

# The Research of Online Customer Reviews on the Base of Text Analysis

Beibei Wang<sup>1</sup>, Maohong Liu<sup>2</sup>

<sup>1</sup>Economics and Management School, Wuhan University, Wuhan

<sup>2</sup>Management School, Wuhan University of Science and Technology, Wuhan

Email: [wangbeibei562080@126.com](mailto:wangbeibei562080@126.com), [maohongliu@sohu.com](mailto:maohongliu@sohu.com)

Received: Dec. 3<sup>rd</sup>, 2014; revised: Dec. 20<sup>th</sup>, 2014; accepted: Dec. 27<sup>th</sup>, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In the era of Web2.0, customer reviews have become more and more important. Many researchers have done a large number of researches. But they have just focused on the quantitative proxy such as volume and ratings. But the customers care more about the text. By text analysis, we can find more detail from the texts. From the empirical study, we find that volume and the valence of the customer reviews both positively influence sales and they have a positive interaction. We also find that the variance of the customer reviews has a positive influence on the sales. Our findings can expand the understanding of customer reviews and help marketers utilize online word-of-mouth.

## Keywords

Customer Reviews, Text Analysis, Sales

---

# 基于文本分析的网上消费者评论影响机制研究

王贝贝<sup>1</sup>, 刘茂红<sup>2</sup>

<sup>1</sup>武汉大学经济与管理学院, 武汉

<sup>2</sup>武汉科技大学管理学院, 武汉

Email: [wangbeibei562080@126.com](mailto:wangbeibei562080@126.com), [maohongliu@sohu.com](mailto:maohongliu@sohu.com)

收稿日期: 2014年12月3日; 修回日期: 2014年12月20日; 录用日期: 2014年12月27日

## 摘要

在Web2.0时代，消费者评论的作用日益凸显。为此，很多学者对消费者评论的影响做了大量的研究。通常，他们采用消费者评论的量和产品打分作为研究对象。事实上，消费者更关注的是评论中的文本。基于文本分析的方法，可以有效地挖掘出评论的文本中消费者对于产品的真实评价。利用对网上交易社区的数据研究发现，消费者产品评论的量对销量有正向影响，消费者评论的极性对销量有正向作用，且消费者评论量正向调节消费者评论机型对销量的作用。同时，本文消费者对产品不同属性的评价差异性越大，对产品越有利。相关研究结论深化了对消费者评论作用的理解，并为企业运用消费者评论提供了指导。

## 关键词

消费者评论，文本分析，产品销量

## 1. 引言

消费者评论已经逐渐成为消费者网上购物时最倚赖的信息来源了。在传统的市场中，企业是产品信息的来源，营销人员把控着产品信息的源头，有的学者称之为 FGC (firm generated content)。而在 Web2.0 时代，消费者已经厌倦了这种自上而下的信息传播模式，因为用户生成内容已经开始大行其道。根据 eMarketer 2013 年的报告，网上消费者更愿意相信其他消费者产生的 UGC 而不是 FGC [1]。因此，网上的零售商们开始注意到消费者评论的重要性。然而，消费者评论庞大的数量和数据的非结构性是现在对消费者评论进行管理的一大挑战。根据 Econsultancy 2011 年的市场研究报告，受调查的 81% 的网络零售商不知道或者不了解消费者评论的作用机制，因此，有必要对消费者评论的影响机制做一个深入的研究。

现有的关于网上消费者评论的研究都将重点放在了消费者评论的数量和产品打分。很多研究者发现产品被大家讨论的越多即口碑量越大，产品的销量越大。而产品打分，研究者通过不同的角度进行了实证分析，发现产品的平均打分越高，对产品越有利，产品打分的概率分布越分散对产品越有利[2]-[5]。然而这些研究都只注意到了消费者评论的数量特征(评论量和打分)，而没有关注消费者评论中的文本内容。

很多研究者也开始关注评论中的文本内容。研究者们通过简单的文本分析提取出消费者评论文本中的关键词，并从不同的理论基础出发，研究了消费者文本对产品销量的作用机制。他们通过实证研究发现，消费者评论中采用的词汇与产品品牌的个性或产品的广告类型越相似，对产品销量越有利[1]。但是这些研究都只是将评论文本进行了简单的分析，分析了消费者评论文本中涉及到的有关营销的关键词，而并没有分析消费者评论中消费者对于产品具体属性的感知，也未能抓住消费者评论中对产品和品牌的真实态度。

以往的研究受限于测量技术未能深入的挖掘用户生成内容。但是，深入了解消费者对产品和品牌的态度并全面的分析消费者评论的作用，需要挖掘出每一条评论涉及的关于产品的所有属性和其真实情感，对技术要求比较高。本文通过中文文本分析技术，提取出消费者在用户评论中更细节的内容，挖掘出消费者评论的文本中对于产品和品牌的真实态度，并结合 MOA 理论，扩大了对消费者评论影响机制的理解。为了增强研究的实践性，本文选择的是一个交易社区，不同于以往研究采用电影，书籍的数据，我们选择的是手机，增强了研究的丰富性，理论的丰富性。

通过我们的研究，我们发现消费者评论的量对产品的销量有积极作用；消费者评论的有关产品的评价越正面，对产品越有利；同时研究发现消费者评论的量和消费者评论的极性之间存在着显著的正向交互作用，消费者评论中对产品不同属性的评价差别越大，对商家来说越有利。

## 2. 理论基础和假设

根据 MOA (motivation-opportunity-ability)理论, 我们知道消费者评论给消费者提供了一个理解和接触产品信息的机会, 但将这些机会转做购买行为, 消费者需要(1) 有能力对这些信息进行处理; (2) 有足够的动力驱使购买[6]。在网络环境中, 动力是指消费者接受和处理相关信息的意愿; 能力指的是在信息处理过程中的认知资源, 认知技巧和知识获取能力。消费者参与和处理产品相关信息的动机受信息新奇性和可信性的影响。本文从 MOA 理论出发, 从不同角度研究消费者评论对产品绩效的影响。

### 2.1. 消费者评论量研究

消费者评论量(volume)指的是消费者对产品的评论数量, 以往的研究发现产品评论量越大, 产品的销量会越大[7]。根据 MOA 理论, 理论上消费者评论的量越大, 越容易引起其他消费者的注意, 可以提高产品的认知程度, 增加了消费者处理消费者信息的人质资源从而使得消费者购买可能性增加。因此, 本文提出如下假设:

H1: 消费者评论量对产品销量有正向影响。

### 2.2. 消费者评论极性研究

消费者评论极性(valence)指的是消费者评论中消费者对于产品的态度或效价。Mayzlin 通过研究发现购物网站上, 消费者对书的平均打分越高, 产品的销量越高[2]。但是, 现有的对 valence 的研究是以消费者打分(rating)为基础, 研究消费者评论中的数量特征(打分)。他们认为产品打分越高, 消费者对产品的满意度越高, 而没有从消费者评论中的文本内容出发探究消费者真实的态度和效价[2] [7]。实际上, 消费者评论的态度都是准对产品某个属性的, 从产品属性角度出发探究消费者对产品的满意可以更深刻的认识消费者评论效价的作用。消费者评论中对产品的满意度越高, 越容易提高其他消费者对该产品质量的感知, 使得消费者具有足够的购买动机, 从而影响产品的销量。从而, 我们提出:

H2: 产品评论越正面, 产品的销量越高。

### 2.3. 产品评论极性的调节作用研究

同时, 以往的口碑研究大部分都单独检验了口碑量和口碑极性的影响, 但是 Volume 和 Valence 可能存在交互作用。网络购物时, 消费者通常会首先观察其他消费者的打分, 在产品打分的基础上, 由于很多情况下消费者并不知晓产品卖出了多少, 产品的评论量越多使消费者知晓到很多消费者购买到此产品, 这时就会发生一种观察学习效应, 即消费者通过他人的行为来做出购买决策, 评论越多就会越相信整体的打分, 这时就会增加消费者购买产品的信心。因此, 我们提出假设:

H3: 用户产品评论量能正向调节评论极性对产品销售的影响。

### 2.4. 产品评论中差异性研究

消费者评论的差异性指的是消费者对产品的评价存在不一致性。大量的研究都从消费者对产品打分的 inconsistency 出发进行研究[8] [9]。例如在淘宝同一家网店上的同一样商品, 有的消费者打“3分”, 而有的消费者打“5分”。Sun (2010)发现消费者对产品的打分在概率上的分布越分散, 对产品越有利[8]。而有的研究者持不一样的结论。本文认为, 消费者评论打分的差异性并不能完全反应消费者对产品评价的不一致性, 如图 1。

我们现在可以考虑一种简单的状况。现在有两类消费者, A 类消费者对摩托车的速度要求高, B 类消费者对摩托车的安全性要求高, 他们在选购同一辆摩托车。现在已经有 X、Y 和 Z 三个消费者对该摩

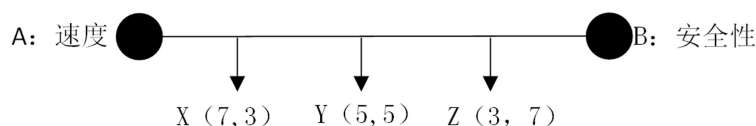


Figure 1. Consumer reviews variance

图 1. 消费者评论差异性

托车进行了评价。为了简化分析，我们可以以产品属性为坐标轴，坐标轴的两端分别是 A 和 B 两种不同需求的消费者。消费者 X 给该摩托车在速度上打 7 分，在安全性上打 3 分；消费者 Y 给该摩托车在速度上打 5 分，在安全性上打 5 分；消费者 Z 给该摩托车在速度上打 3 分，在安全性上打 7 分。可见 X、Y 和 Z 三个消费者对该产品的平均打分都是 5 分，但是三个消费者对产品是持不同意见的。与 Y 的评论相比，消费者 X 和 Z 的评论更能引起 A 和 B 的购买兴趣。

由 MOA 理论可知，由于产品评论的不一致使消费者很难形成对产品的统一态度，使消费者的产品涉入度大大增加，也就造成一种唤起状态(arousal)，促使消费者更加认真地阅读评论，获得对产品更准确的认识，而促成消费者购买产品。另外，研究者提出，双面评论会提升消费者获取产品相关信息的动机和好奇心并驱使更多的产品需求，因为这个产品的存在会使得它变成消费者心目中的首选。这种从消费者评论中所获取的可获得性会更持久、更抗变和更有效。因此，本文提出如下假设：

H4：产品评论的差异性正向影响产品销量。

### 3. 研究数据

#### 3.1. 原始数据

以往的研究多以书和电影数据为研究对象，为了扩大研究的丰富性并且使我们的研究具有一般性和实践性，我们选取了手机产品作为研究对象，并从淘宝中抓取产品数据。从 4 月 1 日到 6 月 13 日，本文跟踪抓取了淘宝上某家专营苹果手机的网店上 iPhone 5S 的数据。

在持续对该产品进行两个多月的数据抓取中，我们主要关注了该产品每日的销量和该产品每日新增的消费者评论。最后，总共抓取了该产品在 73 天内的 1696 条产品评论文本。

#### 3.2. 数据处理

在数据获取和处理的过程中，我们设计了一整套文本分析研究系统。该研究系统整合了大数据处理中数据的抓取、分析和呈现四个功能模块。本文主要就我们在研究过程中所涉及到的方法和技术进行简单说明。

##### (1) 数据抓取模块

我们利用爬虫程序每日凌晨定期执行，抓取上一天的更新数据，然后将抓取的评论数据存放到相应的数据库中供分析使用。本文最终的样本数据为该网店从 4 月 1 号到 6 月 13 号的时间段内的所有数据，有效数据包括每日的销量、每日的评论文本、评论发表的时间、评论者 ID 和产品 ID。数据在抓取之后还要经历一步数据清洗的过程，在网络环境中，用户评论文本在语法上缺乏规范，因此存在大量的有噪音的信息。

##### (2) 数据分析模块

大量学者采用关键词识别技术对消费者评论中所涉及到营销概念进行了测量和研究。但现有的研究未能将消费者评论中所涉及到的属性和消费者对该属性的评价进行结合。本文所采用的可上下文感知的中文文本分析技术，可以将消费者评论所涉及到的属性以及对该属性的评价进行自动分析并量化。

我们设计了一套可上下文感知的文本分析系统，它可以通过中文的语法结构树来识别消费者评论中涉及到的产品属性，感知在该语句中修饰该产品属性的评价词，以此判断消费者对该产品属性的评价并进行自动量化。例如一条消费者评论“苹果的外观超级棒”，中文分词程序将其解析为“苹果/名词，的/结构助词，外观/名词，超级形容词”。根据系统中的产品特征词典库，提取出评论中涉及的产品特征为“外观”，通过语法结构提取出修饰“外观”的形容词为“超级棒”，这是一个非常强烈的正面评价，系统赋值 10 (打分范围为：-10~10)。

有研究者根据消费者评论中文本内容，采用研究人员手工编码的方式将消费者评论分为三种，即产品属性评论、情感评论和推荐评论[10]。本文在 Gopinath 研究的基础上将消费者评论的分类细化，根据消费者满意度的度量，将消费者评论内容中的产品属性分为六大类，即质量、价格、象征、服务、物流和推荐。本文构建了一个满意度词典，将其置入文本分析系统中，系统可以自动的将消费者评论中提及的产品特征归类为相应的产品属性。例如上例中的一条评论“苹果的外观超级棒”，文本分析系统将“外观”自动归类到“质量”中，并将 10 分赋予“质量”这个产品属性。从表 1 中可以看到六个属性在 73 期上的描述性统计。

## 4. 研究模型和讨论

### 4.1. 解释变量的获取和描述性统计

73 期的 1696 产品评论在经过文本分析系统之后，呈现出来的是每一天的评论中评论者所评论的产品属性和其评价。某日的评论文本处理后的结果如表 2 所示。

我们只选择了产品评论中抽取出来的频率较高的产品属性， $B_{ij}$  指代的是第  $i$  条评论中，关于产品的  $j$  属性的打分， $B$  在  $(-10,10)$  范围内。 $D_{ij}$  指代的是第  $i$  条评论中，产品属性  $j$  的提及次数，因为有的属性会被不同的消费者提及多次。为了使 Variance 便于计算，我们将产品属性的打分映射到了  $(0,10)$  的数值范围。

本文所涉及到的解释变量为每日的消费者评论量(volume)、消费者评论的极性(valence)和消费者评论的差异性(variance)。消费者评论量(volume)为已知量，消费者评论极性的计算如公式(1)所示，消费者评论差异性的计算如公式(2)所示。

$$\text{Valence}_n = \sum_{i=n} B_{ij} D_{ij} / \text{features}_n \quad (1)$$

$$\text{Variance}_n = \sum_{i=n} (B_{ij} - \text{Valence}_n)^2 D_{ij} / \text{features}_n \quad (2)$$

式中， $n$  指的是第  $n$  天，指当天所有评论打分的均值， $\text{features}_n = \sum_{i=n} D_{ij}$

解释变量的描述性统计分析如表 3 所示。

### 4.2. 模型建立

本文所采用的数据均是评论当日的更新数据，因此代表每日的动态效果，便于研究解释变量的动态变化对被解释变量的影响。本文采用简单最小二乘法进行估计，为了体现模型中的时间先后顺序，解释变量均采用了滞后一阶。所以本文所得模型如下：

$$\text{sale} = c + a_1 \text{volume}(-1) + a_2 \text{valence}(-1) + a_3 \text{variance}(-1) + a_4 \text{volume}(-1) * \text{valence}(-1)$$

回归模拟在 Eviews 中进行，结果如表 4 所示。

### 4.3. 结果分析

结果显示，消费者评论量对产品的销量有显著的正向影响(H1)。消费者评论的极性对产品销量有显



**Table 1. Product attributes descriptive statistic**  
**表 1. 产品属性描述性统计**

属性	提及数	平均值	方差	最小值	最大值
质量	647	4.22	1.94	-2	7
价格	103	2.12	3.17	-6	8
象征	330	4.94	2.28	-2	10
服务	506	4.03	2.47	-2.48	7.33
物流	471	3.79	2.59	-4.5	8
推荐	98	2.21	3.89	-6	10

**Table 2. Processing data presentation**  
**表 2. 处理数据展示**

评论	质量	价格	象征	服务	物流	推荐
1	(B <sub>11</sub> , D <sub>11</sub> )	(B <sub>12</sub> , D <sub>12</sub> )	(B <sub>13</sub> , D <sub>13</sub> )	(B <sub>14</sub> , D <sub>14</sub> )	(B <sub>15</sub> , D <sub>15</sub> )	(B <sub>16</sub> , D <sub>16</sub> )
2	(B <sub>21</sub> , D <sub>21</sub> )	(B <sub>22</sub> , D <sub>22</sub> )	(B <sub>23</sub> , D <sub>23</sub> )	(B <sub>24</sub> , D <sub>24</sub> )	(B <sub>25</sub> , D <sub>25</sub> )	(B <sub>26</sub> , D <sub>26</sub> )
3	(B <sub>31</sub> , D <sub>31</sub> )	(B <sub>32</sub> , D <sub>32</sub> )	(B <sub>33</sub> , D <sub>33</sub> )	(B <sub>34</sub> , D <sub>34</sub> )	(B <sub>35</sub> , D <sub>35</sub> )	(B <sub>36</sub> , D <sub>36</sub> )
...						

**Table 3. Descriptive statistic and analysis of explanatory variable**  
**表 3. 解释变量的描述性统计分析**

	均值	中位值	最大值	最小值	标准差	偏度	峰度
消费者评论量	23.233	23.000	43.000	6.000	7.382	0.198	3.005
消费者评论极性	4.298	4.390	7.630	0.000	1.348	-0.522	4.293
消费者评论差异性	1.690	1.640	3.990	0.120	0.841	0.563	3.169
产品销量	468.822	450.000	713.000	335.000	87.473	0.962	3.527

**Table 4. Model estimation result**  
**表 4. 模型估计结果**

	模型 1		模型 2	
	系数(显著性)	标准误	系数(显著性)	标准误
<b>主效应</b>				
消费者评论差异性	45.435***	14.454	70.896***	16.402
消费者评论量	7.953***	1.480	17.033***	2.923
消费者评论极性	45.560***	6.975	66.140***	12.373
<b>调节效应</b>				
消费者评论量*			3.097***	1.331
消费者评论极性				

注意: \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.10。

著的正向影响(H2)。同时, 消费者评论量正向调节产品评论量对产品销量的影响(H3)。而消费者评论中

存在的差异性对产品的销量有显著的正向影响(H4)。评论量对销量的正向影响与以往的研究结果是一致的(Chevalier & Mayzlin, 2006), 消费者评论极性对销量也是有显著的正向影响, 与(Chintagunta et al., 2010) 结论相一致。Sun (2010)发现消费者打分的差异性在高于一定阈值时会对产品的销量产生正向的影响。与 Sun的结果相似, 但本文采用的研究方法更能深入的挖掘产品评论中对产品不同属性评价的真实差异性。

## 5. 结论和意义

### 5.1. 结论

以往的文献中, 已经对消费者评论的很多方面都有了一个论述, 但是缺乏对消费者评论一个整体的研究。通过中文文本分析系统, 我们挖掘出了消费者评论中对不同的属性的评价, 并结合以往的研究, 全面探讨了消费者评论对产品销量的影响机制。结果发现消费者评论量对产品销量有一个显著的正向作用, 而且消费者对产品属性的评价越正面, 对产品越有利。根据 MOA 理论, 消费者评论将阅读转为购买, 需要消费者有足够的处理信息的能力和动机。评论量越大提供的信息量越大, 使得消费者对产品的了解越深入, 从而有足够的信息处理能力。消费者评论的极性越高, 即消费者的评论越正面, 表明评论者对产品越满意, 使得消费者更有兴趣了解产品并最终购买。同时, 我们发现, 消费者评论量正向调节产品评论量对产品销量的影响。对企业来说, 应该更多的鼓励消费者对产品进行评价, 消费者评论量越大, 对产品的销售越有利, 而且, 产品的评论量可以扩大正面评价的作用。

同时, 我们发现消费者评论中的不一致性对销量有一个显著的正向影响。一方面, 消费者面对涉及到针对不同属性的不同的评价, 会增加评论的新奇性和客观性, 进而增加消费者了解产品的动机; 另一方面, 信息的丰富性增加消费者对产品的了解程度, 并以此发展出对产品的更系统的认知, 进而增加了消费者了解产品的能力。有了更了解产品的动机和能力, 在消费者进行购买决策时会将该产品列为自己心目中的首选地位, 进而提高该产品的销售业绩。

### 5.2. 意义

本研究通过设计了一套文本分析系统讨论了用户生成内容中产品评论的作用。以往的研究虽然对消费者评论的作用有较多的研究, 但是仍然未能从消费者评论中的属性评价出发对消费者评论的影响做一个系统的研究。我们的研究有以下几点意义。

首先, 本文结合计算机领域的前沿技术, 提出了一套处理中文消费者评论的系统并改进了现有的中文处理技术。通过数据抓取、清洗、评论断句、分词、特征提取和特征自动打分几个环节, 本文的研究系统可以深入地挖掘消费者评论中所涉及到的产品属性, 并较准确的挖掘出消费者对该产品属性的真实态度。本研究系统可以用来对产品进行舆情监控, 实时的观测消费者对于产品的真是评价, 使得营销工作更具有针对性和实效性。

其次, 本文通过测量消费者在不同产品属性上不同的评价, 扩展了现有研究对产品评论的研究。现有研究对评论的整体数值特征研究了很多, 却忽略了消费者可能对产品的不同属性有不同的评价。本文通过文本分析系统测量了评论中对不同属性的打分, 并通过建立了合理的回归模型, 发现消费者评论中产品的不同属性的评价差异性越大, 对产品的销量越有利。

## 参考文献 (References)

- [1] Ludwig, S., Ruyter, K.D., Friedman, M., et al. (2013) More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, **77**, 87-103.
- [2] Chevalier, J.A. and Mayzlin, D. (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, **43**, 345-354.

- [3] Chintagunta, P., Gopinath, S. and Venkataraman, S. (2010) The effects of online user reviews on movie box-office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, **29**, 944-957.
- [4] Zhu, F. and Zhang, X.Q. (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, **74**, 133-148.
- [5] Liu, Y. (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, **70**, 74-89.
- [6] Tang, T., Fang, E. and Wang, F. (2014) Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, **78**, 41-58.
- [7] Sengupta, J. and Johar, G.V. (2002) Effects of inconsistent attribute information on the predictive value of product attitudes: Toward a resolution of opposing perspectives. *Journal of Consumer Research*, **29**, 39-56.
- [8] Sun, M. (2012) How does the variance of product ratings matter? *Management Science*, **58**, 696-707.
- [9] Clemons, E.K., Gao, G. and Hitt, L.M. (2006) When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, **23**, 149-171.
- [10] Gopinath, S., Thomas, J.S. and Krishnamurthi, L. (2014) Investigating the relationship between the content of online word of mouth, advertising, and brand performance. *Marketing Science*, **33**, 241-258.