

C2C 环境中顾客重复购买行为的实证与建模^①

叶作亮¹, 王雪乔², 宝智红¹, 陈滨桐^{1,3}

(1. 西南财经大学工商管理学院, 成都 610074; 2. 西南财经大学经济与管理研究院, 成都 610074;
3. 特拉华州立大学工商管理系, 美国 多佛)

摘要: 通过分析 C2C 商家的交易记录, 发现单个商家的顾客重复购买次数呈幂律分布, 这与重复购买理论以往得到的购买次数为负二项分布的结论不同. 基于 C2C 交易的实际情况, 建立了具有购买强化效应的购买概率模型, 模型很好的解释了顾客重复购买率呈幂律分布的现象. 实证和模型表明, C2C 环境中, 顾客购买次数的分布具有很长的尾部, 大量顾客有低的购买次数, 而少数顾客有很高的重复购买次数; 模型还揭示了巨大的潜在顾客数量和顾客多次购买的强化效应是造成重复购买分布呈幂律形式的重要原因.

关键词: 重复购买行为; 幂律分布; 电子商务; 购物行为; 二项分布模型

中图分类号: F27 **文献标识码:** A **文章编号:** 1007-9807(2011)12-0071-08

0 引言

随着支付手段和物流配送的完善, 网络购物已经成为重要的消费方式. 而研究表明消费者在线购买与线下购买的行为有明显差异, 其原因是购买环境、风险、成本和消费者习惯都发生了变化^[1-2]. 首先, 顾客在网上的可选范围更大, 成熟的 C2C 平台往往拥有海量的产品和众多的商家; 其次, 交易是在网络空间完成, 顾客面临更大购物风险; 最后, 顾客的转换成本很低, 顾客要另选商家几乎不需要付出选择成本.

在上述因素中, 面临的多种风险促使顾客忠诚于可信赖的商家, 但是更大的可选范围和低的转换成本又便于顾客另选商家, 由此形成两个相反的因素. 顾客信赖和重复购买行为对商家非常重要, 一般认为拥有高数量的回头客, 那么商家就具有了较好的顾客忠诚度^[3-4].

到底顾客重复购买情况如何, 对重复购买理论(RBT)进行过深入研究^[5]. 从商家的角度, 有

多少顾客会成为回头客是顾客的多次购买选择行为的结果, 即在多次的买与不买的决策事件中确定了顾客的重复购买行为. 重复购买理论正是从这一个角度来建立分析模型. 采用多次伯努利事件来建立顾客购买的概率模型是自然会想到的方法. RBT 考虑的是一个家庭 (household) 对某一品牌产品的购买情况, 首先假定购买事件是泊松流, 而不同顾客之间的差异用伽玛分布来描述, 这样就得到一段时期顾客重复购买的负二项分布模型 (NBD)^[5].

分析 NBD 模型发现, 当潜在顾客数量很大而顾客购买概率比较小时, NBD 曲线非常偏向纵轴, 而采用只需要 1 个参数的对数级数分布 (LSD) 来拟合也非常有效^[6]. RBT 把顾客购买事件简化为泊松过程, 因此 NBD 模型要求每次购买事件是独立的, 且购买概率是常数. 虽然这些限制较为严格, 并和实际情况有出入, 但是仍然很好地解释了线下的顾客购买现象^[6-8], 并能够和实

① 收稿日期: 2009-12-25; 修订日期: 2010-09-01.

基金项目: 国家自然科学基金资助项目 (70801050); 教育部人文社会科学研究青年资助项目 (07JC630018); 西南财经大学“211 工程”3 期资助项目.

作者简介: 叶作亮 (1975-), 男, 四川金堂人, 博士, 副教授. Email: yezuoiliang@163.com

证数据很好地吻合. 因此后续以 NBD 模型为基础衍生出一系列相关研究与应用^[9-11].

然而在电子商务环境中, 特别是 C2C 情况下, 购买环境和线下购买有显著不同, 一是失去地理的约束, 商家的潜在顾客数量非常巨大; 二是由于风险和信任因素, 顾客上次购买事件会显著地影响到下一次的购买决策. 因此购买的概率不应该是常数; 三是由于低转换成本, 顾客转换到其他商家的可能性增加. 这导致 C2C 顾客重复购买的实际情况可能和线下的重复购买模型存在差异. 由此, 本文对淘宝及 Ebay 等 C2C 平台上的交易数据进行了统计分析, 发现在 C2C 交易中, 某一商家的顾客重复购买次数分布既不是 NBD 分布, 也不是 LSD 分布, 而是呈幂律分布 (power law distribution, PLD).

幂律分布是极不均匀的分布形态. 对服从幂律分布的系统特征的研究表明, 幂律分布是系统规模基本不变、边界确定下的宏观特征, 它对微观结构的改变非常敏感^[12]. 从财富的分布、公司和城市大小的分布, 幂律分布广泛地存在自然和社会系统中^[13-15]. 虽然学者对幂律分布有深入的研究, 但是幂律形成的内在机制并没有得到统一的解释. 文献 [16] 分析表明, 多个体的重复选择行为会导致结果呈幂律分布, 原因是个体每次都会做出其本人视野范围内自认为最有利的决策. 严格的 Zipf 律是一种特殊的幂律分布^[17], 产生的机制被认为是人们希望通过最小的努力来达到期望的目标. 这两种解释都可以用来理解 C2C 交易中的购买次数的幂律分布成因: 1) 顾客的视野是有限的, 难以获得 C2C 平台中的全部相关产品, 可猜想他们会通过最小的努力来获得期望的信息; 2) 顾客会在所有相关产品中做出最有利的决策.

Simon^[18] 提出的 Simon 模型是较通用的幂律生成模型, 模型体现了“成功酝酿成功”的思想, 这使 Simon 模型是当前对幂律解释力最强的基础模型^[19-20]. Stefan 等^[21] 用 Simon 模型来审视互联网上的数量规律; Newman^[22] 也把 Simon 模型作为最基本的幂律产生机制.

综上所述, RBT 理论是对购买行为的直接描述, 便于理解顾客购买行为且能解释很多实际交

易记录, 但是无论是 NBD 还是 LSD 分布都不能够描述 C2C 交易中的重复购买行为. 本文从实证出发, 根据幂律分布相关研究, 通过建立封闭的多用户重复决策行为模型来分析 C2C 商家交易中顾客重复购买的幂律分布及形成机制, 并通过实证与仿真的方法对模型进行检验, 最后讨论模型参数的确定以及模型所揭示的现实意义.

1 实证分析

1.1 实证数据

实证数据源于中国最大的 C2C 电子商务网站淘宝网. 按照成交排名在 32 个产品大类中选取了交易最大的 3 类典型产品: 数码相机、女装、游戏点卡做为统计对象, 每类产品随机抽取 70 个商家, 共计 210 个样本. 为了准确统计重复购买次数, 对样本数据进行如下处理.

1) 伪重复购买. 顾客如果 1 次购买多个产品, 淘宝往往会显示每个产品的交易记录, 这些交易记录并不是真正的重复购买, 因此本文把同一天内顾客的多次交易记录只计 1 次.

2) 重复购买的稳定性. 定义总体重复购买率 R 为购买两次及以上的交易量与总交易量的比值. 对已获得的样本数据进行分析, 发现只有总交易量达到一定规模 R 才趋于稳定, 如图 1 所示. 因此, 剔除交易量未达到 2 000 次的样本数据. 最后得到可用样本 200 个, 其中服装类 62 个, 数码产品类 68 个, 游戏点卡类 70 个.

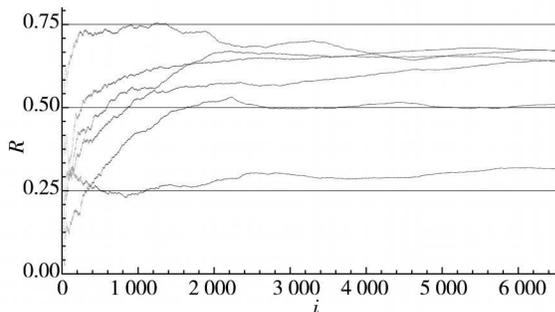


图1 C2C 商家总体重复购买率

Fig. 1 Total repeat buying rate in the sales records

1.2 数据分析

以某典型的 C2C 交易记录为例 (标记为 TF0030), 如表 1 所示. 该商家总交易为 5 231

次,总购买人数为 2 166 人,其中 1 390 人 (64.17%) 购买 1 次,购买 5 次以上的顾客只有 200 人 (9.23%),但是他们实现的交易次数为 2 272 次 (43.43%),在总交易中占有很大的比例.

表 1 典型 C2C 商家顾客重复购买分布数据

Table 1 Distribution data of a typical case

购买次数 i	重复购买次数 $i-1$	购买人数 $f(i)$	人数所占百分比
1	0	1 390	64.17
2	1	316	14.59
3	2	141	6.51
4	3	81	3.74
5	4	38	1.75
6	5	32	1.48
7	6	36	1.66
8	7	17	0.78
9	8	20	0.92
10	9	12	0.55
11	10	13	0.60
12	11	10	0.46
13	12	12	0.55
14	13	6	0.28
15	14	8	0.37
16	15	6	0.28
17	16	4	0.18
19	18	3	0.14
20	19	3	0.14
21	20	4	0.18
22	21	3	0.14
23	22	2	0.09
24	23	1	0.05
25	24	2	0.09
28	27	1	0.05
31	30	1	0.05
32	31	1	0.05
33	32	1	0.05
43	42	1	0.05
51	50	1	0.05

绘制表 1 的购买次数分布图,得到一条向右倾斜得很厉害分布曲线,这表明 C2C 商家重复购买分布可能服从幂律分布,如图 2 所示.

为了确定曲线的分布形式,将表 1 中的数据剔除重复购买次数为零的数据后,得到以购买次数 i 升序排序的顾客人数 $f(i)$,取对数后得到对数坐标系下重复购买次数与购买人数的关系,见图 3. 顾客购买次数 i 和购买人数 $f(i)$ 在对数坐标系下呈现线性关系,对数据进行回归分析,结果显示线性模型拟合较好,拟合优度 $R^2 = 0.974$,非常接近于 1,表明模型对样本数据的拟合优度非常高. T 检验和 F 检验都非常理想,显著水平为 0,拟合直线方程式为 $y = 3.113 - 2.002x$. 因此表明 $\log f(i)$ 与 $\log i$ 满足负线性关系,也即在双对数坐标下,C2C 商家的顾客重复购买分布服从幂律分布,表现为一条斜率为幂指数 $\rho = -2.002$ 的直线

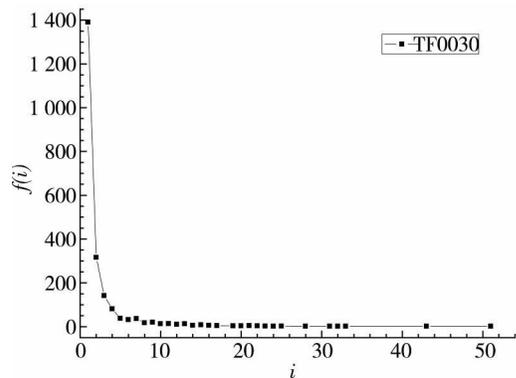


图 2 某 C2C 商家顾客购买次数分布

Fig. 2 Quantity of repeat buying distribution for a seller

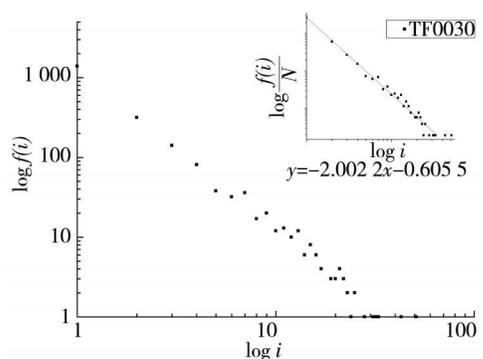


图 3 某 C2C 商家双对数坐标下顾客重复购买人数分布 (右上角为数据进行归一化处理后的数据)

Fig. 3 Quantity of repeat buying distribution in logarithm coordinate.

(Top-right of the figure illustrates normalized value)

对取得的样本研究发现 C2C 商家重复购买都服从这一分布规律, ρ 值集中于 1.4 - 3.8 之间,拟合程度非常高. 部分拟合结果如表 2 所示.

表 2 不同类型商家购买分布的幂律指数 (部分数据)

Table 2 Power index for different types of sellers

	C2C 商家	幂指数 ρ	截距 c	拟合优度
服装类	TF0001	2.917	3.506	0.979
	TF0017	2.948	3.473	0.971
	TF0013	2.511	3.118	0.960
	TF0023	3.541	3.709	0.973
	TF0021	2.292	3.050	0.929
数码产品类	TS0008	3.197	3.532	0.959
	TS0026	2.192	3.008	0.864
	TS0031	2.705	3.265	0.920
	TS0007	2.767	3.192	0.921
	TS0030	2.240	2.996	0.862
游戏点卡类	TD0014	3.138	3.536	0.994
	TD0035	2.994	3.607	0.991
	TD0008	2.766	3.502	0.977
	TD0024	2.608	3.251	0.921
	TD0002	2.557	3.440	0.912

数据表明: 不同的 C2C 商家重复购买分布有着相近的 ρ 值; $\log f(i)$ 与 $\log i$ 呈现显著的线性关系.

2 顾客购买行为建模

2.1 符号定义和基本假设

RBT 建模要求重复购买事件之间必须独立, 但研究表明“记忆性”会支配用户网上行为^[23], 在建模时应考虑顾客以前的购买对再次购买概率的影响, 需要新的建模思路. 社会心理学中的社会交易理论认为, 决定个人关系持续倾向的因素包括关系满意水平、关系投资规模、习惯性思考等“购后效应”^[24], 因此, 可假设 C2C 顾客的多次购买行为是建立在以前购买经历的基础上, 即顾客在同一店铺的购买概率随着购买次数的增加而增大.

设在 C2C 交易平台中, 任意商家 E 有 n 个潜在顾客, 每个顾客有相同的机会光顾该商家的店铺, 顾客每次访问店铺时有 p_i 的概率产生购买行为. 商家 E 的所有销售正是由于 n 个顾客重复这样的行为而产生. 在一段时间后, 顾客中购买 i 次的顾客人数 $f(i)$ 形成一种分布, 可以描述顾客的

重复购买行为特征.

模型的符号表示:

n 为系统中潜在顾客数;

k 为迭代次数, k 每增加 1 表示在 n 个顾客中随机地选出 1 个顾客光顾店铺;

f_k^i 为系统中迭代到 k 次时, 购买次数为 i 的人数;

$p_i = p(i)$ 为已经购买了 i 次的顾客可能再次购买的概率.

建模的目的就是求解 f_k^i , 得到 f_k^i 就得到了购买次数的分布. 相关假设包括:

- 1) 假设整个过程中顾客总数 n 保持不变;
- 2) 当迭代次数 k 较大时, f_k^i 的分布是稳定的, 且和 k 无关.

3) 根据前面的分析并结合文献 [23] [24] 的研究, 假设 $p_i = \alpha_i + \alpha_0$.

2.2 建模过程

1) 每次从 n 个顾客中随机选取 1 个顾客, 顾客被选中的概率为 $1/n$;

2) 由于购买了 i 次的顾客人数是 f_k^i , 因此被选中者是购买了 i 次的顾客的概率是 $\frac{f_k^i}{n}$;

3) 当由 $k \rightarrow k+1$ 时, 从 n 个顾客中随机抽取 1 个顾客, 考虑系统中 f_k^i 的变化:

① 如果选中的是购买了 $i-1$ 次的顾客, 且以 p_{i-1} 的概率产生购买行为, 系统中购买了 i 次的顾客以 $\frac{f_k^{i-1}}{n} p_{i-1}$ 的概率增加 1 个人.

② 如果选中的是购买了 i 次的顾客, 且以 p_i 的概率产生购买行为, 则系统中购买了 i 次的顾客以 $\frac{f_k^i}{n} p_i$ 的概率减少 1 个人.

③ 其他情况不影响 f_k^i 的变化. 因此得到下面的递推关系

$$f_{k+1}^i = f_k^i + \frac{f_k^{i-1}}{n} p_{i-1} - \frac{f_k^i}{n} p_i, \quad (i > 1, k > 0) \quad (1)$$

$$f_{k+1}^1 = f_k^1 + \frac{f_k^0}{n} p_0 - \frac{f_k^1}{n} p_1, \quad (k > 0) \quad (2)$$

2.3 模型求解

式 (1) 和 (2) 与 Simon^[18] 研究 Zipf 律的表达式类似, 根据 Simon 模型中提出的稳态假设, 考虑

重复购买次数随 k 增加的稳定性, 得到如下关系

$$\frac{f_{k+1}^i}{f_k^i} = \frac{k+1}{k} \quad (3)$$

代入式 (1) 得

$$f_k^i = \frac{kp_{i-1}}{n+kp_i} f_k^{i-1} \quad (4)$$

设 $p_i = \alpha_i + \alpha_0$, 得

$$f_k^i = \frac{1 + \frac{\alpha_0}{\alpha} - 1}{\frac{n}{\alpha k} + i + \frac{\alpha_0}{\alpha}} f_k^{i-1}$$

当 k 很大时 f_k^i/k 分布稳定且与 k 无关, 记 f_k^i/k 为 $f(i)$, 则

$$f(i) = \frac{i + \frac{\alpha_0}{\alpha} - 1}{\rho_1 + i + \frac{\alpha_0}{\alpha}} f^{i-1}, \quad \rho_1 = \frac{n}{\alpha k}$$

因此

$$f(i) = \frac{(i + \frac{\alpha_0}{\alpha} - 1)(i + \frac{\alpha_0}{\alpha} - 2) \dots (\frac{\alpha_0}{\alpha} + 2)(\frac{\alpha_0}{\alpha} + 1)}{(\rho_1 + i + \frac{\alpha_0}{\alpha})(\rho_1 + i + \frac{\alpha_0}{\alpha} - 1) \dots (\rho_1 + 1 + \frac{\alpha_0}{\alpha})} f(1)$$

即

$$f(i) = (i + \frac{\alpha_0}{\alpha})^{-(\rho_1 + 1)} f(1)$$

令 $\rho = \rho_1 + 1$, 得

$$f(i) = (i + \frac{\alpha_0}{\alpha})^{-\rho} f(1) \quad (5)$$

即 $f(i)$ 为幂律分布函数.

下面求解 $f(1)$. 根据式 (2) 可知, f_k^0 是系统中迭代到 k 次时购买次数为 0 次的顾客人数. 顾客一次都不购买的概率为 $1 - p_0$, 第 k 次时顾客被选中的平均次数是 k/n 次, 所以有

$$f_k^0 = n(1 - p_0)^{k/n}$$

其中, p_0 为常数, 且与 p_i 无关, $0 < p_0 < 1$. 将 f_k^0 及 p_0 代入式 (2), 得到

$$f_k^1 = \frac{p_0(1 - p_0)^{k/n}}{(n + p_1 k)/(nk)}$$

当 $p_i = \alpha_i + \alpha_0$ 时, $\rho = \frac{n}{\alpha k}$, 且有 n 远大于 α 和 α_0 , 得到

$$f(1) = p_0(1 - p_0)^{\frac{1}{\alpha \rho}} \frac{n}{\alpha \rho} \quad (6)$$

3 模型参数的确定

由式 (5) 知

$$\frac{f(i)}{f(1)} = (i + \frac{\alpha_0}{\alpha})^{-\rho}$$

以表 1 为例对模型进行验证, 并确定模型参数, 对于式 (5) 和 (6), 当 $\rho = 2, p_0 = 0.2$ 时, 模型和实证结果一致, 如表 3.

表 3 某 C2C 商家的重复购买分布

Table 3 Distribution data of a typical case

购买次数	$f(i)$		$f(i)/f(1)$	
	模型	实证	模型	实证
1	1 367	1 390	1.000 0	1.000 0
2	342	316	0.250 0	0.227 3
3	152	141	0.111 1	0.101 4
4	85	81	0.062 5	0.058 3
5	55	38	0.040 0	0.027 3
6	38	32	0.027 8	0.023 0
7	28	36	0.020 4	0.025 9
8	21	17	0.015 6	0.012 2
9	17	20	0.012 3	0.014 4
10	14	12	0.010 0	0.008 6
11	11	13	0.008 3	0.009 4
12	9	10	0.006 9	0.007 2
13	8	12	0.005 9	0.008 6
14	7	6	0.005 1	0.004 3
15	6	8	0.004 4	0.005 8
16	5	6	0.003 9	0.004 3
17	5	4	0.003 5	0.002 9
19	4	3	0.002 8	0.002 2
20	3	3	0.002 5	0.002 2
21	3	4	0.002 3	0.002 9
22	3	3	0.002 1	0.002 2
总购买人数	2 155			
模型参数	$\rho = 2, p_0 = 0.2$			

对表 3 中 $f(i)$ 模型理论值与实际观测值进行了卡方检验. 由于卡方检验要求数据期望频次不小于 5, 只取表 3 中购买次数小于等于 17 次的数据, 得到卡方统计值 Q 为

$$Q = \sum_{j=1}^{17} \frac{(\text{实证}_j - \text{模型}_j)^2}{\text{模型}_j} = 16.52$$

因 $\chi^2_{-\alpha}(k) = \chi^2_{.99}(16) = 32.0$, 所以 $Q < \chi^2_{-\alpha}(k)$, 显示不能拒绝原假设, 即 $\rho_1 = 1$ 的模型与实证数据相一致, 能够很好地解释实证数据.

注意到表 3 后两列的数据为 $f(i)/f(1)$, 那么就可以构建模型 $\log \frac{f(i)}{f(1)} = \beta + \beta_1 \log i + \epsilon$ 来验证

重复购买分布模型是否能够很好地解释实证数据. 根据表 3 中的数据, 可得到该商家的顾客购买分布对数回归模型为

$$\log \frac{f(i)}{f(1)} = -0.053 - 1.952 \log i$$

$$\text{修正 } R^2 = 0.985, F = 1295.49$$

从上面的回归结果可以看到模型在取 $\rho_1 = 1, p_0 = 0.2$ 能够很好地拟合表 3 中所示样本店铺的顾客购买分布, 可以解释 98.5% 的数据.

根据实证数据, ρ 值在 1.4 - 3.8 之间, 这时取不同的 ρ 值, 可以得到与实证相吻合的数据分布. 表 4 是不同参数下模型和实证数据的比较. ρ 值不同, 得到的重复购买曲线就不同.

表 4 C2C 商家店铺的重复购买分布

Table 4 Distribution data of a different case

i	TD0003		TF0032		TS0026		TD0009	
	实证	模型	实证	模型	实证	模型	实证	模型
1	2 915	2 894	1 740	1 776	4 542	4 554	3 737	3 718
2	551	548	363	337	508	496	586	613
3	236	207	143	127	140	135	212	214
4	87	104	67	64	45	54	95	101
5	57	61	41	37	22	26	68	57
6	40	39	24	24	15	15	43	35
7	23	27	16	17	11	9	29	24
8	13	20	13	12	6	6	15	17
9	12	15	6	9	6	4	12	12
10	13	12	5	7	4	3	7	9
> 11	54	55	27	26	12	7	49	43
n	4 001		2 456		5 314		4 853	
ρ	2.4		2.4		3.2		2.6	
p_0	0.2		0.4		0.2		0.2	

为进一步确定参数 α 和 α_0 的取值, 对模型进行仿真分析, 过程如下: 设系统中有 n 个初始顾客, 每次迭代随机选择其中 1 个, 选中的顾客以 p_i 的概率产生购买行为, 不购买的概率为 $(1 - p_i)$. 当模型参数设置为 $\alpha = 0.006, \alpha_0 = 0.2$, 即当

$p_i = 0.006i + 0.2$ 时, 仿真能拟合表 3 实证数据, 如图 4 所示. 此时表示顾客的初次购买概率为 0.2, 并以 0.006 的速度趋势增加.

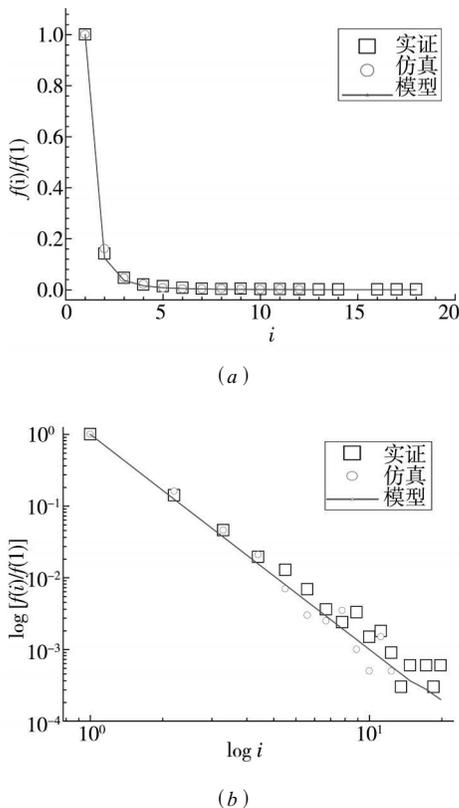


图 4 建模仿真与实证

$$(p_0 = 0.2, \rho = 2, \alpha = 0.006, \alpha_0 = 0.2)$$

Fig.4 Simulation and empirical study of model

4 结束语

PLD 模型能够很好地解释顾客购买次数和购买人数之间的分布关系. 建模分析表明: 在 C2C 商家的潜在顾客随机地光顾商家, 顾客的购买概率随着购买次数的增加而增大的假设下, PLD 模型能很好地拟合实证数据; 模型也表明 C2C 中顾客购买次数分布呈幂律形式, 不同类型的商家都具有相近的幂律指数.

对比 NBD 分布和 LSD 分布, PLD 分布的函数形态更加偏向原点, 这可能是 C2C 交易中, 潜在顾客数量是非常大的, 而购买的总体概率比较低. 另外, 实证数据表明, 购买次数少的人数比例很大, 超过 60% 的人只购买 1 次, 因此 C2C 交易中商家在注重保持和维护客户的忠诚度的

同时,也不能忽视那些购买次数较低的顾客群。

在建模方法上,C2C 交易中的顾客具有购买强化效应,顾客的购买概率 p_i 不应该是常数,而应该为购买次数 i 的函数,反映出了购买强化效应,而这种效应也有利于形成幂律分布。

文献[19]认为互联网是展示幂律行为最佳的场所,产品和顾客的多样性是造成幂律的原因。实际上,造成 C2C 重复购买的幂律分布形态的原

因,到底是顾客的多样性,潜在顾客群体巨大的数量,还是顾客面对的高风险和低转换成本,又或是兼而有之,还需要更深入地研究。另外,今后考虑潜在顾客光顾商家的择优问题,以及深入分析顾客购买决策概率 p_i 的具体形式和影响因素,有利于更好地认识电子商务顾客行为,加强电子商务中的客户关系管理,为电子商务顾客关系管理提供新的理论基础^[25]。

参考文献:

- [1] 黎志成, 刘枚莲. 电子商务环境下的消费者行为研究[J]. 中国管理科学, 2002, 10(6): 21-22.
Li Zhi-cheng, Liu Mei-lian. A study on the consumer behavior in electronic commerce[J]. Chinese Journal of Management Science, 2002, 10(6): 21-22. (in Chinese)
- [2] Lee Pui-Mun. Behavioral model of online purchasers in e-commerce environment[J]. Electronic Commerce Research, 2002, 2(1/2): 75-85.
- [3] Jacoby J, Chestnut R W. Brand Loyalty Measurement and Management[M]. New York: John Wiley And Sons, 1978: 167-175.
- [4] 陈伯成, 叶伟雄, 李英杰. 一类 CRM 模型的构建及其销售策略分析[J]. 管理科学学报, 2007, 10(2): 48-60.
Chen Bo-cheng, Ye Wei-xiong, Li Ying-jie. Kind of CRM modeling and its sales strategy analysis[J]. Journal of Management Sciences in China, 2007, 10(2): 48-60. (in Chinese)
- [5] Chatfield C, Ehrenberg A S C, Goodhard G J. Progress on simplified model of stationary purchasing behaviour[J]. Journal of the Royal Statistical Society, Series A. 1966, 129(3): 317-367.
- [6] Ehrenberg A. Repeat buying[J]. Journal of Empirical Generalisations In Marketing Science, 2000, 5(2): 53-73.
- [7] Schmittlein D C. Counting your customers: Who are they and what will they do next? [J]. Management Science, 1987, 33(1): 1-24.
- [8] Werner R, Kumar V. On the profitability of long life customers in a noncontractual setting: An empirical investigation and implications for marketing[J]. Journal of Marketing, 2000, 64(4): 17-35.
- [9] 马少辉. Pareto/NBD 模型扩展[J]. 系统工程. 2008, 26(8): 123-127.
Ma Shao-hui. The Pareto/NBD model extensions[J]. Systems Engineering, 2008, 26(8): 123-127. (in Chinese)
- [10] Bolton R N, Kannan P K, Bramlett M D. Implications of loyalty program membership and service experiences for customer retention and value[J]. Journal of the Academy of Marketing Science, 2000, 28(1): 95-108.
- [11] Helsen K, Schmittlein D C. Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models [J]. Marketing Science, 1993, 12(4): 395-414.
- [12] Malcai O, Solomon S, Biham O. Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements[J]. Physical Review E, 1999, 60(2): 1299-1303.
- [13] 胡海波, 王林. 幂律分布研究简史[J]. 物理, 2005, 34(12): 889-896.
Hu Hai-bo, Wang Lin. A brief history of power law distributions[J]. Physics, 2005, 34(12): 889-896. (in Chinese)
- [14] Axtell R L. Zipf distribution of US firm sizes[J]. Science, 2001, 293(11): 1818-1820.
- [15] Simon H A. The size distribution of business firms[J]. American Economic Review, 1958, 48(4): 607-617.
- [16] 尉伟峰. 基于选择模式的幂律生成机制[J]. 物理学报, 2009, 58(4): 2127-2136.
Wei Wei-feng. Power law phenomena based on human behavior[J]. Acta Physica Sinica, 2009, 58(4): 2127-2136. (in Chinese)
- [17] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Oxford: Addison-Wesley Press, 1949: 46-77.

- [18] Simon H A. On a class of skew distribution functions [J]. *Biometrika*, 1955, 42 (3/4): 425 – 440.
- [19] 山 石, 邱 红. 长尾分布、幂律的产生机制和西蒙模型[C] // 2008 年中国发展进程中的管理科学与工程(卷 I). 上海: 上海财经大学出版社, 2008: 886 – 890.
Shan Shi, Qiu Hong. Long-tail distribution—the mechanism of power law and Simon model[C] // Proceedings of Management Science and Engineering in China's Development(Part I). Shanghai: Shanghai University of Finacial and Economics Pression, 2008: 886 – 890. (in Chinese)
- [20] Chung K H, Cox R A K. A stochastic model of superstardom: An application of the yule distribution[J]. *The Review of Economics and Statistics*, 1994, 76 (4): 771 – 775.
- [21] Stefan B, Holger E. World wide web scaling exponent from Simon's 1955 mode[J]. *Physical Review E.*, 2001, 64 (3): 035104 – 1/4.
- [22] Newman M E J. Power laws, pareto distributions and Zipfs' law[J]. *Contemporary Physics*, 2005, 46 (5): 323 – 351.
- [23] Hahsler M. Optimizing web sites for customer retention[C] // Proc. of the 2005 International Workshop on Customer Relationship Management: Data Mining Meets Marketing. New York City: Eprint arXiv: 0803.1104, 2005: 18 – 19.
- [24] Thibaut J W, Harold H K. *The Social Psychology of Groups*[M]. New York: John Wiley & Sons. 1956: 238 – 241.
- [25] 齐佳音, 韩新民, 李怀祖. 我国客户关系管理研究的紧迫性和方向分析[J]. *管理科学学报*, 2002, 5 (4): 88 – 95.
Qi Jia-yin, Han Xin-min, Li Huai-zu. Emergency and directions for CRM in China[J]. *Journal of Management Sciences in China*, 2002, 5 (4): 88 – 95. (in Chinese)

Modeling and empirical research of repeat purchase behavior in C2C e-commerce

YE Zuo-liang¹, WANG Xue-qiao², BAO Zhi-hong¹, CHEN Bin-tong^{1, 3}

1. School of Business Administration, Southwestern University of Finance and Economics, Chengdu 610074, China;
2. Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu 610074, China;
3. University of Delaware Civil & Environmental Engineering/Business Administration, USA, Dover

Abstract: By analyzing the C2C sales records, we find the repeat purchase frequency of customers is power-law distribution, which is different from the result of existing repeat purchase research that the purchase frequency is negative binomial distribution. Based on actual situation in C2C transactions, a repeat purchase model with buying reinforcement is built, which explained to the phenomenon of the power-law distribution. Our model and empirical analysis show that, in C2C environment, the distribution of customers' repeat purchase frequency is composed of a long tail and short head, meaning that a large number of customers have a low number of purchases, while a few customers have a high repeat purchase frequency. Our model also reveals that the huge number of potential customers and the repeat purchase with buying reinforcement contributes to the power law distribution characteristics of repeat purchase frequency.

Key words: repeat purchase behavior; power law distribution; e-commerce; shopping behavior; negative binomial distribution