

在线商品评论的效用分析研究^①

杨 铭, 祁 巍, 闫相斌, 李一军
(哈尔滨工业大学管理学院, 哈尔滨 150001)

摘要: 大多数电子商务网站为消费者提供相互交流的平台来发表其针对某件商品的评论。但是, 随着在线商品评论的数量不断增加, 潜在消费者越来越难从中发现有助于制定购买决策的信息。因此, 如何从众多的评论中识别有用的评论, 分析在线评论的效用成为关注的热点。本文对在线商品评论效用分析的最新研究进行评述, 认为该领域的研究需要充分关注消费者的购买决策过程, 进而设计新的数据挖掘方法更好地辅助消费者的购买决策, 同时为电子商务网站的运营商调整营销沟通策略提供决策支持。

关键词: 电子商务; 在线商品评论; 信息过载; 效用评价

中图分类号: F272. 5 文献标识码: A 文章编号: 1007 - 9807(2012)05 - 0065 - 11

0 引 言

许多消费者在电子商务网站进行交易之后, 会对所购商品进行评分, 发表评论, 阐明使用商品和接受服务后的体验。作为 B2C 电子商务网站的显著特征, 在线商品评论不仅增加了消费者对 B2C 电子商务网站的认知有用性, 而且提高了 B2C 电子商务网站的社会认可性^[1]。作为一种反馈机制, 在线商品评论不但促进了 B2C 电子商务网站与消费者的双向互动, 而且促成了大规模的口碑网络^[2]。作为电子口碑 (electronic word-of-mouth), 在线商品评论显著地影响着消费者的购买决策和 B2C 电子商务网站的产品销售^[3]。一方面, 在线消费者常常通过衡量他人对某件商品的评论来制定自己的购买决策; 另一方面, 商品制造商可以从在线商品评论中获得启发, 从而支持广泛的管理活动, 如品牌塑造、客户关系管理、产品研发以及质量管理^[2]。不论买家还是卖家, 作为在线商品评论的读者都希望从中获得有助于制定决策的信息。然而, 针对一

件商品的在线商品评论往往数量众多而且内容质量参差不齐, 这使得消费者很难把目光聚焦在最有价值的评论上, 进而迅速制定购买决策。许多 B2C 电子商务网站, 如亚马逊、当当、京东, 提供在线商品评论的效用评价功能 (如图 1 所示) ——依据每条评论获得的“有用”投票数占总投票数的比例对商品评论进行排序, 获得支持票数越多的评论, 其排名越靠前。但是这种全然依靠人工判断的效用评价机制很难真正地将高品质的在线商品评论准确而及时地呈现在消费者面前。例如, 新近发表的在线商品评论需要很长时间来获得消费者的累计“有用性 (helpfulness)”投票, 这导致很多高品质的评论由于没有及时获得足够的赞成票而被淹没在数量庞大的信息海洋中。另外, 在这种人工效用评价机制下, 绝大多数在线商品评论被认为是“有帮助的”, 消费者依然面临信息过载的挑战。在线商品评论的效用分析是指应用数据挖掘的方法自动识别电子商务网站中那些对消费者制定购买决策有帮助的商品评论。

① 收稿日期: 2010 - 11 - 05; 修订日期: 2011 - 07 - 04.

基金项目: 国家自然科学基金资助项目 (70601009; 70890082).

作者简介: 杨 铭 (1982—), 男, 内蒙古丰镇人, 博士生. Email: yangming_hit@gmail.com

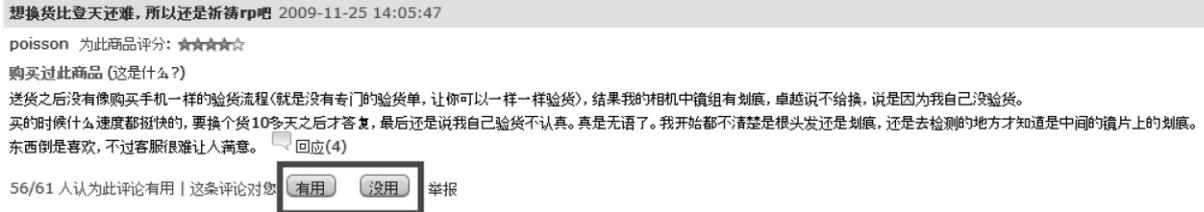


图1 卓越亚马逊网站上的商品评论效用评价

Fig. 1 A review utility evaluation on Joyo Amazon. cn

由于在线商品评论通常具有较强的主观性，消费者不得不揣测每条评论中的观点是否合理，还要综合权衡多方观点——有时甚至是相互冲突的观点^[4]。具体而言，潜在消费者需要对所看过的评论做出综合判断。为了缓解在线商品评论信息过载给消费者造成的负担，很多B2C电子商务网站提供商品的综合得分，即计算所有评论者给某件商品的评分的平均值。然而研究表明，消费者提供的商品评分呈双峰分布，这个得分只是折衷反映了两个评价极端（即好评和差评）^[5]。但是，饱含强烈极性（如赞美、批评）与作者个人观点的在线商品评论也许并不可靠，或者对其他消费者的购买决策并没有帮助^[4]。例如，某件商品的平均得分是4星，只能表明购买该件商品的所有评论者对这件商品的平均评分是积极的，而不能反映消费者对这件商品的意见。因此，B2C电子商务网站所呈现的某件商品的平均得分可能是误导性的推荐。

为解决当前B2C电子商务网站对在线商品评论的效用评价的局限性，学术界提出了一系列基于文本挖掘和信息质量评价理论的解决方案。Liu等^[6]认为对潜在消费者购买决策有用的在线商品评论取决于3个因素：评论者的经验（expertise），评论的写作风格（writing style）和评论的时效（timeliness）。该研究通过从IMDB影评数据集抽取相关属性建模拟合以上3个因素，进而构建非线性的回归模型，预测评论的有用性。Ghose和Ipeirotis^[7]则认为在线商品评论内容中的观点主观性（subjectivity）有助于预测评论的效用。Otterbacher^[8]则认为评论的内容质量（message quality）是评价在线商品评论效用的重要方面。可以看出，以往研究对于影响在线商品评论效用的因素没有统一的看法。在这种

情况下，使用不同的数据挖掘技术，不同的数据集，以及不同的算法评价指标，会得出千差万别的结论。这些结论可能会使电子商务网站IT人员感到无所适从。本文从在线商品评论效用评价涉及的4个维度——评价目标、特征选取、评价方法和评价对象——梳理了近几年该领域的研究，分析了已有研究中存在的问题，并且针对这些问题提供了未来的研究方向和阶段性成果。

1 效用评价研究的分类

在线商品评论的效用评价是基于CMC（computer mediated communication）文本模式的分析。这类研究需要准确地评价、总结、表示CMC文本，尤其是CMC主体内容的文本（如在线商品评论的主体文本内容）^[9]。尽管前人对在线商品评论的文本分析进行了大量研究，但以往的研究主要集中在观点挖掘和情感分析两方面。这主要有两方面的原因：首先，在线商品评论的内容富含观点与情感相关的内容^[10]；其次，主观性是商品评论的固有特性，其中蕴含的商业价值使得商品评论成为情感分析与观点提取的重要研究对象^[4]。与观点挖掘和情感分析类似，在线商品评论的效用评价研究要素包括评价目标、评价特征、评价技术和评价对象的选取^[11]。按照这些要素和相应的维度划分，表1分别对以上4个要素进行了解释。在此基础上，表2对在线商品评论效用评价的国内外研究现状加以概述。以下将从在线商品评论效用评价涉及的4个维度：评价目标、特征选取、评价方法和评价对象，梳理近几年该领域的研究，进而分析当前研究中存在的问题。

表 1 在线商品评论效用评价的研究分类

Table 1 A taxonomy of online review utility evaluation

研究要素	维度划分	描述/举例	标识
评价目标	排序	依据评论的效用得分以降序呈现评论内容	G1
	汇总	归纳汇总某件商品的所有在线商品评论	G2
	分类	区分高品质的评论	G3
评价特征	语法特征	Word/POS tag <i>n</i> -grams, 措辞模式, 停顿	F1
	语义特征	极性标识, 评价组, 文脉特征	F2
	体裁特征	文体风格的词汇和结构测度	F3
	元数据	评论的得票数, 发表时间, 商品评分	F4
评价技术	机器学习	支持向量机, 支持向量回归	T1
	相似度得分	短语模式匹配, 词频数	T2
评价对象	实用型商品	数码相机, 电脑, MP3 播放器等	D1
	享用型商品	书籍, 电影, 音乐等	D2

表 2 在线商品评论效用评价部分研究汇总

Table 2 Selected previous studies in utility evaluation of online product reviews

研究论文	评价特征				评价技术		评价对象		评价目标		
	F1	F2	F3	F4	T1	T2	D1	D2	G1	G2	G3
Kim 等 ^[13]	√	√	√	√	√		√		√		
Liu 等 ^[6]	√	√	√	√		√	√	√	√		
Liu 等 ^[17]	√	√	√		√		√			√	√
Jindal 和 Liu ^[18]	√	√	√	√	√		√	√	√		√
Zhang ^[4]	√	√	√	√	√	√	√	√	√	√	
Weimer 和 Gurevych ^[12]	√		√	√	√	√		√			√
Miao 等 ^[14]	√	√		√		√	√		√		
Mudambi 等 ^[38]		√		√	√		√	√	√		

1.1 在线商品评论效用评价的目标

由于在线商品评论不仅数量众多, 而且信息质量参差不齐. 解决这一信息过载 (information overload) 问题, 并把有助于潜在消费者制定购买决策的有用评论及时呈现出来, 涉及对在线商品评论的效用进行评价. 当前研究中, 对在线商品评论的效用评价目标主要体现在以下 3 个方面: 对评论进行自动排序、归纳汇总和内容分类.

对评论进行自动排序就是按照评论的有用性得分, 将高品质的评论及时呈现在潜在消费者面前. 例如, 一些研究人员提出了两类基于效用得分的商品评论排序机制: 面向顾客的排序机制——依照顾客的预期效用对评论进行排序^[7, 13]; 面向厂商的排序机制——依据厂商对某商品的预期销量对评论进行排序^[7]. 还有的研究人员提

出了观点搜索系统, 该系统在对评论进行排序时考虑了信息质量的时间维度和主题相关性, 从而可以满足潜在顾客的个性化需求^[14].

即使网站能自动对在线商品评论的效用进行排序, 消费者仍需要仔细揣测每条评论中包含的观点, 有时甚至要权衡多方提出的对立观点. 很多消费者可能会因此而无所适从, 因此, 消费者还需要了解所有评论者对某件商品的综合评价以及评论的归纳汇总. 这有助于消费者迅速做出购买决策, 缓解信息过载给消费者造成的负担. 当前 B2C 电子商务网站是基于消费者评分的平均值, 归纳汇总某件商品评论信息, 但这种综合评价的合理性受到以下两个事实的挑战. 首先, 实证研究表明商品评分呈现双峰分布, 即消费者要么给予商品很高的评价, 要么给予差评^[5]. 也就是说, 消费者给某件商品的评分主要集中在

“4 星”以上和“2 星”以下。第二，蕴含较强极性（赞美的或者批评的）和个人观点的商品评论可能并不可靠，或者对消费者的购买决策没有帮助^[4]。例如，某位消费者在对某品牌豆浆机的评论中写道：“因为价格便宜，因为女儿爱喝豆浆，所以买了这款豆浆机。现在每天都在用，真的很好用，物有所值。”这篇评论的极性显然是积极的，作者持赞成的观点。但是，这篇评论没有详细说明这款豆浆机究竟如何好用。所以这篇评论可能对其他消费者的购买决策基本没有帮助。因此，当前 B2C 电子商务网站提供的商品评价汇总并不能真正反映所有消费者的观点，而只是折衷反映了两个评价极端（即好评和差评）^[5]。对某件商品的合理评价汇总需要考量每条评论的效用和极性。Zhang^[4]利用以下加权方法计算所有评论者给予某件商品的综合评分

$$V(P) = \frac{\sum_{i=1}^n \{u[T_i(P)]Polarity[T_i(P)]\}}{\sum_{i=1}^n u[T_i(P)]} \tag{1}$$

其中， $V(P)$ 表示所有评论者对某件商品 P 的综合评分， $T_i(P)$ 表示针对商品 P 的 1 篇消费者评论， $u[T_i(P)]$ 表示评论 $T_i(P)$ 的效用， $Polarity[T_i(P)]$ 表示评论 $T_i(P)$ 的极性。 $V(P)$ 是商品 P 的所有评论极性的加权平均值。另一种解决评论信息过载的手段是归纳所有评论。Hu 和 Liu^[15]提出了对客户评论进行观点挖掘和归纳的方法，该方法由以下 3 步构成：1) 识别客户评论中出现的产品特征；2) 就每一产品特征，识别评论中含有积极观点的语句或者消极观点的语句；3) 对上述产生的信息进行归纳。在此研究基础上，一些学者通过分析评论内容的内部主题结构对评论进行概括，从而归纳消费者所关注的事物^[16]。但是，对在线商品评论进行概括的过程中很少有人强调评论的内容质量控制。

此外，对评论内容按照特定标准进行分类也是评论效用评价的目标之一。比如，由于在线商品评论在文本编辑方面缺乏质量控制，垃圾评论（review spam）比比皆是，这就有必要区分所发表的评论是否是垃圾评论。Liu 等^[17]给出了测度在线商品评论质量的规范。依据这一规范，人工标记

从亚马逊网站收集的在线商品评论，再用标记后的在线商品评论训练和验证基于统计学习理论的分类算法——支持向量机，最后利用该分类器自动侦测垃圾评论。Jandal 和 Liu^[18]则认为在线商品评论中存在故意误导其他消费者的垃圾信息，他们把垃圾评论分成 3 类：① 评论中包含错误的观点；② 仅对某一品牌进行评论；③ 评论没有包含任何观点。依据这一分类，他们制定了以下侦测垃圾评论的策略：1) 侦测抄袭或者近似抄袭的评论；2) 用人工标记的样本训练有监督的机器学习算法，再用该算法侦测第 ② 类和第 ③ 类垃圾评论；3) 利用 3 类垃圾评论的副本以及其他相关信息侦测第 ① 类垃圾评论^[18]。但是侦测垃圾评论面临着相当大的困难^[18]，因为垃圾评论本身很难明确界定。例如，为了评价一篇在线商品评论的质量，研究者通常需要给出具体的评价准则或者规格（一般是描述商品的若干方面，如性能，价格等），然后观察某篇评论满足这些规格的程度，进而评估该评论的效用或质量。但是，前人提出的评价准则通常主要依据观察和经验，缺乏理论依据，这些准则有可能存在偏差。另外，目前还没有证据表明，在线消费者依据（或者应该依据）这些准则来判定某篇评论是否对自己有帮助。

1.2 评价在线商品评论效用的特征选取

使用数据挖掘的方法对在线商品评论的效用进行预测，需要构建一系列数据属性来表征其效用。也就是需要明确，一篇在线商品评论具备了哪些数据特征即表明它是有用的。以往的研究主要使用了两类数据特征，分别是：文本特征（text feature）与元数据特征（meta-data feature）。

文本特征包括语法特征（syntactic feature）、语义特征（semantic feature）和体裁特征（stylistic feature）。语法特征一直被广泛地用于表征在线商品评论。这类特征包括：word n -grams，词性标记，标点符号以及措辞模式。Zhang^[4]研究发现简单的语法特征，包括专有名词、数字、情态动词、感叹词、以 wh 开头的单词、形容词比较级和最高级，在预测在线商品评论的效用时贡献最显著。

语义特征需要结合人工 / 半自动或者全自动的标注技术，添加词语 / 短语的极性或者情感程度的刻度^[11]。这类特征包括极性标记（积极的 / 消极的，主观的 / 客观的），评价组（态度、倾向性、

等级以及词组的极性) 和用来表示上下文语义倾向性的文脉特征.

体裁特征的使用也较为普遍, 在表征在线商品评论效用评价方面的使用频率仅次于语义特征. 这类特征包括词汇和结构属性, 这些属性又与作者身份识别相关联, 如文献 [20 - 22]. 值得一提的是体裁特征适合用于文本的主观性判别和观点判别^[23].

元数据特征和在线商品评论的语言特征无关^[13], 它能够表征商品评论作者与读者之间的互动属性. 这类特征包括评论员给自己所购买商品的评分, 一篇评论所获得的累计总投票数及赞成票数, 产品的发布时间, 评论的发表时间, 和产品的描述等. 绝大多数研究者使用了累计赞成票数对投票数的比例 (x/y) 作为元数据特征. 这一特征直接而简便地近似表明了一篇评论的效用^[4], 因而被广泛地作为效用评价模型的预测变量.

但是, 目前学术界对于什么特征能够有效地刻画一篇评论的效用还没有形成一致的观点, 有时这些观点甚至是相互冲突的. 例如, Kim 等^[14]研究表明, 刻画在线商品评论效用的最佳特征是评论篇幅, 把评论文本中的词汇按曲折变化形式进行归类得到的 tf-idf 得分, 以及评论者给商品的评分. 也就是说, Kim 等^[13] 认为元数据特征和简单的文体特征足以刻画一篇评论的效用. 而 Weimerh 和 Gurevych^[12] 却认为文体特征不如元数据特征的作用显著; Zhang^[4] 则认为表面语法特征 (shallow syntactic feature) 在预测一篇评论的效用中说服力最强. 郝媛媛等^[24] 认为评论的正负情感倾向对评论的有用性影响显著. 或许这些没有达成一致的观点是由于评论样本选取^[19], 评价技术的选择以及评论效用界定的差异所引起的.

1.3 在线商品评论效用评价技术

在线商品评论效用评价的技术可以分为基于机器学习和基于相似度得分的两类方法. 基于机器学习的评价方法主要有支持向量机 (support vector machines, SVM) 和支持向量回归 (support vector regression, SVR). 支持向量机在文本分类算法中占有主导地位, 而支持向量回归则被广泛用于处理连续性刻度, 如情感强度^[25], 评论的效用^[4, 6, 13]. 这是因为支持向量机和支持向量回归算

法可以使用功能强大的核函数 (kernel functions), 而核函数可以捕获数据中的大量结构特征.

Liu 等^[17] 利用支持向量机侦测低质量的在线商品评论. 设 x 表示一篇在线商品评论的特征向量, 该篇评论的效用值可由方程 (2) 计算,

$$f(x) = w^T x + b \quad (2)$$

其中 w 表示特征权重向量; b 表示截距. $f(x)$ 函数值越大, 商品评论 x 的效用就越大. 如果函数值为正, 这篇评论就被判定为高质量; 否则, 就被判定为低质量. Chen 和 Tseng^[26] 依据信息质量理论 (information quality theory) 提出了评价在线商品评论质量的 9 个维度, 分别是可信度 (believability), 客观性 (objectivity), 声誉 (reputation), 相关性 (relevancy), 及时性 (timeliness), 完整性 (completeness), 信息翔实性 (appropriate amount of information), 容易理解程度 (ease of understanding), 简明扼要程度 (concise reputation). 然后从在线评论中提取文本特征和元数据特征分别表示以上 9 个维度. 最后构造多类支持向量机 (multi-class SVM) 模型对在线评论进行分类, 从而区分高质量的评论.

运用支持向量回归对某篇在线商品评论的效用进行评价可以用方程 (3) 表示

$$u(x) = F(f_1, f_2, \dots, f_i, \dots, f_p) \quad (3)$$

其中 $f_1(x), \dots, f_i(x), \dots, f_p(x)$ 表示在线商品评论的特征向量; $u(x)$ 表示评论的效用, 函数的值域是 $u \in [0, 1]$, 连续型的函数值可以尽可能准确的反应评论的真实效用. 另外, 使用支持向量回归作为评价技术时往往采用径向基函数 (radial basis function, RBF) 作为核函数. 因为径向基函数能够处理函数目标值与常用特征之间的潜在非线性关系^[6].

基于相似度得分的方法通常依照文本反映的总体情感倾向进行分类, 例如一篇评论主要批评某件商品, 这篇评论就体现了消极的倾向性. 有时评论中有褒有贬, 这就需要综合比较褒贬的程度来确定整体的情感倾向性. 有的学者通过评论中体现的主观性和客观性对评论的效用进行预测. 例如, Ghose 和 Ipeirotis^[7] 研究了评论的主观程度得分和评论效用之间的关系. 他们将一篇评论中的每个语句分为“主观的”或者“客观的”, 记录每

个语句 s 的主观程度概率 $Pr_{\text{subj}}(s)$. 这样, 每篇评论 x 都有一个主观程度得分. 语句主观程度得分的标准差为

$$DevProb(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[Pr_{\text{subj}}(s_i) - \frac{1}{n} \sum_{i=1}^n Pr_{\text{subj}}(s_i) \right]^2} \quad (4)$$

他们认为语句主观程度得分的标准差 $DevProb(x)$ 与一篇评论的效用有很强的统计显著性. Zhang^[4] 使用基于得分的方法——向量空间中的标准余弦相似度, 测量在线商品评论与商品描述之间的相似性, 认为一篇有用的在线商品评论应该是原创的而不是对商品描述的忠实转载. Miao 等^[14] 使用基于得分的方法计算评论中的观点与消费者查询词汇之间的相似度, 从而为消费者提供个性化的在线商品评论.

Lau 等^[27] 认为垃圾评论或者不真实的评论 (fake review) 的重要特点是过度泛滥, 即内容雷同的评论针对不同产品反复出现. 在这一假设下, 他们提出了新的语义语言模型 (semantic language model) 来估计不同评论的相似度, 高相似度的评论被划分为垃圾评论. 然后, 作者借助计量经济学的分析方法检验了垃圾评论对亚马逊网站上图书销售排名的影响. 结果发现, 仅仅 1 条负面的垃圾评论可以在 1 个月内给电子商务网站造成的经济损失达 76.4 美元.

1.4 在线商品评论的评价对象

在线商品评论的评价对象可以分为实用型 (utilitarian) 和享用型 (hedonic) 两类商品^[28]. 实用型商品包括电脑、手机、数码相机、MP3 播放器等等, 享用型商品包括音乐、电影、书籍等等. 商品的评论内容相对较为复杂, 针对同一件商品的不同方面既有消极的评价又有积极的评价^[11]. 此外, 对享用型商品的在线商品评论进行效用评价比对实用型商品的评论更为复杂. 它不仅包含冗长的情节概要, 而且评论者常常使用复杂的文学修饰手法, 既有华丽的修辞, 也不乏讽刺挖苦^[11, 19].

对享用型商品发表的在线评论通常主观性很强, 主要表现为很多评论爱憎情感分明. 享用型商品的评论一般描述的是个人的独特体验. 在对这

类评论的效用进行评价时, 潜在消费者的个人品味发挥着重要作用, 而消费者往往对自己的品味和主观判断很自信, 而对那些饱含爱憎情感的评论存在质疑. 因此, 对享用型商品而言, 消费者更倾向于相信那些相对客观的评论.

当消费者购买实用型商品时, 他们更倾向于看到描述商品的具体属性 (如电脑的配置, 数码相机的像素) 的在线评论. 因为相对于享用型商品, 实用型商品的属性是具体可见的, 所以潜在消费者一般认为对商品具体属性的评价是令人信服的.

不同的商品类型, 其对应的评论内容差异很大, 不同商品类型的评论, 影响消费者的感知效用的因素也不同^[29]. 因此, 对在线商品评论的效用进行分析时, 需要考虑商品的类型.

2 现有研究存在的问题

通过对国内外研究现状进行分析, 可以发现当前对在线商品评论的效用评价研究主要存在以下问题. 第 1, 很少从理论上解释在线商品评论的哪些属性会影响消费者对它们的感知有用性. 第 2, 所使用效用评价方法主要依赖机器学习算法, 很少考虑消费者对在线商品评论产生认知有用性的行为过程. 第 3, 很少考虑两个偏差: 即那些没有获得投票的在线商品评论以及由此造成的样本分布的偏度. 以下将分别解释当前研究中存在的主要问题.

2.1 影响在线商品评论效用的因素缺少理论依据

以往的研究很少从理论上解释在线商品评论的哪些属性会影响消费者对它们的感知有用性. 针对一件商品的在线商品评论往往具有多重属性, 不同消费者给予同一篇评论的评价也不尽相同. 另外, 与其他形式的 CMC 文本 (如电子邮件、网络论坛以及博客) 类似, 在线商品评论也富含社交线索, 如观点、情感、文体和文风. 这些事实使得表征在线商品的效用非常困难. 同时, 目前就如何合理刻画在线商品评论的效用还没有形成一致观点. 值得注意的是, 潜在消费者是在线商品评论的主要读者. 然而, 以往的研究却很少从消费者购

买决策行为的角度出发研究在线商品评论的效用。

2.2 在线商品评论效用评价的方法单一

以往研究中所使用的效用评价方法主要依赖机器学习算法和相似度度量方法,很少考虑消费者对在线商品评论产生认知有用性的行为过程。机器学习算法需要使用计算机可以识别的特征来表征在线商品评论。但是,在线商品多种多样,而不同消费者针对同一件商品发表的评论也是仁者见仁智者见智。尽管某些机器学习算法(如支持向量机)的核函数可以捕获数据中的大量结构特征,但是这种方法仅仅是启发性的,特征选取大多凭直觉和经验,另外相同的机器学习算法面向不同的数据集(如来自不同网站的评论,中文和英文评论)的时候性能差异很大。而相似度度量的方法虽然在评论效用评价中表现稳定,但其计算复杂度(computing complexity)要高于机器学习算法^[30]。因此,在线商品评论的效用自动评价迫切需要性能稳定、计算复杂度较低的数据挖掘方法。

2.3 人工效用评价的偏差

以往对在线商品评论进行自动效用评价的时候,通常使用网站的人工评价结果作为参照,即将在线商品评论获得的累计赞成票与得票总数之比作为模型或算法的因变量。但是,使用这个比值作为因变量是有偏差的,而过去的研究很少强调这个偏差。

这个偏差表现在两方面。首先,B2C电子商务网站中的很多在线商品评论没有获得人工效用评价,即没有获得任何投票。例如,有人曾对亚马逊网站上的在线商品评论进行过调查,样本是20000篇有关MP3播放器的在线商品评论。研究发现,近38%的评论少于3票,在发表时间超过3个月的评论中,31%的评论不足3票^[13]。此外,越是新近发布的评论,获得的投票越少,因而在网站上被呈现的位次越靠后^[17]。或许有些没有获得投票的评论确实没有帮助,消费者不愿意去给它们投票。但是,这并不意味着所有没有获得投票的在线评论肯定毫无帮助,尤其不能武断地认为新近发表的评论由于没有获得足够的投票就毫无帮助^[19]。然而,以往研究通常只选择那些获得10张以上累计投票的评论作为样本,这样选取的样本

很可能缺乏代表性。

另一方面,偏差还体现在人工效用评价结果的分布上。具体而言,经过人工评价,绝大多数的在线商品评论显示“有用”。通过从亚马逊网站上抓取大约4百万条在线书评(涉及670000本不同书籍,并且这些书评所获累计投票都在10票以上),发现这些评论的人工效用评价平均值在70%~74%之间。也就是说,平均来看,每条评论被认为“有用”的概率大约在70%~74%之间^[31]。无独有偶,通过从亚马逊网站上抓取大约23000条有关数码相机的在线商品评论,发现50%的评论的人工效用评价值(即“你是否认为该评论有用?有用:没用)超过9:1,而且随着评论发表日期更新,在线商品评论获得人工效用评价的数量呈指数下降趋势^[17]。

在样本以及因变量选取方面,以往研究很少强调以上两个方面的偏差。尽管这样可以简化算法设计以及计算过程,但这可能会使在线商品评论的自动效用评价结果出现偏差。通过以上对人工效用评价偏差的分析发现,当前B2C电子商务网站提供的效用评价机制很不完善。这一点从平时浏览在线商品评论的直觉中也可以做出判断,那就是排名靠前的评论未必是有帮助的评论,排名靠后的评论未必是垃圾评论,而且这种现象非常普遍。因此,在线商品评论的自动效用评价有必要考虑以上分析的两方面偏差,从而提高对评论效用预测的准确性。

3 研究设计与研究进展

针对在线商品评论效用分析已有研究中存在的问题,本文认为今后可以通过3个阶段性研究进行扩展。第1阶段,以信息经济学和消费者行为学理论为依据,研究在线商品评论对消费者购买决策行为的影响机理;第2阶段,按照之前形成的理论依据,运用更加先进的数据挖掘和统计学方法,改进现有的在线商品评论效用评价技术,克服人工效用评价所造成的偏差;第3阶段,通过评价在线商品评论的效用,为B2C电子商务网站的网络营销提供决策支持。

3.1 探究影响在线商品评论效用的理论依据

在线商品评论的效用主要体现在其内容是否

有助于消费者制定购买决策。因此,对在线商品评论的效用进行分析可以看作是信息诊断的过程。将在线商品评论的效用作为消费者决策过程中的测度,符合信息诊断在信息经济学研究中的观点^[32-35]。事实上,在线商品评论可以在消费者购买决策过程中的不同阶段发挥作用。消费者的购买决策由6个阶段组成,包括需求识别、信息搜索、备选方案评估、购买决策制定、购买和购买后的评价^[36]。一旦确定了消费需求,消费者能够利用相关的在线商品评论收集充分的信息来制定购买决策,同时评估各种备选方案。通过信息搜寻和备选方案对比,在线消费者可以制定满意度更高的决策^[37]。在购买行为结束之后,一些消费者会返回网站,发布对购买的商品使用后的体验。通过阅读其他消费者发布的商品评论,消费者可能会意识到自己需要这件商品,继而开始另一个新的购买决策过程。

在线消费者的购买决策过程表明,电子商务网站可以通过提供高品质的在线商品评论为消费者提供更多潜在的价值。如果电子商务网站能向消费者提供便捷的途径来获得这些高品质的在线商品评论,将极大地改善消费者的决策水平并加快消费者的决策进程,从而提高消费者对网站的满意度和忠诚度。在现实中,鼓励高品质的在线商品评论一直是一些电子商务网站开展内容营销(content marketing)的重要策略。为此,需要首先分析在线商品评论的哪些因素会影响消费者的购买决策,进而分析这些因素之间的相互关系是什么,它们是如何作用于消费者的购买过程中的。在这方面,Susan和David^[38]进行了探索性的研究,他们的对亚马逊网站上1587条在线商品评论进行分析后发现,在线商品评论的内容深度(review depth)显著影响消费者对在线商品评论的感知有用性,进而影响消费者的购买决策。

3.2 在线商品评论效用评价技术和方法创新

研究在线商品评论对消费者购买决策行为的影响机理,为合理选取表征在线商品评论效用的特征和数据挖掘方法提供了理论依据。而在线商品评论效用的评价技术和方法创新是逾越理论和实践鸿沟的必由之路。

管理信息系统研究的主要目标是扩展人的能

力与组织的能力的边界,而这一目标是通过创造新的IT制品(IT artifact)来实现的。新的IT制品可以是构件(constructs)、模型(models)、方法(methods)和实例(instantiations)^[38]。因而,信息系统设计本身不仅包含设计出的产物(product)也包含设计过程(process)^[39-40]。信息系统设计的产物包括一系列设计需求和开发IT制品构件的必要设计准则。而信息系统的设计过程则是构造和评估设计产物的一系列反复迭代的专门活动^[41]。因此,如何表示所要设计的问题,如何构建并评价设计方案,是信息系统设计研究的两个主要问题^[42]。

在线消费者通常必须在信息不完整的情形下制定决策。他们通常不了解所购买产品的质量、销售商的信誉以及其他备选方案。他们也知道在网上搜索这些信息需要花费很多时间和精力。在线消费者需要权衡感知成本和信息搜索带来的潜在效益。他们遵循以下决策过程:通过信息搜索减少决策的不确定性,同时承认这种不确定性在购买决策中不可排除^[43]。因此,在线商品评论的效用评价技术和方法需要在此前提下进行创新。以往的研究主要关注在线商品评论的静态属性,如评论的语法、语义和观点等文本特征,以及评分(rating)和投票数(votes)等元数据特征,而很少考虑其动态属性,例如,与时间相关的特征(如评论发表的时间)和消费者之间的互动特征(如其他消费者对某条评论的反馈意见)。另外,以往的研究在选取效用评价模型的因变量(即helpfulness)时一般忽略人工效用评价的偏差。而这一偏差可能是由忽略在线商品评论的动态属性造成的。对亚马逊网站上随机收集的1500篇在线评论进行分析,发现有用性较强的评论往往伴随着众多回应。因此,结合在线商品评论动态属性构建效用评价模型,进而解决人工效用评价带来的偏差,将是今后研究的重要方向。

3.3 基于在线商品评论效用的网络营销策略

当前研究在线商品评论效用自动评价的动因比较简单,其主要研究目标是对在线商品评论进行排序、归纳汇总以及分类,而很少强调如何将在线商品评论的呈现方式与电子商务网站的运营策略结合起来。通过创新的数据挖掘技术对在线

商品评论的效用进行自动分析的最终目标是为电子商务网站提供网络营销决策支持。

在实际应用中, B2C 电子商务网站在处理不同效用的在线商品评论时, 面临的问题要复杂得多。例如, 电子商务网站该如何依据不同效用的在线商品评论适时地调整自己的营销沟通策略? 在什么情况下, 网站通过为消费者提供便利的评论平台能促进消费者的参与, 从而使网站盈利^[44]? 此外, 有实证研究发现了所谓的不一致效应: “商品评论数量比质量更能影响低参与度的消费者……当商品评论的质量较高时, 商品评论的数量更能影响高参与度的消费者……在评论数量相当的前提下, 评论的质量显著影响高参与度的消费者; 而当评论的数量有限时, 评论数量对高参与度的消费者作用更显著”^[3]。

因此, 基于在线商品评论的效用分析的商务智能开发需要立足于 B2C 电子商务网站运营中遇到的实际问题。依据在线商品评论的数量和效用适时调整 B2C 电子商务网站的营销策略, 不仅能够加强网站与消费者的互动, 而且有助于吸引较低参与度的消费者, 维持高参与度的消费者, 从而提高网站的盈利能力和运作绩效。另外, 本文发现“有用性”得票数较多的在线评论中一般会列举商品的若干优缺点, 以及与其他

同类商品进行对比。商家可以利用这些信息对产品展开联合分析 (conjoint analysis)^[45], 同时发掘利基市场。

4 结束语

对在线商品评论进行效用评价是电子商务领域研究的重要问题, 得到了电子商务网站运营商的普遍关注和实际应用。本文以在线商品评论效用评价涉及的 4 个维度 (即评价目标、特征选取、评价方法和评价对象) 为研究框架, 梳理了近几年该领域的研究。针对已有研究中存在的问题, 分别从理论和实践的角度提供了可行的解决方案和阶段性成果。本文认为今后该领域的研究可以通过 3 个阶段进行扩展。首先, 依据信息经济学和消费者行为理论, 探究在线商品评论对消费者购买决策行为的影响机理, 分析在线商品评论的哪些因素会影响消费者的购买决策。在此基础上, 开发新的商务智能应用技术对在线商品评论的效用进行自动评价, 及时呈现高品质的在线评论从而辅助消费者制定购买决策。最后, 电子商务运营商可以充分利用萃取的高品质评论, 从中洞悉消费者对产品的偏好, 进而制定营销战略和发掘利基市场。

参 考 文 献:

- [1] Kumar N, Benbasat I. The influence of recommendations and consumer reviews on evaluations of websites [J]. *Information Systems Research*, 2006, 17(4): 425 - 439.
- [2] Dellarocas C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms [J]. *Management Science*, 2003, 49(10): 1407 - 1424.
- [3] Park D H, Lee J, Han I. The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement [J]. *International Journal of Electronic Commerce*, 2007, 11(4): 125 - 148.
- [4] Zhang Z. Weighing stars: Aggregating online product reviews for intelligent e-commerce applications [J]. *IEEE Intelligent Systems*, 2008, 23(5): 42 - 49.
- [5] Hu N, Pavlou P A, Zhang J. Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication [C]// *Proceedings of the 7th ACM Conference on Electronic Commerce*, Ann Arbor, Michigan, USA: Association for Computing Machinery, 2006: 324 - 330.
- [6] Liu Y, Huang X J, An A J, et al. Modeling and predicting the helpfulness of online reviews [C]// *Proceedings of the 8th IEEE International Conference on Data Mining*, Washington, DC, USA: IEEE Computer Society, 2008: 443 - 452.
- [7] Ghose A, Ipeiortis P G. Designing novel review ranking systems: Predicting the usefulness and impact of reviews [C]// *Proceedings of the 9th ACM Conference on Electronic Commerce*, Minneapolis, MN, USA: Association for Computing Machinery, 2007: 303 - 310.
- [8] Otterbacher J. "Helpfulness" in online communities: A measure of message quality [C]// *Proceedings of the 27th Interna-*

- tional Conference on Human Factors in Computing Systems , Boston , MA , USA: Association for Computing Machinery , 2009: 955 – 964.
- [9] Abbasi A , Chen H C. Cybergate: A design framework and system for text analysis of computer-mediated computation [J]. *MIS Quarterly* , 2008 , 32(4) : 811 – 837.
- [10] Ye Q , Zhang Z Q , Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches [J]. *Expert Systems with Applications* , 2009 , 36(3) : 6527 – 6535.
- [11] Abbasi A , Chen H C , Salem A. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums [J]. *ACM Transactions on Information Systems* , 2008 , 26(3) : 1 – 34.
- [12] Weimer M , Gurevych I. Predicting the perceived quality of web forum posts [C]// *Proceedings of RANLP 2007 Borovets , Bulgaria: Publication Post of Subjectivity* , 2007.
- [13] Kim S M , Pantel P , Chklovski T , et al. Automatically assessing review helpfulness [C]// *Proceedings of the Conference Empirical Methods in Natural Language Processing* , Morristown , NJ , USA: Association for Computational Linguistics , 2006: 423 – 430.
- [14] Miao Q L , Li Q D , Dai R W. AMAZING: A sentiment mining and retrieval system [J]. *Expert Systems with Applications* , 2009 , 36(3) : 7192 – 7198.
- [15] Hu M Q , Liu B. Mining and summarizing customer reviews [C]// *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , New York , NY , USA: Association for Computing Machinery , 2004: 168 – 177.
- [16] Zhan M J , Loh H T , Liu Y. Gather customer concerns from online product reviews : A text summarization approach [J]. *Expert Systems with Applications* , 2009 , 36(2) : 2107 – 2115.
- [17] Liu J J , Cao Y B , Lin C Y , et al. Low-quality product review detection in opinion summarization [C]// *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* , Prague: Association Computational Linguistics , 2007: 334 – 342.
- [18] Jindal N , Liu B. Analyzing and detecting review spam [C]// *7th IEEE International Conference on Data Mining* , Omaha , USA: IEEE Computer Society , 2007: 547 – 552.
- [19] Pang B , Lee L. Opinion mining and sentiment analysis [J]. *Foundations and Trends in Information Retrieval* , 2008 , 1(2) : 1 – 135.
- [20] Abbasi A , Chen H C. Applying authorship analysis to extremist-group web forum messages [J]. *IEEE Intelligent System* , 2005 , 20(5) : 67 – 75.
- [21] Abbasi A , Chen H C , Nunamaker J F. Stylometric identification in electronic markets: Scalability and robustness [J]. *Journal of Management Information Systems* , 2008 , 25(1) : 49 – 78.
- [22] Abbasi A , Chen H C. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace [J]. *ACM Transactions on Information Systems* , 2008 , 26(2) : 1 – 29.
- [23] Wiebe J , Wilson T , Bruce R , et al. Learning subjective language [J]. *Computational Linguistics* , 2004 , 30(3) : 277 – 308.
- [24] 郝媛媛 , 叶 强 , 李一军. 基于影评数据的在线评论有用性影响因素研究 [J]. *管理科学学报* , 2010 , 13(8) : 78 – 96.
Hao Yuanyuan , Ye Qiang , Li Yijun. Research on online impact factors of customer reviews usefulness based on movie reviews data [J]. *Journal of Management Sciences in China* , 2010 , 13(8) : 78 – 96. (in Chinese)
- [25] Abbasi A , Chen H C , Thoms S , et al. Affect analysis of web forums and blogs using correlation ensembles [J]. *IEEE Transactions on Knowledge and Data Engineering* , 2008 , 20(9) : 1168 – 1180.
- [26] Chen C C , Tseng Y D. Quality evaluation of product reviews using an information quality framework [J]. *Decision Support Systems* , 2011 , 50(4) : 755 – 768.
- [27] Lau R Y K , Liao S S Y , Xu K. An empirical study of online consumer review spam: A design science approach [C]// *31st International Conference on Information Systems (ICIS)* , St. Louis , USA: Association of Information Systems , 2010: 103.
- [28] Sen S , Lerman D. Why are you telling me this? An examination into negative consumer reviews on the web [J]. *Journal of Interactive Marketing* , 2007 , 21(4) : 76 – 94.
- [29] Hao Y Y , Ye Q , Li Y J , et al. How does the valence of online consumer reviews matter in consumer decision making? Differences between search goods and experience goods [C]// *43rd Hawaii International Conference on System Science*

- (HICSS) , Washington DC , USA: IEEE Computer Society , 2010: 1 – 10.
- [30] Gamon M. Sentiment classification on customer feedback data: Noisy data , large feature vectors , and the role of linguistic analysis [C] // Proceeding of COLING-04 , the 20th International Conference on Computational Linguistics , Geneva , USA: COLING , 2004: 841 – 847.
- [31] Hu N , Pavlou P A , Zhang J. Overcoming the J-shaped distribution of product reviews [J]. Communications of the ACM , 2009 , 52(10) : 144 – 147.
- [32] Jiang Z H , Benbasat I. Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping [J]. Journal of Management Information Systems , 2004 , 21(3) : 111 – 147.
- [33] Jiang Z H , Benbasat I. Investigating the influence of the functional mechanisms of online product presentations [J]. Information Systems Research , 2007 , 18(4) : 454 – 470.
- [34] Pavlou P A , Fygenon M. Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior [J]. MIS Quarterly , 2006 , 30(1) : 115 – 143.
- [35] Pavlou P A , Liang H G , Xue Y J. Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective [J]. MIS Quarterly , 2007 , 31(1) : 105 – 136.
- [36] Kotler P , Keller K. Marketing Management [M]. 12 ed. , Upper saddle River , NJ: Prentice-Hall , 2005.
- [37] Kohli R , Devaraj S , Mahmood M A. Understanding determinants of online consumer satisfaction: A decision process perspective [J]. Journal of Management Information Systems , 2004 , 21(1) : 115 – 135.
- [38] Mudambi S M , Schuff D. What makes a helpful online review? A study of customer reviews on Amazon. com [J]. MIS Quarterly , 2010 , 34(1) : 185 – 200.
- [39] Hevner A R , March S T , Park J , et al. Design science in information systems research [J]. MIS Quarterly , 2004 , 28(1) : 75 – 105.
- [40] Walls J G , Widmeyer G R , Sawy A E. Building an information system design theory for vigilant EIS [J]. Information Systems Research , 1992 , 3(1) : 36 – 59.
- [41] March S T , Smith G F. Design and natural-science research on information technology [J]. Decision Support System , 1995 , 15(4) : 251 – 266.
- [42] March S T , Storey V C. Design science in the information systems discipline: An introduction to the special issue on design science research [J]. MIS Quarterly , 2008 , 32(4) : 725 – 730.
- [43] Stigler G J. The economics of information [J]. Journal of Political Economy , 1961 , 69(3) : 213 – 225.
- [44] Chen Y B , Xie J H. Online consumer review: Word-of-mouth as a news element of marketing communication mix [J]. Management Science , 2008 , 54(3) : 477 – 491.
- [45] Lee T , Bradlow E. Automatic Construction of Conjoint Attributes and Levels from Online Customer Reviews [R]. Florida Decision and Information Sciences Workshop , 2007.

Utility analysis for online product review

YANG Ming , QI Wei , YAN Xiang-bin , LI Yi-jun

School of Management , Harbin Institute of Technology , Harbin 150001 , China

Abstract: Most e-commerce websites provide platforms for consumers to express their opinions on a specific product by writing online reviews. Due to the huge volumes of online product reviews available in e-commerce websites , it is a mentally exhausting , if not infeasible , process for potential consumers to go through all the reviews to make purchasing decisions. Hence , it is necessary to find a solution capable of presenting helpful reviews to consumers automatically. This paper describes various limitations of current utility analysis of online product review and discusses possible extensions that can improve utility evaluation capabilities based on consumer behavior theories , innovative data mining approaches , and applications to decision-making in network marketing.

Key words: electronic commerce; online product review; information overload; utility evaluation