

周静^①, 沈俏蔚^②, 涂平^③, 王汉生^④

摘要 在线用户创造内容 (User Generated Content, UGC) 已经成为人们在社交平台上进行交流与信息分享的主要方式。高质量的 UGC 能够吸引更多的广告主, 为平台带来可观的收入。因此, 如何鼓励人们在社交平台上贡献优质的内容已成为研究人员关心的重要问题之一。用户为何发帖? 产生 UGC 的动机是什么? 这类问题已经受到越来越多学者的关注。本文提出了一个全面的效用理论模型, 用于研究用户的发帖动机。在已有的内在效用和形象效用的基础上, 本文提出了交互效用的概念。具体地, 交互效用是指用户通过与社交平台上的好友进行互动而获得的效用。其次, 本文在效用最优化的方程中加入了时间约束这一条件, 通过该约束条件, 我们可以进一步分析用户是如何分配发帖与阅读他人帖子的时间的。最后, 本文用新浪微博的数据对理论模型进行了实证检验。

关键词 社交网络; 交互效用; UGC; 负二项回归

(稿件审理信息 终审主编/联合主编: 陈煜波; 专业主编: 黄敏学; 收稿日期: 2017 年 07 月 12 日; 修订次数: 3 次)

原创还是转发 ——基于社交媒体 UGC 的交互效用研究^⑤

0 引言

随着在线社交平台 (如 Facebook、Twitter 以及微博) 的蓬勃发展, 在线用户创造内容 (User Generated Content, UGC) 已经成为社交平台用户进行交流与信息分享的主要手段。UGC 的产生和网络活跃度息息相关, 一般来说, 一个社交网站活跃度越高, 广告主进行广告展现的机会就越多, 平台的收入也会越高。例如, Facebook 2014 年的年报显示, 其营业收入超过 120 亿美元, 而

其中有 90% 的收入来自广告销售。Facebook 之所以有如此多的广告收入, 是因为每天有数以亿计的用户活跃在其平台上。活跃用户的 UGC 对 Facebook 来说是潜在的商业价值。因此, 如何提高 UGC 的数量与质量成为平台运营者关心的主要问题之一。

UGC 的种类众多, 本文重点讨论文本形式的 UGC。已有文献对用户发帖动机的研究表明, 人们在 Twitter 上发帖是为了获得内在效用 (Intrinsic Utility) 和形象效用 (Image-Related U-

①周静, 中国人民大学统计学院, 讲师, E-mail: zhoujing_89@126.com

②沈俏蔚, 北京大学光华管理学院市场营销系, 教授, E-mail: qshen@gsm.pku.edu.cn

③涂平, 北京大学光华管理学院市场营销系, 教授, E-mail: tuping@gsm.pku.edu.cn

④王汉生, 北京大学光华管理学院商务统计与经济计量系, 教授, E-mail: hansheng@pku.edu.cn

⑤基金项目: 本研究得到国家自然科学基金项目“原创还是转发? 社交网络视角下的 UGC 产生动机研究”(71702185) 的资助以及中国人民大学 2018 年度“中央高校建设世界一流大学 (学科) 和特色发展引导专项资金”的支持。非常感谢两名评审专家提出的宝贵意见。

tility) (Toubia and Stephen, 2013)。然而, 还有很多其他效用驱动的动机因素并未被很好地研究。例如, Shriver 等 2013 年的一项调查研究显示, 有相当比例的用户表示在网上发帖是为了扩大自己的社交圈以及和圈里的好友进行互动。因此本文认为, 社交网络的独特性使得相互联系的用户之间产生了更多的互动, 而这种基于互动效用的动机研究正是之前文献中未涉及的, 因此本文的第一个贡献是研究基于交互效用 (Interaction Utility) 的发帖动机。本文的第二个贡献是探究了用户的时间成本, 即用户在发帖与阅读帖子之间是如何进行时间分配的。之前的文献只从发帖的角度进行研究, 而忽略了时间成本因素。假设人们用在社交平台上的时间是一定的, 即用户发帖和阅读帖子的时间一定, 在有限的时间内, 用户需要平衡发帖与阅读他人的帖子的时间。本文在优化用户的效用函数时, 加入了时间成本这一条件, 使决策者可以考虑不同情况下的最优决策。

本文的结构如下, 第一部分进行文献回顾, 主要回顾和 UGC 以及社会交互相关的文献, 接下来第二部分是理论基础与研究假设, 重点阐述效用理论模型, 模型的推导以及相关结论的讨论。第三部分用新浪微博的实际数据对理论模型进行了验证。在文章的最后即第四部分给出了本文的管理启示和研究局限。

1 文献回顾

近年来学术界关于 UGC 的研究曾掀起一阵热潮。在 2012 年, 国际著名的营销学杂志《营销科学》(Marketing Science) 曾专门刊出一期关于 UGC 研究的特刊 (Special Topic)。在所有发表的研究论文中, 学者们关注的研究主题可以大体划分为以下三个方面: UGC 是如何产生的以及用户为什么要贡献 UGC、UGC 带来的影响、分析 UGC 数据所产生的新的方法 (Fader and Winer,

2012)。本文将重点回顾和社交网站发帖动机相关的研究。

Shriver 等 (2013) 的一篇文章研究了发帖数与好友数之间的因果关系。作者研究了社会网络效应对 UGC 的影响, 尽管他们并未直接研究效用或动机, 但是该研究发现了一个重要的结论: 社会关系可以对是否发帖及发帖数量产生重要影响。同年, Toubia 和 Stephen (2013) 的文章从效用的角度探讨在 Twitter 上人们的发帖动机。他们指出内在效用 (Intrinsic Utility) 和形象效用 (Image-Related Utility) 可以驱动用户发帖。其中, 内在效用假设人们的行为受到自身内在满足感的驱使, 人们可以从发帖中直接获得效用。而形象效用则是受他人感知的驱使, 即人们获得形象效用主要来自于他人的认可, 例如, 拥有更多的粉丝。研究表明, 当粉丝数增加时, 人们的发帖行为会因受到不同效用的驱使而不同。如果用户受内在效用驱使多一些, 那么粉丝数的增加会带来发帖数的增加。如果用户受形象效用的驱使多一些, 那么粉丝数的增加会带来发帖数的减少。他们的实证研究结果表明, 在解释用户发帖动机时, 形象效用比内在效用更重要。最近的一篇分析文章, Iyer 和 Katona (2016) 从受关注的角度研究了影响社会化交流的因素。作者指出, 为了获得接收者的注意, 来自信息发送者之间的竞争是影响信息产生的一个重要因素。该研究的一个重要结论是: 随着信息交流范围的扩大, 信息交流的动机会逐渐减弱。

综上所述, 以上回顾的三篇文献都直接或间接地研究了发帖动机或由效用驱动的动机。然而, 关于为了扩大社交圈以及和他人进行互动的动机, 似乎还没有与之对应的研究, 因为这样的研究需要研究者能够观察到网络结构。而事实上, 在 Toubia 和 Stephen (2013) 的文章里已明确指出, 如果有网络结构数据, 那么会更加丰富现有的研究结果。用户的社交网络结构对于评价 UGC 的产生具有重要意义。这是因为人们的行为

很容易受到与他在同一个网络里的好友的影响。实际上,近年来在营销领域有很多相关的研究在探索社会交互对行为预测的影响 (Goel and Goldstein, 2013; Iyengar et al., 2011; Dover et al., 2012; Nitzan and Libai, 2011; Wang et al., 2013; 施卓敏, 等, 2015; 周志民, 等, 2011)。这些研究的结果表明,加入社会交互信息之后的行为预测要比没加入这些信息的预测精度提高很多。

随着网络结构数据越来越受到重视,许多传统的营销问题在这个新的情境下都进行了重新检验,例如,新产品扩散研究 (Iyengar et al., 2011; Dover et al., 2012; Hu and Van den Bulte, 2014; Iyengar et al., 2015; Risselada et al., 2014)、客户流失 (Nitzan and Libai, 2011)、信用评分 (Wei et al., 2016)、网络扩散 (Hasan and Bagde, 2015) 以及消费者选择 (Narayan et al., 2011)。因为通过好友关系可以获得传统变量无法获取的信息。在预测消费者行为上,社会交互也起了很大的作用,作为个体的我们并不是独立存在于这个世界上,每个人都会形成不同的复杂的社会网络,来自个体网络信息的变量对于预测个体行为将是一个强有力的指标。为了研究互动行为,需要获得网络结构数据和好友信息。互动行为在解释传统的营销问题上发挥了重要的作用,因此本文认为如果在解释用户发帖动机时引入社会交互,会更加丰富人们对 UGC 的理解。所以本文在理论上有以下创新:首先,从微观层面上解释发帖动机时引入了社会交互的概念。其次,发帖可以进一步被区分为原创帖和转发帖,每种发帖动机背后的机制是不同的。按照常识,原创帖比转发帖需要付出更多的成本,因为原创帖需要更多原创性的工作和富有创造力的想法。接下来我们将具体讨论理论模型。

2 理论基础及研究假设

2.1 内在效用 vs. 形象效用 vs. 交互效用

内在效用被定义为个体进行某种活动的一种

自然需求,这是一种先天的兴趣,不依赖于任何其他的外在因素 (Ryan and Deci, 2000)。当人们被内在效用驱使时,个体的行为只和他们的内在特质有关。形象效用被认为是社会地位的测量 (Toubia and Stephen, 2013),该效用关注的是别人对自己的看法。当用户受形象效用驱使时,他们更希望在别人面前表现自己,通过被别人注意和尊重而获得自我实现。形象效用更强调结果层面,而对于实现这一结果的过程并没有过多限制,例如,粉丝数增加被认为是形象效用的一种体现 (Toubia and Stephen, 2013)。根据形象效用的定义,其他用户对某个用户发的帖子进行了评论或转发,表明其他用户对该用户的发帖和这个用户的认可,那么也可以认为是形象效用的一种体现。

交互效用是本文新提出的概念,将它定义为从某些特定的社会互动活动中因满足感而获得的效用。根据 Goffman (1983) 经典的社会交互文献,这些社会互动包括交换、竞争、合作、冲突和强迫。本文认为,在社交媒体上的互动基本可以分为两类。第一类是来自他人的反馈,例如,一个典型的情形是其他用户对某个用户发的帖子进行了评论或转发。在这种情形下,用户通过获得他人的反馈 (评论或转发) 而获得交互效用。该效用可以帮助用户保持和增强与网络上好友的联系。在某种程度上,该效用可以被看作是形象效用的一个补充,因为获得交互效用的一个前提是有很多粉丝关注。假设用户 A 和用户 B 都拥有 1000 个粉丝,当 A 发帖时,几乎超过一半的粉丝对他的帖子进行了回应 (包括转发、评论点赞等),然而,当 B 发帖时,几乎没人对 B 的帖子产生回应。很显然 A 比 B 获得更多的交互效用。因此,我们可以将交互效用操作化为转发数。第二类交互效用是指用户阅读他人的帖子。例如,社交平台上的每个人都可以选择关注他们感兴趣的人或团体,这样就可以收到来自这些人或团体的发帖信息。用户接收到的帖子,内容多

种多样,一些高质量的帖子会引起人们的阅读和创作兴趣,然而一些被转发了多次、内容重复的帖子或者没有任何营养的帖子不仅占用用户时间,还会大大削减人们的阅读兴趣。所以本文认为,阅读高质量的帖子更有助于用户获得交互效用。因此,交互效用另一个操作化的指标可以是用户关注的人的发帖数。

2.2 理论模型

假设在一个大的社交网络里,每个用户都面临两个选择,发原创帖或是转发帖,定义 Y_1 为发原创帖的数量, Y_2 为发转发帖的数量。于是一个用户发帖(原创或转发)所获得的效用由内在效用、形象效用和交互效用构成。与此同时,用户在发帖和阅读帖子时有一定的时间成本,假设每个用户花费在社交网络上的时间是一定的,那就意味着用户需要平衡发帖与阅读帖子之间的成本。因此本文将讨论在有时间成本约束的条件下最优化用户获得的效用。

首先,定义 $\{U_1^{Intr}, U_2^{Intr}\}$ 分别代表用户发一条原创帖和一条转发帖带来的内在效用。 U_1^{Intr} 和 U_2^{Intr} 可以是用户人口统计信息变量的函数,因此,一个人获得的总内在效用 U^{Intr} 可以表示为

$$U^{Intr} = Y_1 U_1^{Intr} + Y_2 U_2^{Intr} \quad (1)$$

其次,定义 $\{U_1^{Imag}, U_2^{Imag}\}$ 为用户发一条原创帖和转发帖分别获得的形象效用,借鉴 Toubia and Stephen (2013) 的做法, U_1^{Imag} 和 U_2^{Imag} 可以表示成用户粉丝数的函数。因此,一个人获得的总形象效用 U^{Imag} 可以表示为

$$U^{Imag} = Y_1 U_1^{Imag} + Y_2 U_2^{Imag} \quad (2)$$

最后,定义 $\{U_1^{Inter}, U_2^{Inter}\}$ 为用户发一条原创帖或转发帖分别获得的交互效用,交互效用除了依赖于用户从他的粉丝中收到的反馈,也会受到他关注的人(即好友)的发帖的影响。本文认为,用户受好友影响的机制源于网络进化动力的驱动。一个快速发展的网络总是能吸引很多用户进行互动。如果一个网络拥有很强的进化动力,

那么用户会获得更多的交互效用,因为此时他们有更多的机会和别人互动。组织行为领域的研究表明创新是一个组织进化的动力 (Strebel, 1987),对于在线社交平台来说,创新来源于用户的原创帖。如果一个社交平台每天被重复的信息充斥着,那么久而久之该平台就会沉寂下去,因此平台的原创性直接决定了它的创新程度,进而影响网络进化的驱动力。这说明网络进化的驱动力可以由用户的原创帖数和转发帖数决定,因此用 R_1 和 R_2 分别表示用户好友的原创帖数和转发帖数,这里再定义一个新的变量 q ,用以表示网络进化的驱动力。所以 q 是 R_1 和 R_2 的一个函数,即 $q = Q(R_1, R_2)$ 。例如,假设 $q = R_1 / (R_1 + R_2)$,这说明网络进化驱动力会随着好友原创帖的增多而增加,随着转发帖的增多而减少。因此交互效应 U^{Inter} 可以表示为

$$U^{Inter} = Y_1 U_1^{Inter}(q) + Y_2 U_2^{Inter}(q) \quad (3)$$

综上,一个用户发帖所获得的总效用 U^{total} 可以表示为

$$U^{total} = U^{Intr} + U^{Imag} + U^{Inter} \\ U^{total} = \{U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q)\} Y_1 \\ + \{U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q)\} Y_2 \quad (4)$$

接下来,分析用户受到的时间成本的约束。假设每个人都面临一个总的时间成本,我们将其定义为 T ,包括发帖的时间成本以及阅读帖子的时间成本。如果用户将大部分时间用在发帖上,那么留给阅读和消化他人帖子的时间就会变少,即用户需要在发帖与阅读帖子之间进行合理的时间分配。基于此,假设用户每发一条原创帖的成本是 $C_1(Y_1, q)$,是 Y_1 和 q 的函数。同理,发一条转发帖的成本是 $C_2(Y_2, q)$,这里 C_1 和 C_2 代表成本函数, Y_1 和 Y_2 分别代表原创帖数和转发帖数。引入网络进化驱动力变量 q ,是因为该变量反映了在社交平台上人们发帖和阅读帖的动机来源。阅读别人的帖子通常会带来创作的灵感,这会降低用户发原创帖的成本。因此从某种程度

上说,网络进化驱动力 q 会影响用户的成本函数。最后将阅读他人帖子的成本定义为 $C_3(R_1+R_2)$, 该成本函数是 R_1 和 R_2 的函数。这表明,好友发帖越多,用户用于阅读帖子的时间就越多。因此一个用户面临的成本约束可以表示为

$$C_1(Y_1, q) + C_2(Y_2, q) + C_3(R_1 + R_2) \leq T \quad (5)$$

在进行给定约束条件下的效用最优化求解前,先讨论 3 个基本的假设。

2.3 基本假设

(A₁) 边际成本递增。假设用户发帖(原创或转发)的成本是边际递增的。这是因为随着原创帖的增加,每多发一条原创帖,留给用户进行原创的资源就会减少,也就是说用户需要更努力去寻找更多的资源与灵感,表现在成本上就是边际成本递增。因此有如下的假设:

$$\begin{cases} \frac{\partial C_1(Y_1, q)}{\partial Y_1} > 0 \\ \frac{\partial^2 C_1(Y_1, q)}{\partial^2 Y_1} > 0 \end{cases} \quad \begin{cases} \frac{\partial C_2(Y_2, q)}{\partial Y_2} > 0 \\ \frac{\partial^2 C_2(Y_2, q)}{\partial^2 Y_2} > 0 \end{cases} \quad (6)$$

(A₂) 激励效应。由前文可知,成本函数 $C_1(Y_1, q)$ 和 $C_2(Y_2, q)$, 其中 q 衡量的是网络进化的驱动力,它受原创帖和转发帖的影响。一个活跃度很高的网站通常拥有大量且高质量的 UGC, 这些可以成为用户发帖的资源,从而产生自己的原创帖。当一个网络处在上升阶段时,用户每天都会接触到高质量的 UGC, 因而他们的创作成本较低,即网络进化驱动力越强,用户的创作成本越低。这是因为此时的网络会给用户带来更多的创作灵感。因此将第二个假设称为激励效应。有如下的形式:

$$\begin{cases} \frac{\partial C_1(Y_1, q)}{\partial q} < 0 \\ \frac{\partial C_2(Y_2, q)}{\partial q} < 0 \end{cases} \quad (7)$$

(A₃) 进化驱动力。最后一个假设是关于 q 的。正如前文所讨论的,网络进化驱动力和好友

的发帖数以及转发帖数相关。原创性的内容通常被认为是创新的源泉,而转发的内容一般是重复信息的传播。而网络的进化主要和创新程度相关,换句话说,创新性决定了一个网络的生存时间。因此,如果一个网络里原创的内容多,那么网络就处在一个不断上升发展的状态。反之,如果一个网络每天都充斥着重复的信息,用户无法从网络中获得新的信息和知识,那么网络很有可能就会走向消亡。因此网络进化驱动力会随着原创帖比例的增加而增强,主要受潜在创新性的驱使。而网络进化驱动力会随着转发帖比例的增加而下降,主要受获得新信息能力的限制。因此该假设概括为随着原创帖比例的增加,网络进化动力增强,随着转发帖比例的增加,网络进化动力减弱。

$$\begin{cases} \frac{\partial Q(R_1, R_2)}{\partial R_1} > 0 \\ \frac{\partial Q(R_1, R_2)}{\partial R_2} < 0 \end{cases} \quad (8)$$

2.4 理论模型结果讨论

根据经济学里理性人的假设并结合本研究的设定,在给定时间成本约束的条件下,通过选择最优的 Y_1 和 Y_2 来最大化用户的效用。用数学表达式则可以表示为

$$\begin{aligned} \max U^{total} = & \{ U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q) \} Y_1 + \\ & \{ U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q) \} Y_2 \\ \text{s. t. } & C_1(Y_1, q) + C_2(Y_2, q) + C_3(R_1 + R_2) \leq T \end{aligned}$$

其中 $q = Q(R_1, R_2)$ 。运用拉格朗日方法对该问题进行优化,定义 λ 为拉格朗日乘子, L 为最大化效用,则有

$$L = \{ U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q) \} Y_1 + \{ U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q) \} Y_2 + \lambda \{ T - C_1(Y_1, q) + C_2(Y_2, q) + C_3(R_1 + R_2) \}$$

上式中,分别对 Y_1 , Y_2 和 λ 求一阶导,可得:

$$\begin{cases} \frac{\partial L}{\partial Y_1} = U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q) - \lambda \frac{\partial C_1(Y_1, q)}{\partial Y_1} = 0 \\ \frac{\partial L}{\partial Y_2} = U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(2) - \lambda \frac{\partial C_2(Y_2, q)}{\partial Y_2} = 0 \\ \frac{\partial L}{\partial \lambda} = T - C_1(Y_1, q) + C_2(Y_2, q) + C_3(R_1 + R_2) = 0 \end{cases}$$

整理后, 得到

$$\begin{cases} \frac{\partial C_1(Y_1, q)}{\partial Y_1} = \frac{U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q)}{\lambda} \\ \frac{\partial C_2(Y_2, q)}{\partial Y_2} = \frac{U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q)}{\lambda} \\ C_1(Y_1, q) + C_2(Y_2, q) + C_3(R_1 + R_2) = T \end{cases} \quad (9)$$

假设 Y_1 和 Y_2 的最优解分别为 Y_1^* 和 Y_2^* , 方程组 (9) 的第三个式子可以改写为

$$C_1(Y_1^*, q) + C_2(Y_2^*, q) + C_3(R_1 + R_2) = T \quad (10)$$

进而对方程 (10) 中的 R_1 和 R_2 求导, 得到

$$\begin{cases} \frac{\partial C_1}{\partial Y_1^*} \frac{\partial Y_1^*}{\partial R_1} + \frac{\partial C_1}{\partial q} \frac{\partial q}{\partial R_1} + \frac{\partial C_2}{\partial Y_2^*} \frac{\partial Y_2^*}{\partial R_1} + \frac{\partial C_2}{\partial q} \frac{\partial q}{\partial R_1} + C_3'(R_1 + R_2) = 0 \\ \frac{\partial C_1}{\partial Y_1^*} \frac{\partial Y_1^*}{\partial R_2} + \frac{\partial C_1}{\partial q} \frac{\partial q}{\partial R_2} + \frac{\partial C_2}{\partial Y_2^*} \frac{\partial Y_2^*}{\partial R_2} + \frac{\partial C_2}{\partial q} \frac{\partial q}{\partial R_2} + C_3'(R_1 + R_2) = 0 \end{cases}$$

因此有

$$\begin{aligned} & \left[\frac{U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q)}{\lambda} \right] \frac{\partial Y_1^*}{\partial R_1} + \left[\frac{U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q)}{\lambda} \right] \frac{\partial Y_2^*}{\partial R_1} \\ & + \frac{\partial q}{\partial R_1} \left(\frac{\partial C_1}{\partial q} + \frac{\partial C_2}{\partial q} \right) < 0 \end{aligned} \quad (11)$$

$$\begin{aligned} & \left[\frac{U_1^{Intr} + U_1^{Imag} + U_1^{Inter}(q)}{\lambda} \right] \frac{\partial Y_1^*}{\partial R_2} + \left[\frac{U_2^{Intr} + U_2^{Imag} + U_2^{Inter}(q)}{\lambda} \right] \frac{\partial Y_2^*}{\partial R_2} \\ & + \frac{\partial q}{\partial R_2} \left(\frac{\partial C_1}{\partial q} + \frac{\partial C_2}{\partial q} \right) < 0 \end{aligned} \quad (12)$$

在前文的讨论中有如下假设: $\frac{\partial q}{\partial R_1} > 0$, $\frac{\partial q}{\partial R_2} < 0$, $\frac{\partial C_1}{\partial q} < 0$, 以及 $\frac{\partial C_2}{\partial q} < 0$ 。根据方程 (11) 和 (12), 本文讨论下面两种有意思的情形。

情形 1: Y_1^* 和 Y_2^* 如何随着 R_1 的变化而变化? 这里评价用户的原创或转发是如何随着好友的原创帖数的变化而变化的。从式子 (11) 我们

可以看到, 这等价于评价 $\frac{\partial Y_1^*}{\partial R_1}$ 和 $\frac{\partial Y_2^*}{\partial R_1}$ 的取值情况。

根据假设 (A₂) 和 (A₃) 可知, $\frac{\partial q}{\partial R_1}$

$\left(\frac{\partial C_1}{\partial q} + \frac{\partial C_2}{\partial q} \right)$ 取值为负, 效用部分的符号和 λ 的符号都为正, 因此为了保证不等式 (11) 右边为

负, $\frac{\partial Y_1^*}{\partial R_1}$ 和 $\frac{\partial Y_2^*}{\partial R_1}$ 或者取正或者取负。例如, 如果

原创相较于转发的比例高 (即 $\frac{\partial q}{\partial R_1}$ 的值较大), 说明这是一个网络进化动力较强的活跃网络。用户

在这样的环境下更容易发帖, 因为此时受到周围活跃气氛的影响, 发帖的成本也比较低。在这种情形下, 随着 R_1 的增加, Y_1^* 和 Y_2^* 也会增加。

相反, 如果网络进化动力很弱, (例如, $\frac{\partial q}{\partial R_1}$ 是一个很小的正数)。

那么人们发帖的动机或许不那么强, 此时会看到 R_1 和 Y_1^* 或 Y_2^* 负相关的关系。

情形 2: Y_1^* 和 Y_2^* 如何随着 R_2 的变化而变化? 这里评价用户的原创或转发是如何随着好友的转发帖数的变化而变化的。从式子 (12) 我们

可以看到, 这等价于评价 $\frac{\partial Y_1^*}{\partial R_2}$ 和 $\frac{\partial Y_2^*}{\partial R_2}$ 的取值情况。

根据假设 (A₂) 和 (A₃), 可以得知 $\frac{\partial q}{\partial R_2}$

$\left(\frac{\partial C_1}{\partial q} + \frac{\partial C_2}{\partial q} \right)$ 取值为正。效用部分的符号和 λ 的符号都为正。因此为了保证不等式 (12) 右边为

负, $\frac{\partial Y_1^*}{\partial R_2}$ 和 $\frac{\partial Y_2^*}{\partial R_2}$ 当中至少保证有一个取值为负。

例如, 如果网络中充斥着很多重复的转发帖, 网络进化的动力就会变小, 此时用户没有任何发帖的动机, 所以发帖数量会减少。这也可以解释为太多的重复信息占据了用户的阅读时间, 从而挤占了其用于创作的时间。因此在这种情形下, 随着 R_2 的增加, 社交平台上用户的发帖动机减少。

3 实证结果

3.1 数据收集

实证分析的数据来自新浪微博^①，基于研究目的，本文收集了一个包含动态网络结构的固定样本组时间序列数据。为了获取固定样本组用户，本文采取滚雪球抽样的方法。由于网络结构的稀疏性，任意两个个体间联系的概率几乎为 0，如果采用随机抽样，抽出的将会是独立的个体，而不具有网络结构，这样的样本无法满足研究的需要。而滚雪球抽样能在一定程度上保留抽出个体的网络结构。具体的实施程序如下：首先，我们从微博上某官方账号的粉丝中随机抽取了一批用户 ID，以此为初始抽样种子，运用滚雪球的方法，每次抽取上一批用户粉丝的 ID，去除重复的 ID。如此反复，最终形成了一个 8340 个用户的网络结构，该网络结构的密度为 1.09%。以此作为本研究的固定样本组，从 2013 年 11 月 11 日开始一直到 2014 年 3 月 20 日持续记录这些用户的以下信息：①网络结构信息，即用户之间的关注与被关注的关系；②用户的个人信息，如昵称、粉丝数、好友数、发帖数、账号创建时间、地理位置、性别以及个人描述；③微博文本信息，即记录他们每天发微博（原创与转发）的文本内容。数据通过网络传输，由于新浪服务器的原因或者当时的网络状况，一些请求的丢失都是正常的。这就出现了每天的观测人数不会正好是 8340，最终收集到的用于本研究的数据有 947 662 个，包含所有用户 116 天的观测数据^②。

由于本文的研究目的是分析个体用户发帖（原创与转发）的动机，因此有必要删去公众账号的用户（如一些企业的官方微博），因为这些

账号的发帖动机不同于普通用户，因此不应该被包含在研究中。为此，项目组聘请了 40 个助理（助理研究员的简称）去手动检查每一个用户的个人描述，如果出现了诸如“某某公司官方微博”的字样，即认为这是一个公众账号，会将其剔除。最终，留下用于分析的样本为 862 118 个，横跨 116 天。

3.2 变量构建

根据本文的理论模型并结合实证研究的数据，在本小节我们将详细叙述内在效用、形象效用和交互效用相关变量的计算方法，并给出控制变量的计算方法。

①内在效用相关的变量。内在效用相关的变量可以操作化为和用户个体特征相关的一些变量。本文考虑了四个和用户特征相关的变量：第一个变量为用户的入网时长，记录了用户从注册微博开始到收集数据为止使用微博的时间，单位为天。第二个变量为性别。第三个变量为是否微博认证，即一个用户是否通过了微博的官方认证。最后一个变量是地理位置，即在注册时用户填写的地理位置信息，这里只能记录地址是否为北京。

②形象效用相关的变量。Toubia 和 Stephen (2013) 将用户的粉丝数间接地表示为形象效用的大小。本研究也使用这样的设定，并且只考虑纯粹粉丝数，删除了互粉的人数，因为对于互粉的人来说，用户除了可以获得形象效用之外，还可能获得其他的效用。第二个变量为是否为名人，本文收集了 2013 年和 2014 年两年的中国名人福布斯排行榜里的名人名单，用它作为一个客观标准评判样本里的微博用户是不是名人，该变量为 0~1 变量。

^①2013 年年底，得益于和新浪的合作，新浪微博向我们开放了 API 接口（Application Programming Interface），使我们可以爬取程序自动获取想要的数据库。

^②注：这里的观测数并不等于用户数乘以观测天数，这是因为一些请求的失败，导致有些用户并没有完整的 116 天的记录。

③交互效用相关的变量。将用户前一期发帖的被转发数近似代表用户和粉丝之间的互动，如果用户发的帖子总是被关注被转发，那么下次发帖的动机就会增强。除此之外，交互效用也和用户关注的好友发帖（原创与转发）数有关。因此，交互效用可以用三个变量来测量，分别是上一期用户发的帖子被转发数、用户关注的好友的原创帖数、转发帖数。

④控制变量。除了以上介绍的关键变量之外，本文还考虑了四个控制变量，首先是前一天用户发的原创帖数和转发帖数，用以衡量用户发帖的惯性效应。第三个控制变量为是否为节庆日，根据收集数据的时间，将圣诞节（2013年12月24、25、26），元旦（2014年1月1日）和春节（2014年1月30至2月6日）这几天定义为节庆日，否则为非节庆日。最后一个控制变量是网络密度，网络密度十分依赖抽样方式的选择，为了避免抽样方式带来的误差，我们计算了每天的网络密度，将其作为控制变量放到模型里^①。具体的变量说明如表1所示。

表1 变量说明

	变量名称	类型
控制变量	上一期原创帖数	连续型
	上一期转发帖数	连续型
	是否节日	1=节日, 0=非节日
	网络密度	连续型
内在效用相关变量	入网时长	连续型
	性别	1=男性, 0=女性
	是否微博认证	1=认证, 0=非认证
形象效用相关变量	地理位置	1=北京, 0=非北京
	粉丝数	连续型
交互效用相关变量	是否为名人	1=名人, 0=非名人
	上一期被转帖数	连续型
	好友的原创帖数	连续型
	好友的转发帖数	连续型

3.3 描述性分析

在具体建模之前，本小节对各个变量进行描述性分析。首先，对不随时间变化的个体指标统计如下：样本中微博认证的用户占比为43.7%，男性占比为65.2%，名人用户的占比为1.4%，地理位置为北京的用户占比为58.8%，微博用户的平均入网时长3.28年。其次，对于随时间变化的个体指标（主要是计数数据），直接对原始变量进行描述统计，这些变量包括：原创帖数、转发帖数、被转发帖数、好友原创帖数和好友转发帖数。结果详见表2。再次，网络密度也是一个随时间变化的变量，经过计算我们发现，在本数据集中，平均网络密度为1.011%，标准差为0.0003%，这说明平均每天新增或取消的关注数变化极小，网络密度变化比较平稳。最后，为了进一步展示原创帖数和转发帖数的情况，我们对每人每天的发帖数进行了频数统计，如表3所示。

表2 随时间变化的个体指标的描述性统计分析

变量名称	均值	标准差	最小值	最大值
原创帖数	0.419	2.022	0	50
转发帖数	0.677	2.655	0	50
被转发帖数	115.254	2192.952	0	465223
好友原创帖数	225.779	237.000	0	2597
好友转发帖数	172.912	172.583	0	2067

表3 原创帖数和转发帖数统计

发帖数	原创帖数		转发帖数		
	频数	占比	转发帖数	频数	占比
0	739 656	85.80%	0	700 724	81.28%
1	66 904	7.76%	1	70 783	8.21%
2	22 250	2.58%	2	30 701	3.56%
3	10 173	1.18%	3	16 663	1.93%
4	5 904	0.68%	4	10 214	1.18%
5	3 810	0.44%	5	6 966	0.81%
6	2 359	0.27%	6	4 768	0.55%
7	1 642	0.19%	7	3 617	0.42%
8	1 251	0.15%	8	2 770	0.32%
9	1 061	0.12%	9	2 134	0.25%
≥10	7 108	0.82%	≥10	12 778	1.48%

^①网络密度=网络中已有关注关系的边的个数/所有可能边的个数。

从表 2 的分析结果我们可以看出：第一，无论是原创帖还是转发帖，均值都不超过 1，这说明平均每人每天的发帖量很小，大部分人都处在不发帖的状态。相较于均值，标准差比较大，说明用户之间的发帖行为存在比较大的差异。进一步地，从表 3 中我们可以看出发帖数的分布情况，有超过 80% 的用户是不发帖的，说明发帖数（原创和转发）是一个严重的右偏分布。第二，被转发帖数的均值为 115，最大值达到 465223，而最小值只有 0，说明热帖可以越来越热，而有些帖子却永远无人问津。从标准差中我们也可以看出，帖子能否被转发千差万别。最后，好友的原创帖数平均来说大于好友转发帖数。

3.4 模型结果

根据理论模型的设定，本文探究的是发帖数的影响因素。发帖数可以看成是一种计数数据，在营销研究中，通常用泊松回归（Poisson）或负二项回归（Negative Binomial Regression）来对计数数据进行建模。但是在使用时，二者又有一定的区别，相较于负二项回归，泊松回归有更多的限制，尤其是期望等于方差这一假设，而事实上大多数应用都很难满足这一条件，如果被解释变量的方差大于期望，则存在过度分散（Overdispersion）的情况，此时需要用负二项回归来进行建模。因此在实证研究中，我们首先对原创帖数和转发帖数分别进行负二项回归，根据似然比检验来判断是否可以进行泊松回归。回归结果如表 4 所示，其中 LR （似然比）检验显著（ $p < 0.001$ ），说明存在过度分散，因此只需要考虑负二项回归即可。

从表 4 中我们可以得出以下结论：首先，看因变量是否为原创的情形，对于和内在效用相关的变量，微博使用时间越长的用户越容易发帖。和女性相比，男性更愿意发帖，微博认证用户比非认证用户更愿意发帖，地理位置上没有显著差异。对于和形象效用相关的变量，粉丝数多的用

户更愿意发帖；与普通用户相比，名人更不愿意发帖。对于和交互效用相关的变量，上一期被转发帖数越多，发原创帖的意愿越强，好友的原创帖数会促进用户的原创帖，而好友的转发帖数会抑制用户的原创帖数。这一现象可以用时间成本有限的理论来解释，因为用户花在社交网络上的时间是有限的，好友的转发占据了阅读时间，因此用于创作的时间就会相对减少，因此会产生抑制的效应。其次，再来看因变量是否为转发的情形。对于和内在效用相关的变量，各个变量的结论与因变量和是否为原创时一致。对于和形象效用相关的变量，是否为名人这个变量变成显著为正，相较于普通用户，名人更愿意转发，这也从侧面印证了之所以很多大 V 都被用作营销账号，是因为他们的转发效应。对于和交互效用相关的变量，好友的转发帖促进了用户的转发（该结果不同于因变量是否为原创），因为转发的成本很低，用户和用户的好友在很大概率下会关注共同感兴趣的话题，因此好友的转发会极大地促进用户的转发行为。对于控制变量，用户前一期的原创帖数（转发帖数）越多，越能促进他们发原创帖和转发帖的行为。对于节日效应，研究发现相较于平日，节假日的原创增多，而转发减少，这说明在节假日用户比较在意原创内容。网络密度对原创和转发都具有显著作用，越稠密的网络，用户越容易发帖。

表 4 负二项回归结果

	变量名	是否原创	是否转发
控制变量	前一期原创帖数	1.520*** (0.006)	0.462*** (0.007)
	前一期转发帖数	0.447*** (0.005)	1.445*** (0.005)
	是否节日	0.043*** (0.011)	-0.057*** (0.010)
	网络密度	35.188*** (9.633)	18.872* (8.803)

续表

	变量名	是否原创	是否转发
内在效用 相关变量	入网时长	0.173*** (0.012)	0.196*** (0.011)
	性别-男性	0.016* (0.007)	0.050*** (0.006)
	是否微博认证	0.387*** (0.007)	0.156*** (0.006)
	注册地是否 为北京	0.009 (0.006)	0.087*** (0.006)
形象效用 相关变量	粉丝数	0.062*** (0.002)	0.005** (0.002)
	是否为名人	-0.118*** (0.023)	0.220*** (0.023)
交互效用 相关变量	被转发帖数	0.045*** (0.003)	0.023*** (0.003)
	好友原创帖数	0.049*** (0.005)	-0.063* (0.005)
	好友转发帖数	-0.016** (0.006)	0.271*** (0.006)
常数项	Cons (常数项)	-4.165*** (0.133)	-4.337*** (0.122)
	时间的固定效应	加入	加入
	观测个数	862 118	
似然比检验 $\alpha=0$, P value = 0			

注：显著性水平：0.001 “***”；0.01 “**”；0.05 “*”；0.1 “.”。

4 结论

4.1 理论贡献与营销意义

本文提出了一个综合全面的基于效用理论解释用户发帖动机的理论模型。在提出的效用理论模型中，除了之前研究的内在效用和形象效用，本文提出了交互效用。交互效用指用户通过与其好友进行互动而获得的效用。理解交互效用驱动的动力在理论和实际应用中都具有重要的意义。首先从理论上，本文丰富了用户发帖动机的效用理论，利用交互效用可以解释人们使用社交网站，主要的动机是为了和他人进行互动。在实践中，理解交互效用的作用可以帮助营销人员更好

地设计营销方案。交互效用动机表明用户之所以喜欢在社交平台上发帖，是因为希望与他人交流并且扩大自己社交圈。因此平台运营者可以对一群同质的用户进行集体营销，这会带来事半功倍的效果。以往的营销行为往往基于个体，由于有交互效用的存在，群体刺激也非常重要。

本文的第二个贡献是讨论了用户是如何在发帖与阅读帖子之间进行权衡的。之前的研究都把关注点放在了发帖上，而忽略了发帖时需要的成本。一个用户花在社交平台上的时间是有限的，这意味着需要在发帖与阅读帖子之间做出权衡。因此，在优化用户效用函数时，本文引入了时间成本约束的限制，可以评价不同情形下的最优选择。这对管理人员具有重要的实践意义。例如，本文发现好友的发帖数对用户的发帖行为有抑制作用，这说明帖子占用了用户太多的阅读时间，从而导致用户没有时间留给自己发帖。正因为每个用户都有这样一个约束条件，因此为了鼓励人们更多地发帖，企业实践者应该密切关注平台上帖子的数量和质量，合理有效地管理这些帖子可以带来更大的客户价值。例如，一种可能的方式是让用户尽可能多地接触好的且有限的资源环境，如一些自我创建的小组，这样用户发帖会更加积极。

4.2 局限性与未来研究展望

本文的研究仍然存在一些不足，希望在今后的研究中可以改进。首先，本文的数据来自固定样本组的局部网络结构数据，并没有观察到全网的数据。本文用局部网络数据来近似全网数据，这是研究的一个局限。在未来，我们可以获取更加全面的网络结构数据，但是在现实中，想要获得全网的数据显然是不现实的，并且也很难做到，只能尽可能地用比较全面的网络数据来做近似。其次，本文重点关注了交互效用驱动下的发帖动机，然而用户的发帖动机是多种多样的，除了之前文献提到的内在效用和形象效用动机，

以及本文提出的交互效用动机, 还有其他效用驱动的动机未被探索。在未来, 应该在这方面做进一步的探索, 挖掘出更多用户发帖背后的动机。

参考文献

- [1] 施卓敏, 陈永佳, 赖连胜. 2015. 网络面子意识的探究及其对社交网络口碑传播意愿的影响 [J]. 营销科学学报, 11 (2), 133-151.
- [2] 周志民, 贺和平, 苏晨汀, 周南. 2011. 在线品牌社群中 E-社会资本的形成机制研究 [J]. 营销科学学报, 7 (2), 1-22.
- [3] Dover, Y., Goldenberg, J., and Shapira, D. 2012. Network traces on penetration: Uncovering degree distribution from adoption data [J]. *Marketing Science*, 31 (4), 689-712.
- [4] Fader, P. S., and Winer, R. S. 2012. Introduction to the special issue on the emergence and impact of user-generated content [J]. *Marketing Science*, 31 (3), 369-371.
- [5] Goel, S., and Goldstein, D. G. 2013. Predicting individual behavior with social networks [J]. *Marketing Science*, 33 (1), 82-93.
- [6] Goffman E. 1983. The interaction order [J]. *American Sociological Review*, 48 (1), 1-17.
- [7] Hasan, S., and Bagde, S. 2015. Peers and network growth: Evidence from a natural experiment [J]. *Management Science*, 61 (10), 2536-2547.
- [8] Hu, Y., and Van den Bulte, C. 2014. Nonmonotonic status effects in new product adoption [J]. *Marketing Science*, 33 (4), 509-533.
- [9] Iyengar, R., Van den Bulte, C., and Lee, J. Y. 2015. Social contagion in new product trial and repeat [J]. *Marketing Science*, 34 (3), 408-429.
- [10] Iyengar, R., Van den Bulte, C., and Valente, T. W. 2011. Opinion leadership and social contagion in new product diffusion [J]. *Marketing Science*, 30 (2), 195-212.
- [11] Iyer, G., and Katona, Z. 2016. Competing for attention in social communication markets [J]. *Management Science*, 62 (8), 2304-2320.
- [12] Narayan, V., Rao, V. R., and Saunders, C. 2011. How peer influence affects attribute preferences: A bayesian updating mechanism [J]. *Marketing Science*, 30 (2), 368-384.
- [13] Nitzan, I., and Libai, B. 2011. Social effects on customer retention [J]. *Journal of Marketing*, 75 (6), 24-38.
- [14] Risselada, H., Verhoef, P. C., and Bijmolt, T. H. 2014. Dynamic effects of social influence and direct marketing on the adoption of high-technology products [J]. *Journal of Marketing*, 78 (2), 52-68.
- [15] Ryan, R. M., and Deci, E. L. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions [J]. *Contemporary Educational Psychology*, 25 (1), 54-67.
- [16] Shriver, S. K., Nair, H. S., and Hofstetter, R. 2013. Social ties and user-generated content: Evidence from an online social network [J]. *Management Science*, 59 (6), 1425-1443.
- [17] Strebler P. 1987. Organizing for innovation over an industry cycle [J]. *Strategic Management Journal*, 8 (2), 117-124.
- [18] Toubia, O., and Stephen, A. T. 2013. Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? [J]. *Marketing Science*, 32 (3), 368-392.
- [19] Wang, J., Aribarg, A., and Atchad'e, Y. F. 2013. Modeling choice interdependence in a social network [J]. *Marketing Science*, 32 (6), 977-997.
- [20] Wei, Y., Yildirim, P., Van den Bulte, C., and Dellarocas, C. 2016. Credit scoring with social network data [J]. *Marketing Science*, 35 (2), 234-258.

Tweet or Retweet——Interaction Utility Derived from User Generated Content in Social Media

Zhou Jing^①, Shen Qiaowei^②, Tu Ping^②, Wang Hansheng^②

(^① School of Statistics, Renmin University of China

^② Guanghua School of Management, Peking University)

Abstract User generated content (UGC) is becoming a dominating way for people to communicate and share with each other on social platforms. For platform providers, a good UGC performance can attract more advertisers and directly influence their revenue. Therefore, how to encourage people to contribute content becomes a problem of interest. However, the underlying motivation on user generated content is still not well understood. Therefore, we propose in this paper a more comprehensive utility framework. Combining with previous intrinsic and image-related utility, we propose an interaction utility concept to describe the utility that users derive from interacting with their friends. Then incorporating interaction utility is our first contribution in this paper. As our second contribution, we exert a time budget constraint in our utility framework and this enables us to evaluate the tradeoff between generating and consuming content. Finally, a Sina Weibo dataset is used for illustration.

Key words social networks; interaction utility; user generated content; negative binomial regression