论为例, 计算评论集的话题分布. 笔者获取了 2014-10-19~2014-12-30 的 1 000 条手机评论, 利用中科院计算技术研究所的汉语词法分析系统 ICTCLAS 对其进行分词处理, 并将评论中的“了”, “呢”等没有意义的词语及标点符号去掉. 取话题数 $K = 80$ (可根据 LDA 模型困惑度^[20]来选取话题数, 话题数越大困惑度越小(趋向最优), 但计算量也随之增大). 模型迭代初始参数取经验值 $\alpha = 50/K, \beta = 0.01$. 用 Gibbs 抽样迭代 5 000 次, 得到每条评论在话题上的分布及话题在词上的分布. 表 1 列出话题在词上的概率分布情况, 限于篇幅, 表中只显示两个话题.

表 1 中 Topic 1 和 Topic 2 代表评论集中的两个潜在话题, 每个话题右面是该话题在词上的概率分布, 将词按概率从大到小排列, 列出前十个词. Topic 1 讨论了手机包装, Topic 2 讨论了手机性能及整体评价.

2.2 在线评论相似度计算及评论网络的构建

本节利用上述 LDA 模型计算得到的在线评论的话题分布, 确定评论间的相似度, 从而构建评论网络. LDA 模型计算话题分布其实是将评论文本向量空间映射到话题分布上, 进而利用与文本对应的话题概

率来计算两个评论文本的相似度. 计算相似度常用的方法有 KL 距离和 Pearson 相似度等. 而在文本相似度计算方面, Pearson 相似度公式表现相对较好^[24], 与 LDA 模型结合可以有效地度量文本相似程度^[21]. 因此, 选取 Pearson 相似度来度量评论间相似度.

表 1 评论话题

Table 1 Topics of reviews

Number	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
Topic 1	包装	手机	盒子	知道	包	外包装	一分	保障	箱子	最后
	0.191 8	0.075 4	0.068 5	0.041 1	0.034 3	0.027 4	0.027 4	0.020 6	0.020 6	0.020 6
Topic 2	好	希望	性能	良好	相信	货真价实	发票	一个	细节	好说
	0.357 9	0.205 2	0.115 8	0.100 1	0.031 6	0.010 5	0.005 3	0.005 3	0.005 3	0.005 3

评论间的 Pearson 相似度为

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

其中 x_i, y_i 表示两条不同的评论在第 i 个话题上的概率分布, \bar{x} 和 \bar{y} 表示两条评论分别在所有话题上概率的均值, 即 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. r 表示两条评论在话题分布上的相关相似度, $0 \leq r \leq 1$. 当 r 接近 1 时, 表示两条评论在话题分布上正相关且相似. 当 r 接近 0 时, 表示两条评论在话题分布上不相关且不相似. 当 r 接近 -1 时, 表示两条评论在话题分布上负相关, 但话题分布不相似. 由此, 就确定了以评论为节点, Pearson 相似度为连边关系的构建复杂网络的原则.

以前小节中 iPhone 6 plus 手机评论为例构造评论网络. 根据所得的评论话题分布概率, 利用 Pearson 相似度公式计算评论间话题相似度. 评论话题相似度概率密度见图 2, 可以看出评论相似度密度分布接近钟形分布, 相似度集中在 0 附近. Pearson 相似度大于 0.5 表示变量有强的正相关线性关系, 如果两个文本之间的相关度低于 0.5, 那么它们描述话题相似的概率较小. 因此, 取阈值 τ 为 0.5, 相似度大于阈值连边构造的复杂网络如图 3.

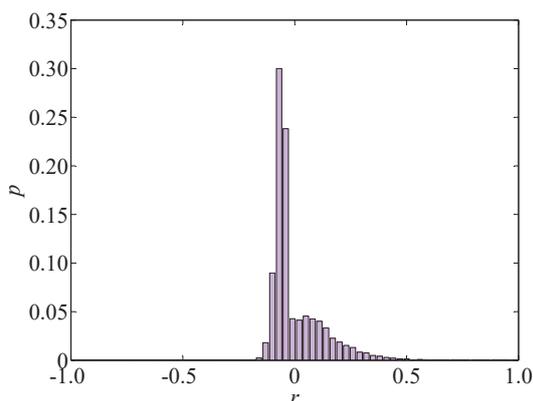


图 2 iPhone 6 plus 评论相似度概率密度

Fig. 2 Probability density of similarity about iPhone 6 plus' reviews

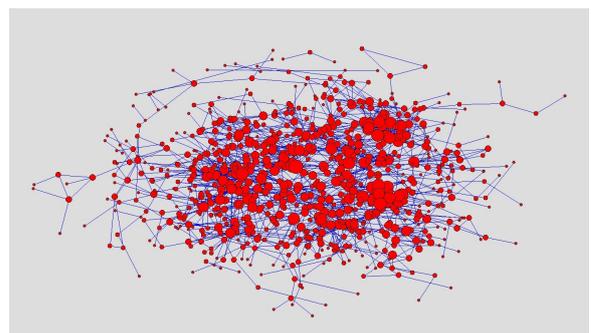


图 3 iPhone 6 plus 评论网络

Fig. 3 The reviews network of iPhone 6 plus

3 手机评论网络及其统计性质分析

3.1 实际手机在线评论网络构建

本文从苏宁易购, 天猫, 京东商城三个不同平台上分别获取 1 300 条 iPhone 7 手机在线评论, 以及京东

商城 iPhone 6 plus 的 5 617 条在线评论数据(参见表 2 中评论数), 每一种手机评论从产品销售初期开始选取。

对上述不同平台的 iPhone 7 在线评论, 以及京东商城 iPhone 6 plus 前 1 300 条在线评论和 5 617 条在线评论分别建立评论网络。首先, 对每种手机产品评论数据集按 2.1 节方法计算评论的话题分布, LDA 模型困惑度, 迭代初始参数同 2.1 节所取。通过 Gibbs 抽样迭代 5 000 次, 得到话题分布及词分布。然后, 利用 2.2 节中介绍的 Pearson 相似度公式, 对评论进行话题相似度计算。设定阈值为 $\tau = 0.5$, 相似度大于等于阈值时节点对进行连边且权值取为 1; 相似度小于阈值时节点对不相连。由此构造出 5 个不同的评论网络。

3.2 统计性质分析

为分析在线评论的相互关系及整体演化规律, 对 3.1 节构造的不同平台不同规模评论网络的统计特性进行计算, 结果见表 2。

表 2 评论网络统计性质计算结果
Table 2 Statistical properties of review network

评论网络	节点数 (评论数)	边数	平均度	直径	随机网络平均 路径长度	平均路径 长度	随机网络 聚类系数	聚类系数
苏宁易购 iPhone 7	1 300	7 075	10.885	12	3.270	4.609	0.008 5	0.469
天猫 iPhone 7	1 300	3 423	5.266	16	7.391	6.069	0.001 2	0.448
京东 iPhone 7	1 300	5 056	7.797	12	3.726	4.884	0.006 0	0.444
京东 iPhone 6 plus	1300	3 458	5.320	13	4.487	5.480	0.004 2	0.388
京东 iPhone 6 plus	5 617	92 564	32.959	7	3.394	3.598	0.002 9	0.423

1) 平均路径长度

平均路径长度描述网络中任意两个节点之间距离的平均值, 反映了网络的集中程度。从表 2 可以看出, 不同平台上相同产品相同规模评论网络的平均路径长度都与相同规模随机网络的平均路径长度相似, 网络直径较小。这表明, 消费者在线评论网络连接较强, 评论内容在话题上具有普遍相似性。

2) 聚类系数

聚类系数反映相邻节点之间联系的紧密程度^[25]。从表 2 可以看出, 不同平台上相同产品相同规模的评论网络都具有较高聚类系数, 远大于相同规模随机网络的聚类系数。说明评论讨论的产品特性及服务话题较为聚集, 大部分节点因讨论相似话题而相互聚集。反映了消费者在线评论普遍联系, 话题内聚的特性。

3) 度分布

无向网络中节点度定义为与节点直接相连的边的数目。在本文构建的评论网络中, 节点度表示与所关注评论节点具有话题相似的节点数量。网络中所有节点的度的平均值称为网络的平均度。节点度的分布情况用分布函数 $P(d)$ 描述, $P(d)$ 表示网络中度为 d 的节点占有节点数的比例, 即一个随机选定的节点度恰好为 d 的概率^[25]。最常见的度分布为钟形分布和长尾分布, 分别对应均匀网络和无标度网络。上述不同平台上相同规模的手机评论网络度分布概率如图 4 所示。图 5 为京东 iPhone 6 plus 的 5 617 条评论网络度的分布概率。

由图 4 可以看出, 不同平台上相同规模在线评论网络的度分布均为长尾分布。这表明, 网络中存在少数度极高的节点, 对应的评论与其它评论具有较高相似性, 即存在少数评论能够包含大多数评论所讨论的有关产品及服务话题。同时, 不同平台度分布形状略有不同。天猫评论网络比其它三个评论网络更接近幂律分布; 苏宁易购和天猫评论节点的最大度较京东的大, 说明少数度大的评论所包含的话题较京东更丰富; 而天猫度小的评论节点比其它两个平台多, 说明评论的话题差别较其它平台大; 京东上不同产品评论网络差异较小。

对比图 4(d)和图 5 可以看出, 网络规模较小时, 其度分布为长尾分布, 而规模较大时则表现为钟形分布。小世界网络的特性是聚类系数都远大于同规模随机网络的聚类系数, 网络的平均路径长度较小, 与相同规模的随机网络平均路径长度接近^[26]。而无标度网络具有网络规模不断扩大和优先连接特性, 优先连接即新的节点更倾向于与那些具有较高度度的节点相连, 度分布为幂律分布^[27]。结合表 2 统计结果, 京东 iPhone 6

plus 评论网络有 561 7 个节点, 边数为 92 564 条, 比其它小规模评论网络的边明显多; 随着节点数增多, 网络平均度, 平均路径长度和聚类系数明显增大. 可见, 节点少时评论网络为具有优先连接的无标度网络, 而随着节点数增加, 评论网络逐渐演化为一个钟形分布的小世界网络. 其内在原理为消费者在产品销售初期购买产品时参考已有评论, 新发表评论讨论的话题受已发表评论影响, 讨论话题趋于一致, 即评论行为有相互影响作用, 使得新发表的在线评论与已有评论中讨论频率高的话题所在评论相似性高, 这恰好符合优先连接特性, 评论形成一个无标度网络. 而产品销售到中后期, 消费者受周围人对产品舆论影响, 已发表评论对消费者影响变小, 即评论行为相互影响作用变小, 使得评论整体话题内聚, 评论具有普遍相似性, 形成具有较大聚类系数和较短平均路径且具有钟形分布的小世界网络.

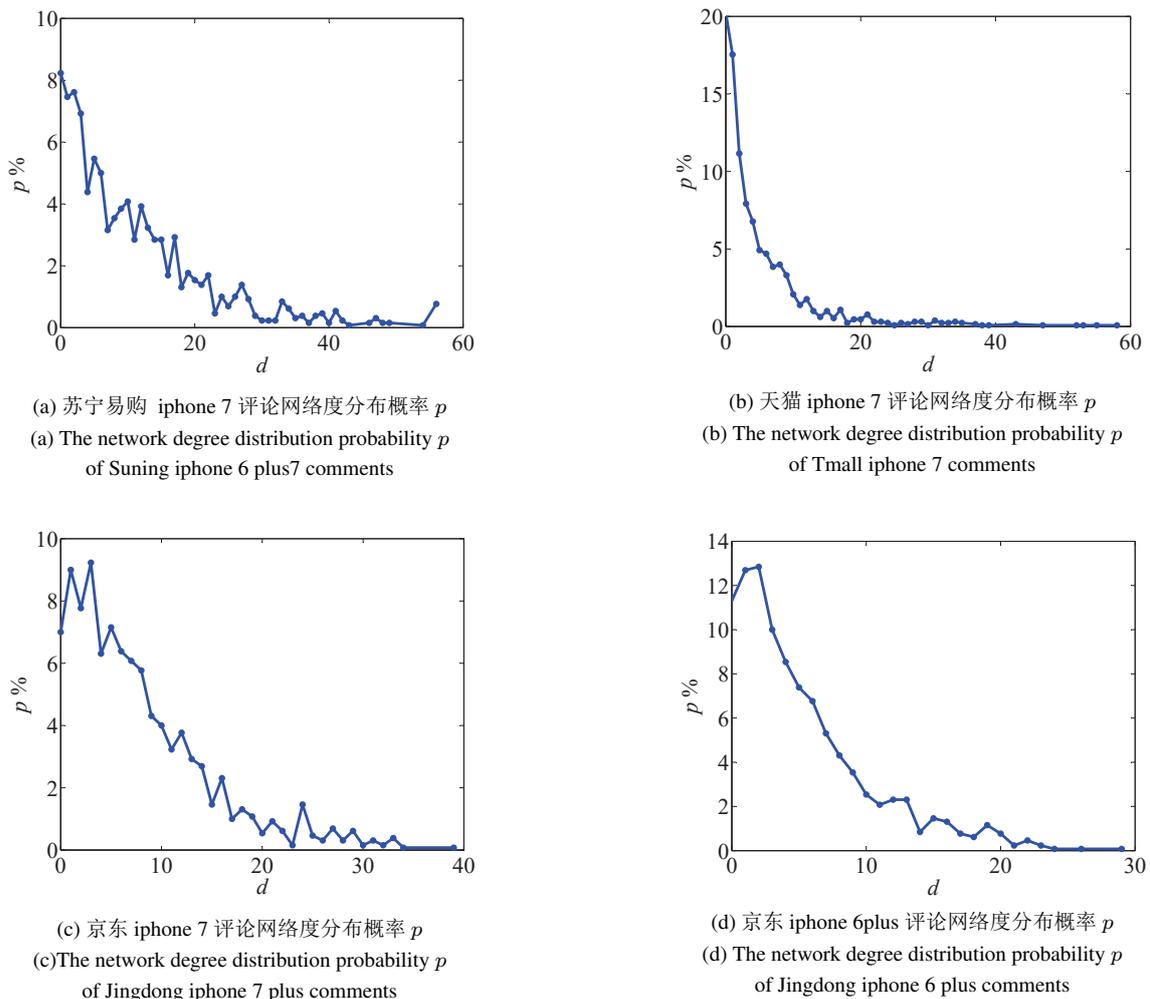


图4 不同平台上相同规模的手机评论网络度分布概率

Fig. 4 Degree distribution probability p of same scale review network on different platforms

4 结束语

本研究构建了以消费者在线评论为节点, 评论间话题相似关系确定连边的评论网络. 以苏宁易购, 天猫, 京东商城等平台上的手机评论语料库为例, 对相应评论网络的统计特性及网络结构进行分析. 研究表明, 不同平台上相同产品相同规模的评论网络统计特性相近, 反映了在线评论的普遍联系性及话题内聚性特征; 而同一产品不同规模的评论网络性质差异明显, 小规模评论形成一个无标度网络, 而随着规模增大, 评论网络则逐渐演化为一个度分布为钟形分布的小世界网络, 说明产品销售初期, 已有在线评论对消费者发表评

论行为影响较大,而在中后期则影响较小.可见,产品销售初期的消费者在线评论满意度尤为重要.因此,企业在制定营销策略时,应考虑如何提高初期消费者的满意度,促使他们通过在线评论影响未来消费者的评论行为,进而提高产品的好评率,促进产品销售.

本研究丰富了在线评论文本分析方法,为研究消费者在线评论提供了新的研究视角.同时,验证了在线评论对消费者决策的重要参考价值,反映了中国乃至东方的集体主义倾向引发的从众心理和跟随效应,对于企业市场营销和管理实践具有较强的指导意义.

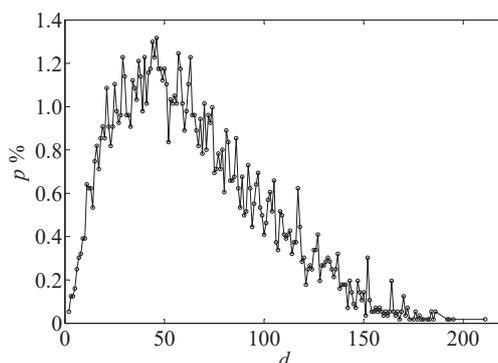


图5 京东 iphone 6 plus 的 5 617 条评论网络度分布概率(p)

Fig. 5 Degree distribution probability p of review network with 5 617 nodes about Jingdong iphone 6 plus

参考文献:

- [1] Schiffman L G, Kanuk L L. Consumer behavior. *Consumer Behavior*, 2004, 43(19): 1121–1139.
- [2] 张艳辉, 李宗伟. 在线评论有用性的影响因素研究: 基于产品类型的调节效应. *管理评论*, 2016, 28(10): 124–133.
Zhang Y H, Li Z W. Analysis of the factors that influence online reviews helpfulness: Based on the regulating effect of product type. *Management Review*, 2016, 28(10): 124–133. (in Chinese)
- [3] 岳中刚, 王晓亚. 在线评论与消费者行为的研究进展与趋势展望. *软科学*, 2015, 29(6): 90–93.
Yue Z G, Wang X Y. Research progress and future prospect on online reviews and consumer behavior. *Soft Science*, 2015, 29(6): 90–93. (in Chinese)
- [4] 刘洋, 廖貅武, 刘莹. 在线评论对应用软件及平台定价策略的影响. *系统工程学报*, 2014, 29(4): 560–570.
Liu Y, Liao X W, Liu Y. The impact of online review on software and platform's strategies. *Journal of Systems Engineering*, 2014, 29(4): 560–570. (in Chinese)
- [5] Ho Y C, Wu J, Tan Y. Disconfirmation Effect on Online Rating Behavior: A Dynamic Analysis. *Social Science Electronic Publishing*, 2014.
- [6] Moe W W, Schweidel D A. Online product opinions: Incidence, evaluation and evolution. *Marketing Science*, 2011, 31(3): 372–386.
- [7] Lee Y J, Hosanagar K, Tan Y. Do I follow my friends or the crowd. *Information cascades in online movie ratings*. *Management Science*, 2015, 61(9): 2241–2258.
- [8] Nbsp A, Nbsp E, Schlosser. Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*. 2005, 32(2): 260–265.
- [9] Shen W, Hu Y J, Ulmer J R. Competing for attention: An empirical study of online reviewers' strategic behavior. *MIS Quarterly*. 2015, 39: 683–696.
- [10] Tsekouras D. Variations on a Rating Scale: The Effect on Extreme Response Tendency in Product Ratings. *Münster: ECIS, Working Paper 186*, 2015.
- [11] 姜巍, 张莉, 戴翼, 等. 面向用户需求获取的在线评论有用性分析. *计算机学报*, 2013, 36(1): 119–131.
Jiang W, Zhang L, Dai Y, et al. Analyzing helpfulness of online reviews for user requirements elicitation. *Chinese Journal of Computers*, 2013, 36(1): 119–131. (in Chinese)
- [12] Taboada M, Tofiloski M, Brooke J, et al. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 2011, 37(2): 267–307.

- [13] 王伟, 王洪伟. 特征观点对购买意愿的影响: 在线评论的情感分析方法. 系统工程理论与实践, 2016, 36(1): 63–76.
Wang W, Wang H W. The influence of aspect-based opinions on user's purchase intention using sentiment analysis of online reviews. *Systems Engineering: Theory & Practice*, 2016, 36(1): 63–76. (in Chinese)
- [14] 尹裴, 王洪伟. 面向产品特征的中文在线评论情感分类: 以本体建模为方法. 系统管理学报, 2016, 25(1): 103–114.
Yin P, Wang H W. Sentiment classification for Chinese online reviews at product feature level through domain ontology method. *Journal of Systems and Management*, 2016, 25(1): 103–114. (in Chinese)
- [15] 李倩倩, 顾基发. 用户行为驱动的在线社交网络建模. 系统工程学报, 2015, 30(1): 9–15.
Li Q Q, Gu J F. Activity driven modelling of online social network. *Journal of Systems Engineering*, 2015, 30(1): 9–15. (in Chinese)
- [16] 李永奎, 周宗放. 基于无标度网络的关联信用风险传染延迟效应. 系统工程学报, 2015, 30(5): 575–583.
Li Y K, Zhou Z F. Contagion delayed effects of associated credit risk based on scale-free network. *Journal of Systems Engineering*, 2015, 30(1): 9–15. (in Chinese)
- [17] 周文, 陈伟, 郎益夫. 集群创新网络知识动态增长研究: 基于过程视角. 系统工程学报, 2015, 30(4): 431–441.
Zhou W, Chen W, Lang Y F. Knowledge dynamic growth in the innovation network of industrial cluster: Based on process perspective. *Journal of Systems Engineering*, 2015, 30(4): 431–441. (in Chinese)
- [18] 索琪, 郭进利. 在线社交网络资源评论关系超网络演化模型. 系统管理学报, 2016, 25(5): 852–857.
Suo Q, Guo J L. An evolving hypernetwork model for the relationship of resources comments in online social networks. *Journal of Systems and Management*, 2016, 25(5): 852–857. (in Chinese)
- [19] 王伟, 王洪伟. 面向竞争力的特征比较网络: 情感分析方法. 管理科学学报, 2016, 19(9): 109–126.
Wang W, Wang H W. Comparative network for product competition in feature-levels through sentiment analysis. *Journal of Management Science in China*, 2016, 19(9): 109–126. (in Chinese)
- [20] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, 3: 993–1022.
- [21] Schaal M, O'Donovan J, Smyth B. An analysis of topical proximity in the twitter social graph // *International Conference on Social Informatics*. Berlin, Heidelberg: Springer, 2012: 232–245.
- [22] Rus V, Niraula N, Banjade R. Similarity measures based on latent Dirichlet allocation // *International Conference on Intelligent Text Processing and Computational Linguistics*. 2013: 459–470.
- [23] 李文波, 孙乐, 张大鲲. 基于 Labeled-LDA 模型的文本分类新算法. 计算机学报, 2008, 31(4): 620–627.
Li W B, Sun L, Zhang D K. Text classification based on labeled LDA model. *Chinese Journal of Computers*, 2008, 31(4): 620–627. (in Chinese)
- [24] Huang A. Similarity measures for text document clustering // *Proceedings of the Sixth New Zealand Computer Science Research Student Conference*. Christchurch, New Zealand: 2008: 49–56.
- [25] 汪小帆, 李翔, 陈关荣. 网络科学导论. 北京: 高等教育出版社, 2012.
Wang X F, Li X, Chen G R. *Network Science: An Introduction*. Beijing: Higher Education Press, 2012. (in Chinese)
- [26] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. *Nature*, 1998(6684): 440–442.
- [27] Barabasi A, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509–512.

作者简介:

刘晓君(1987—), 女, 内蒙古乌兰察布人, 博士生, 研究方向: 文本挖掘与复杂网络, Email: cool.sky.jun@163.com;

那日萨(1970—), 男, 内蒙古乌兰浩特人, 教授, 博士生导师, 研究方向: 文本挖掘与复杂网络, Email: nmgnrs@dlut.edu.cn;

崔雪莲(1989—), 女, 山东青岛人, 博士生, 研究方向: 文本挖掘与复杂网络, Email: cxuelian@mail.dlut.edu.cn.