

依据在线评论的商品排序方法

毕建武, 刘 洋*, 樊治平

(东北大学工商管理学院, 辽宁 沈阳 110167)

摘要: 提出了一种依据商品在线评论的基于多粒度情感强度分析和随机逼近理想点排序法的商品排序方法. 使用爬虫软件和 ICTCLAS 对消费者关注的备选商品的在线评论进行获取和预处理. 依据预处理后的评论, 通过提出多粒度情感强度分析算法确定每条评论针对商品属性的情感强度值. 通过对得到的情感强度值进行统计分析, 得到备选商品针对商品属性的多粒度情感强度分布形式的属性值. 最后, 依据得到多粒度情感强度分布形式的属性值, 采用随机逼近理想点排序法确定备选商品的排序. 基于中关村在线中的数码相机在线评论, 给出了提出方法应用的实例分析.

关键词: 商品排序; 在线评论; 多粒度情感强度; 累积分布; 随机逼近理想点排序法

中图分类号: C934 文献标识码: A 文章编号: 1000-5781(2018)03-0422-11

doi: 10.13383/j.cnki.jse.2018.03.013

Method for ranking products through online reviews

Bi Jianwu, Liu Yang*, Fan Zhiping

(School of Business Administration, Northeastern University, Shenyang 110167, China)

Abstract: How to automatically analyze the huge amounts of online reviews and rank products is a new important research topic. This paper proposes a method based on multi-granularity sentiment strength analysis and stochastic technique to order preferences for products through online reviews according to the closeness to an ideal solution (TOPSIS). In this method, online reviews of alternative products are first crawled by web crawler software and processed by ICTCLAS software. Then, according to the processed online reviews, an algorithm is given to calculate the sentiment strengths of online reviews concerning product features. Furthermore, according to the results of sentiment strength analysis, the feature values in the form of distribution concerning multi-granularity sentiment strengths can be obtained by statistical analysis. According to the obtained feature values, the ranking of alternative products can be determined by stochastic TOPSIS method. Finally, based on the online reviews on digital camera from the Zhongguancun online, a case analysis is given to illustrate the proposed method.

Key words: goods ranking; online reviews; multi-granularity sentiment strength; cumulative distribution; stochastic technique for order preference by similarity to an ideal solution

收稿日期: 2016-11-22; 修订日期: 2017-01-30.

基金项目: 国家自然科学基金资助项目(71771043; 71571039; 71271049; 71371002); 中央高校基本科研业务经费资助项目(N170605001).

*通信作者

1 引言

随着互联网的不断发展,越来越多的电子商务网站和社交媒体平台鼓励消费者在网站上发表他们已购买或使用过的商品的相关评论信息^[1,2]. 与商品卖方提供的商品描述相比,这些由消费者提供的商品在线评论信息能够更加客观的反应商品的真实情况. 一些研究结果表明,大众消费者在购买商品(尤其是价格较高的商品)之前通常会阅读关于该商品的在线评论信息,并依据商品在线评论信息做出最终的购买决策^[3-6]. 然而,由于商品在线评论属于非结构化文本数据而且其数量往往较大,如果让消费者逐条阅读和分析大量在线评论信息进而做出购买决策将会非常繁琐和困难^[7]. 因此,为了支持消费者的购买决策,如何客观的对大量的商品在线评论进行自动化分析并对相关商品进行排序是一个非常值得关注的研究问题.

目前,基于在线评论的商品排序已经引起了国内外一些学者的关注,并取得了一些研究成果. 例如 Zhang 等^[8]较早的关注到了基于在线评论的商品排序问题并提出了相应的排序方法,在其方法中,首先提出了一种动态规划技术来识别评论中的比较语句和评价语句,然后依据情感分析技术对比较语句和评价语句的情感倾向进行判定,在此基础上构建了针对同类商品比较的有向加权图,依据有向加权图采用改进的 PageRank 算法确定了商品排序. 在文献[8]的基础上, Zhang 等^[9,10]通过考虑在线评论所涉及的商品属性,评论获得的有用性投票(点赞)数量和评论发表时间等因素,对文献[8]所提出的方法进行了改进. Peng 等^[11]通过统计商品在线评论中不同属性特征词出现的频率,确定了消费者所关注的重要商品属性,在此基础上,提出了一种基于模糊 PROMETHEE 的商品排序方法. Chen 等^[12]依据同类商品的在线评论信息,通过集成主题模型, TOPSIS 和多维尺度分析提出了一种基于在线评论的市场结构可视化方法,通过使用该方法可以确定同类商品的排序. Najmi 等^[13]通过考虑商品的品牌,评论的情感倾向和评论的有用性等因素,提出了一种基于在线评论的商品综合排序方法. Yang 等^[14]同时考虑了消费者针对商品的打分评级,文本评论和对比性投票等三类信息,提出了一种基于有向加权图的商品排序方法.

已有研究对于基于在线评论的商品排序做出了重要贡献. 然而,目前针对此方面研究所取得的研究成果非常有限,仍显不足. 在已有的研究中^[8-14],通常是首先识别在线评论的正向或负向的情感倾向,再依据得到正向和负向的情感倾向进行商品排序. 目前,一些学者已经明确指出,仅识别在线评论的正向和负向的情感倾向而不考虑相同情感倾向评论的不同情感强度,是过于简化的处理方式,容易导致信息损失^[15,16]. 如果可以识别商品在线评论不同的情感强度,则可以基于大量在线评论的情感强度来确定商品的排序. 为此,本文提出一种基于多粒度情感强度分析和随机 TOPSIS 的商品排序方法. 在该方法中,首先通过提出多粒度情感强度分析算法确定每条评论针对商品属性的情感强度值,然后构建备选商品针对商品属性的多粒度情感强度分布形式的属性值,进而采用随机 TOPSIS 方法确定备选商品的排序.

2 依据在线评论的商品排序问题及排序方法

2.1 依据在线评论的商品排序问题

图1展示了一类依据商品在线评论的商品排序问题. 由图1可以看出,消费者在购买商品之前,通常会根据自身的需要和商品的价格等相关信息,初步确定几款备选商品. 为了从多个备选商品中选择最适合的商品,消费者可能会通过商品销售和评论网站获取备选商品的相关在线评论信息. 依据消费者关注的备选商品和属性,如何依据在线评论确定备选商品针对属性的评价结果,并依据属性评价结果和属性权重确定备选商品的排序,这是本文所要解决的问题. 为了便于分析说明,下面的符号用来描述该问题中所涉及的集合和变量.

$A = \{A_1, A_2, \dots, A_n\}$ 表示消费者关注的 n 个备选商品集合,其中 A_i 表示消费者关注的第 i 个备选商品, $i = 1, 2, \dots, n$. $F = \{f_1, f_2, \dots, f_m\}$ 表示消费者关注的 m 个商品属性集合,其中 f_j 表示消费者关注的

第 j 个属性, $j = 1, 2, \dots, m$. $w = (w_1, w_2, \dots, w_m)$ 表示消费者关注的备选商品属性权重向量, 其中 w_j 表示属性 f_j 的权重, 且满足 $\sum_{j=1}^m w_j = 1, w_j \geq 0, j = 1, 2, \dots, m$. 备选商品属性权重向量可以由消费者根据自身偏好预先给定. $Q = (q_1, q_2, \dots, q_n)$ 表示备选商品的评论数量向量, 其中 q_i 表示针对备选商品 A_i 的评论数量, $i = 1, 2, \dots, n$. $R_i = \{R_{i1}, R_{i2}, \dots, R_{iq_i}\}$ 表示针对备选商品 A_i 的评论集合, 其中 R_{ik} 表示针对商品 A_i 的第 k 条评论, $i = 1, 2, \dots, n, k = 1, 2, \dots, q_i$. 本文所要解决的问题是, 依据消费者提供的属性 F , 属性权重 w 和在线评论信息 $R_i, i = 1, 2, \dots, n$, 如何确定备选商品 $A_1, A_2, \dots, A_n$ 的排序, 以支持消费者进行商品购买决策.

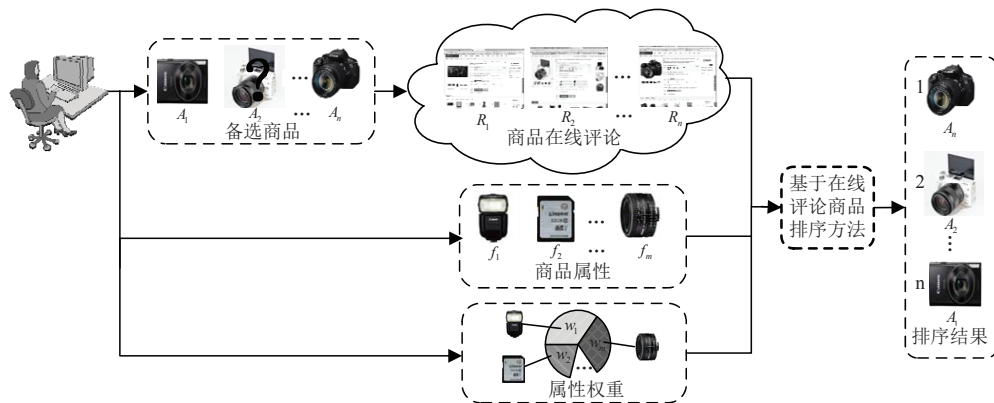


图1 基于商品在线评论的商品排序问题

Fig. 1 The goods ranking problem based on online reviews

2.2 依据在线评论的商品排序方法

为了解决上述问题, 这里给出一种基于多粒度情感强度分析和随机 TOPSIS 的商品排序方法. 该方法的基本思想是: 首先, 采用爬虫软件和 ICTCLAS 软件对消费者关注的备选商品的在线评论进行获取和预处理; 然后, 依据预处理后的评论, 通过提出多粒度情感强度分析算法确定每条评论针对消费者关注的商品属性的情感强度值; 再次, 通过对得到的情感强度值进行统计分析, 构建备选商品针对商品属性的多粒度情感强度分布形式的属性值; 最后, 依据得到属性值, 可以采用随机 TOPSIS 方法确定备选商品的排序. 下面给出该方法的具体描述.

2.2.1 备选商品在线评论获取和预处理

备选商品在线评论获取和预处理是备选商品在线评论多粒度情感强度分析的基础工作. 本文采用爬虫软件对商品在线评论进行获取, 采用 ICTCLAS 汉语分词系统(<http://ictclas.nlpir.org/>)对获取的评论进行预处理, 具体过程如下:

1) 备选商品在线评论获取

根据消费者关注的备选商品集合 $A = \{A_1, A_2, \dots, A_n\}$, 采用爬虫软件按照设定的规则对备选商品在线评论进行获取, 可以得到备选商品在线评论 $R_i = \{R_{i1}, R_{i2}, \dots, R_{iq_i}\}, i = 1, 2, \dots, n$.

2) 备选商品在线评论预处理

针对备选商品在线评论的预处理包括两个步骤, 即分词和词性标注和停用词删除. 下面分别针对这两个步骤给出具体的描述.

(a) 分词和词性标注. 采用 ICTCLAS 汉语分词系统对备选商品在线评论进行分词和词性标注. 通过分词和词性标注能够将句子形式的评论分解成若干词语并且在每个词后标注了相应的词性. 例如, 评论“画质非常好”经过分词和词性标注得到的结果为“画质/ n 非常/ d 好/ a”, 其中 n, d 和 a 分别表示名词, 副词和形容词.

(b) 停用词删除. 停用词通常是指出现频率高, 但又没有实际意义的词, 例如“了”, “的”, “呢”等. 为了提高情感强度分析的效率, 通常需要对评论中的停用词进行删除. 本文使用中文停用词表对停用词进行删除. 具体的, 将分词及词性标注处理后的评论与中文停用词表中的停用词(保留标点符号)进行比对, 并删除在停用词表中出现的词.

将经过预处理后得到的关于备选商品 A_i 的第 k 条评论的词集合记为 $WS_{ik} = \{W_{ik}^1, W_{ik}^2, \dots, W_{ik}^{q_{ik}}\}$, 其中 W_{ik}^v 表示 WS_{ik} 中的第 v 个词, q_{ik} 表示 WS_{ik} 中的词总数, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, q_i$, $v = 1, 2, \dots, q_{ik}$.

2.2.2 备选商品属性评论多粒度情感强度分析

依据备选商品在线评论预处理结果, 为了进行商品排序, 需要确定备选商品在线评论针对消费者关注的商品属性的多粒度情感强度值. 为此, 这里给出一种基于情感词典的在线评论多粒度情感强度分析方法, 该方法主要包括三个步骤, 包括备选商品属性的同义词合并, 备选商品领域情感词典建立和备选商品属性评论多粒度情感强度识别. 下面给出每个步骤的具体描述.

1) 备选商品属性的同义词合并

不同的评论者在发表评论时描述同类商品的同一属性可能使用不同的词, 因此有必要首先对描述同一商品属性的同义词进行合并. 本文采用基于词语相似度的方法来对描述同一商品属性的同义词进行合并. 具体过程如下.

首先, 依据备选商品在线评论预处理结果, 从所有备选商品的评论词集合中提取名词. 令 $WS_{\text{noun}} = \{W_{\text{noun}}^1, W_{\text{noun}}^2, \dots, W_{\text{noun}}^{q_{\text{noun}}}\}$ 表示从所有备选商品评论中提取的名词集合, 其中 W_{noun}^l 表示 WS_{noun} 中第 l 个名词, q_{noun} 表示名词集合中名词的数量, $l = 1, 2, \dots, q_{\text{noun}}$.

然后, 令 W_{f_j} 表示针对商品属性 f_j 的标准用词, $j = 1, 2, \dots, m$. 通常, 针对消费者可能关注的属性, 电子商务网站可以预先设定商品属性的标准用词, 并且要求消费者通过对所提供的标准用词进行勾选来确定消费者所关注的商品属性. 因此, 考虑针对各商品属性的标准用词为已知条件. 这样, 通过文献[17]提出的方法, 可以计算词 W_{noun}^l 与商品属性标准用词 W_{f_j} 的相似度 $\text{sim}(W_{f_j}, W_{\text{noun}}^l)$, 即

$$\text{sim}(W_{f_j}, W_{\text{noun}}^l) = \frac{\alpha}{\text{Dis}(W_{f_j}, W_{\text{noun}}^l) + \alpha}, \quad j = 1, 2, \dots, m, \quad l = 1, 2, \dots, q_{\text{noun}}, \quad (1)$$

其中 $\text{Dis}(W_{f_j}, W_{\text{noun}}^l)$ 表示依据同义词词林^[17] 计算得到的 W_{f_j} 和 W_{noun}^l 之间的距离; α 为可调参数, α 的默认取值为 1.6.

由 $\text{sim}(W_{f_j}, W_{\text{noun}}^l)$ 的值的大小可确定词 W_{f_j} 和 W_{noun}^l 是否为同义词. 具体的, 若 $\text{sim}(W_{f_j}, W_{\text{noun}}^l) \geq 0.5$, 则认为词 W_{f_j} 是 W_{noun}^l 的同义词; 如果 $\text{sim}(W_{f_j}, W_{\text{noun}}^l) < 0.5$, 则认为词 W_{f_j} 不是 W_{noun}^l 的同义词^[11,17]. 通过相似度计算, 可以得到针对词 W_{f_j} 的同义词集合 $WS_j = \{W_{f_j}^1, W_{f_j}^2, \dots, W_{f_j}^{q_{f_j}}\}$, 其中, $W_{f_j}^p$ 表示 WS_j , 将评论词集合 WS_{ik} 中的 $W_{f_j}^p$ 替换为 W_{f_j} , 即可完成同义词合并, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$, $p = 1, 2, \dots, q_{f_i}$.

通常一条商品评论中可能包含针对多个属性的评论信息, 为了识别一条评论中关于不同属性的多粒度情感强度值, 需要首先识别一条评论中针对不同属性的评论信息. 记 WS_{ik}^j 为 WS_{ik} 中的针对属性 j 的评论信息, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$. 关于 WS_{ik}^j 的确定方式如下: 将替换同义词后得到的 WS_{ik} 中的词与属性标准用词 W_{f_j} 进行比对, 如果 $W_{f_j} \in WS_{ik}$, 则提取 WS_{ik} 中包含词 W_{f_j} 的两个相邻标点符号之间的评论所包含的形容词, 动词和副词^[18], 可以得到 $WS_{ik}^j = \{W_{ik1}^j, W_{ik2}^j, \dots, W_{ikq_j}^j\}$, 其中 W_{iku}^j 表示 WS_{ik}^j 中的第 u 个词, q_j 表示 WS_{ik}^j 中词的总数, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$. 特别的, 如果 $W_{f_j} \notin WS_{ik}$, 则记 $WS_{ik}^j = \text{“ - ”}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$.

2) 备选商品领域情感词典建立

考虑到针对不同商品的情感词集合可能不同, 为了提高多粒度情感强度分析的准确性, 有必要建立商品领域情感词典. 备选商品领域情感词典建立的具体过程如下.

令 $WS' = \{W_1, W_2, \dots, W_{q'}\}$ 为针对备选商品评论的意见词集合, 其中 W_g 表示 WS' 中的第 g 个意见词, q' 表示 WS' 中词总数, $g = 1, 2, \dots, q'$. WS' 可以通过式(2)确定, 即

$$WS' = WS_{11}^1 \cup WS_{12}^1 \cup \dots \cup WS_{nq_n}^m. \quad (2)$$

依据得到的 WS' , 本文利用 HowNet (<http://www.keenage.com/>) 来构建商品领域的正向情感词典和负向情感词典. 具体的, 令 WS_{HowNet}^+ 和 WS_{HowNet}^- 分别为 HowNet 中通用的正向评价词语和负向评价词语的集合. 根据 WS_{HowNet}^+ , WS_{HowNet}^- 和 WS' 初步构建备选商品领域正向情感词典 WS_P^+ 和负向情感词典 WS_P^- , 其中

$$WS_P^+ = WS_{HowNet}^+ \cap WS', \quad (3)$$

$$WS_P^- = WS_{HowNet}^- \cap WS'. \quad (4)$$

需要指出的是, 由于可能出现 WS' 中的部分词同时不属于 WS_{HowNet}^+ 和 WS_{HowNet}^- 的情况, 针对以上情况需要通过人工识别来确定相应词所隶属的领域情感词典, 并最终得到备选商品领域情感词典. 记 WS^+ 和 WS^- 分别表示最终确定的备选商品领域正向情感词典和备选商品领域负向情感词典.

3) 备选商品属性评论多粒度情感强度识别

令 s_{ik}^j 表示商品 A_i 的评论词集合 WS_{ik} 中针对属性 f_j 的情感强度值, $i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$. 在本文中, 将评论的情感强度划分为 7 个粒度. 特别的, 若商品 A_i 的评论 WS_{ik} 中不包含针对属性 f_j 的评论, 即 $WS_{ik}^j = \text{"-"}$, 则记 $s_{ik}^j = \text{"-"}$, 因此 s_{ik}^j 的值有 8 种可能的情况, 即 $s_{ik}^j \in \{-3, -2, -1, 0, 1, 2, 3, \text{"-"}\}$, 其中 $-3(3)$, $-2(2)$ 和 $-1(1)$ 分别表示负(正)向情感倾向中的“非常”, “一般”和“稍微” 3 个情感强度级别, 0 表示中性情感倾向, $i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$. 令 WS_{neg} 表示常用的中文否定词的集合(<http://www.datatang.com/data/44198>), WS_d 表示 HowNet 程度词集合(<http://www.keenage.com/>). 根据 HowNet 中划分的中文程度词级别, WS_d 可以被进一步划分为两个子程度词集合, 即“情感强度增强词集合”和“情感强度减弱词集合”, 分别用 WS_d^+ 和 WS_d^- 表示. s_{ik}^j 的值主要是依赖于将 WS_{ik}^j 中的词与 $WS^+, WS^-, WS_{neg}, WS_d, WS_d^+$ 和 WS_d^- 中的词进行比对来确定, $i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$. 令 s_{ik}^{j+} 表示 WS_{ik}^j 与 WS^+ 交集的指示变量, s_{ik}^{j-} 表示 WS_{ik}^j 与 WS^- 交集的指示变量, s_{ik}^{jneg} 表示 WS_{ik}^j 与 WS_{neg} 交集的指示变量, s_{ikd}^j 表示 WS_{ik}^j 与 WS_d 交集的指示变量, $i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$. 确定 s_{ik}^j 的情感强度分析算法的步骤如下:

步骤 1 判断 WS_{ik}^j 是否为空集, 若 $WS_{ik}^j = \emptyset$, 则 $s_{ik}^j \leftarrow 0$; 否则跳转到步骤 2;

步骤 2 判断 WS_{ik}^j 是否为“-”, 若 $WS_{ik}^j = \text{"-"}$, 则 $s_{ik}^j \leftarrow \text{"-"}$; 否则跳转到步骤 3;

步骤 3 判断 WS_{ik}^j 与 WS^+ 的交集是否为空集, 若 $WS^+ \cap WS_{ik}^j \neq \emptyset$, 则 $s_{ik}^{j+} \leftarrow 1$; 否则 $s_{ik}^{j+} \leftarrow 0$;

步骤 4 判断 WS_{ik}^j 与 WS^- 的交集是否为空集, 若 $WS^- \cap WS_{ik}^j \neq \emptyset$, 则 $s_{ik}^{j-} \leftarrow 1$; 否则 $s_{ik}^{j-} \leftarrow 0$;

步骤 5 判断 WS_{ik}^j 与 WS_{neg} 的交集是否为空集, 若 $WS_{neg} \cap WS_{ik}^j \neq \emptyset$, 则 $s_{ik}^{jneg} \leftarrow 1$; 否则 $s_{ik}^{jneg} \leftarrow 0$;

步骤 6 判断 WS_{ik}^j 与 WS_d 的交集是否为空集, 若 $WS_{ik}^j \cap WS_d = \emptyset$, 则 $s_{ikd}^j \leftarrow 0$; 若 $WS_d^+ \cap WS_{ik}^j \neq \emptyset$, 则 $s_{ikd}^j \leftarrow 1$; 否则 $s_{ikd}^j \leftarrow -1$;

步骤 7 若 $s_{ik}^{j+} = s_{ik}^{j-} = 0$ 或者 $s_{ik}^{j+} = s_{ik}^{j-} = 1$, 则 $s_{ik}^j \leftarrow 0$; 若 $s_{ik}^{j+} = 1, s_{ik}^{j-} = 0$ 和 $s_{ik}^{jneg} = 1$ 或者 $s_{ik}^{j+} = 0, s_{ik}^{j-} = 1$, 和 $s_{ik}^{jneg} = 0$, 则 $s_{ik}^j \leftarrow -2 - s_{ikd}^j$, 否则 $s_{ik}^j \leftarrow 2 + s_{ikd}^j$.

令 $T_{ik}^j = (\alpha_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \varepsilon_{ik}^j, \zeta_{ik}^j, \eta_{ik}^j)$ 表示情感强度值 s_{ik}^j 的指示向量, $\alpha_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \varepsilon_{ik}^j, \zeta_{ik}^j, \eta_{ik}^j =$

0 或 1, 且满足 $\alpha_{ik}^j + \beta_{ik}^j + \gamma_{ik}^j + \delta_{ik}^j + \varepsilon_{ik}^j + \zeta_{ik}^j + \eta_{ik}^j \leq 1, i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$. T_{ik}^j 可以通过式(5) 来确定, 即

$$T_{ik}^j = \begin{cases} (1, 0, 0, 0, 0, 0), & \text{如果 } s_{ik}^j = -3 \\ (0, 1, 0, 0, 0, 0), & \text{如果 } s_{ik}^j = -2 \\ (0, 0, 1, 0, 0, 0), & \text{如果 } s_{ik}^j = -1 \\ (0, 0, 0, 1, 0, 0), & \text{如果 } s_{ik}^j = 0 \\ (0, 0, 0, 0, 1, 0), & \text{如果 } s_{ik}^j = 1 \\ (0, 0, 0, 0, 0, 1), & \text{如果 } s_{ik}^j = 2 \\ (0, 0, 0, 0, 0, 1), & \text{如果 } s_{ik}^j = 3 \\ (0, 0, 0, 0, 0, 0), & \text{如果 } s_{ik}^j = \text{"-"}, \end{cases} \quad (5)$$

其中 $i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$.

2.2.3 备选商品排序

基于备选商品属性评论情感强度分析结果, 即 $T_{ik}^j = (\alpha_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \varepsilon_{ik}^j, \zeta_{ik}^j, \eta_{ik}^j), i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$, 可以构建备选商品针对属性的多粒度情感强度分布形式的属性值. 在此基础上, 可以采用随机 TOPSIS 方法, 对备选商品进行排序. 具体计算过程描述如下.

令 $T_i^j = (\rho_i^j, \varsigma_i^j, \sigma_i^j, \tau_i^j, \nu_i^j, \varphi_i^j, \chi_i^j)$ 表示针对商品 A_i 的关于属性 f_j 的评论情感强度值统计向量, 根据 $T_{ik}^j = (\alpha_{ik}^j, \beta_{ik}^j, \gamma_{ik}^j, \delta_{ik}^j, \varepsilon_{ik}^j, \zeta_{ik}^j, \eta_{ik}^j), i = 1, 2, \dots, n, j = 1, 2, \dots, m, k = 1, 2, \dots, q_i$, 可计算

$$T_i^j = \left(\sum_{k=1}^{q_i} \alpha_{ik}^j, \sum_{k=1}^{q_i} \beta_{ik}^j, \sum_{k=1}^{q_i} \gamma_{ik}^j, \sum_{k=1}^{q_i} \delta_{ik}^j, \sum_{k=1}^{q_i} \varepsilon_{ik}^j, \sum_{k=1}^{q_i} \zeta_{ik}^j, \sum_{k=1}^{q_i} \eta_{ik}^j \right), \quad (6)$$

其中 $i = 1, 2, \dots, n, j = 1, 2, \dots, m$.

令 κ_i^j 表示备选商品 A_i 的评论中包含属性 f_j 的评论的数量, 则

$$\kappa_i^j = \rho_i^j + \varsigma_i^j + \sigma_i^j + \tau_i^j + \nu_i^j + \varphi_i^j + \chi_i^j. \quad (7)$$

依据 $T_i^j = (\rho_i^j, \varsigma_i^j, \sigma_i^j, \tau_i^j, \nu_i^j, \varphi_i^j, \chi_i^j)$ 和 κ_i^j 可以确定商品 A_i 关于属性 f_j 的多粒度情感强度分布形式的属性值 $P_{ij}(x), i = 1, 2, \dots, n, j = 1, 2, \dots, m$, 即

$$P_{ij}(-3) = \rho_i^j / \kappa_i^j, \quad P_{ij}(-2) = \varsigma_i^j / \kappa_i^j, \quad P_{ij}(-1) = \sigma_i^j / \kappa_i^j, \quad P_{ij}(0) = \tau_i^j / \kappa_i^j,$$

$$P_{ij}(1) = \nu_i^j / \kappa_i^j, \quad P_{ij}(2) = \varphi_i^j / \kappa_i^j, \quad P_{ij}(3) = \chi_i^j / \kappa_i^j, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m.$$

针对 $P_{ij}(x), x = -1, -2, -3, 0, 1, 2, 3$ 的表达式, 相应的累积分布函数可以写为

$$F_{ij}(x) = \begin{cases} 0, & x < -3 \\ \rho_i^j / \kappa_i^j, & -3 \leq x < -2 \\ (\rho_i^j + \varsigma_i^j) / \kappa_i^j, & -2 \leq x < -1 \\ (\rho_i^j + \varsigma_i^j + \sigma_i^j) / \kappa_i^j, & -1 \leq x < 0 \\ (\rho_i^j + \varsigma_i^j + \sigma_i^j + \tau_i^j) / \kappa_i^j, & 0 \leq x < 1 \\ (\rho_i^j + \varsigma_i^j + \sigma_i^j + \tau_i^j + \nu_i^j) / \kappa_i^j, & 1 \leq x < 2 \\ (\rho_i^j + \varsigma_i^j + \sigma_i^j + \tau_i^j + \nu_i^j + \varphi_i^j) / \kappa_i^j, & 2 \leq x < 3 \\ 1, & x \geq 3, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m. \end{cases} \quad (8)$$

在此基础上, 可以采用随机 TOPSIS 方法^[19], 对备选商品进行排序.

令 $\mathbf{F}^+ = (F_1^+(x), F_2^+(x), \dots, F_m^+(x))$ 和 $\mathbf{F}^- = (F_1^-(x), F_2^-(x), \dots, F_m^-(x))$ 分别表示正理想方案

和负理想方案所对应的属性值向量. 依据文献[19]可知, $F_j^+(x)$ 和 $F_j^-(x)$ 可分别采用式(9)和式(10)确定, 即

$$F_j^+(x) = \min\{F_{ij}(x) | i = 1, 2, \dots, n\}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad (9)$$

$$F_j^-(x) = \max\{F_{ij}(x) | i = 1, 2, \dots, n\}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m. \quad (10)$$

备选商品 A_i 关于属性 f_j 的多粒度情感强度分布形式的属性值 $P_{ij}(x)$ 的累积分布函数 $F_{ij}(x)$ 到 $F_j^+(x)$ 和 $F_j^-(x)$ 的距离可以采用式(11)和式(12)进行计算, 即

$$D(F_{ij}(x), F_j^+(x)) = \sum_{h=-3}^3 (F_{ij}(h) - F_j^+(h)), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad (11)$$

$$D(F_{ij}(x), F_j^-(x)) = \sum_{h=-3}^3 (F_j^-(h) - F_{ij}(h)), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad (12)$$

其中 $F_{ij}(h)$, $F_j^+(h)$ 和 $F_j^-(h)$ 分别表示 $F_{ij}(x)$, $F_j^+(x)$ 和 $F_j^-(x)$ 在 $x = h$ 处的取值.

依据 $F_j^+(x)$, $F_j^-(x)$, $D(F_{ij}(x), F_j^+(x))$, $D(F_{ij}(x), F_j^-(x))$ 和消费者给出的商品属性权重向量 w , 分别计算备选商品 A_i 到正、负理想累积分布向量 F^+ 和 F^- 的距离, 即

$$D_i^+ = \sum_{j=1}^m w_j D(F_{ij}(x), F_j^+(x)), \quad i = 1, 2, \dots, n, \quad (13)$$

$$D_i^- = \sum_{j=1}^m w_j D(F_{ij}(x), F_j^-(x)), \quad i = 1, 2, \dots, n. \quad (14)$$

最后, 依据 D_i^+ 和 D_i^- 可以计算备选商品 A_i 的贴近度

$$C_i = D_i^- / (D_i^- + D_i^+), \quad i = 1, 2, \dots, n. \quad (15)$$

可见, C_i 的值越大, 备选商品 A_i 越优. 因此按照计算得到的 C_i 的值的大小对所有备选商品排序, 可以得到所有备选商品的优劣排序结果.

综上, 下面给出依据商品在线评论的基于多粒度情感强度分析和随机TOPSIS的商品排序方法的具体计算步骤.

步骤 1 采用爬虫软件获取备选商品在线评论 $R_i = \{R_{i1}, R_{i2}, \dots, R_{qi}\}$, 对 R_i 进行预处理得到每条评论的词集合 $WS_{ik} = \{W_{ik}^1, W_{ik}^2, \dots, W_{ik}^{q_{ik}}\}$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, q_i$;

步骤 2 根据式(1), 对备选商品属性的同义词进行合并, 并确定每条评论中针对不同属性的评论信息 WS_{ik}^j , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$;

步骤 3 根据式(2)~式(4), 建立商品领域正向情感词典 WS^+ 和负向情感词典 WS^- ;

步骤 4 根据算法1, 计算备选商品属性评论多粒度情感强度值 s_{ik}^j , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, q_i$;

步骤 5 根据式(5)~式(8), 构建备选商品针对商品属性的多粒度情感强度分布形式的属性值 $P_{ij}(x)$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$;

步骤 6 根据式(9)~式(10), 构建正、负理想累积分布向量 F^+ 和 F^- ;

步骤 7 根据式(11)~式(15), 计算备选商品 A_i 的贴近度 C_i , 并根据 C_i 由大到小对备选商品进行排序, $i = 1, 2, \dots, n$.

3 算例分析

为了进一步说明本文提出方法的潜在应用, 本部分给出一个依据在线评论信息对多款相机商品排序的

算例分析. 考虑某消费者欲购买一款价格在 1 万元左右的数码相机, 通过多方面了解信息, 初步确定了 4 款备选数码相机, 即

A_1 : 佳能 6D 套机(24 mm~105 mm);

A_2 : 佳能 7D 套机(18 mm~135 mm);

A_3 : 尼康 D610 套机(24 mm~120 mm);

A_4 : 尼康 D750 套机(24 mm~120 mm).

该消费者关注的备选相机属性为: 性价比(f_1), 操控(f_2), 画质(f_3), 电池(f_4), 镜头(f_5), 对焦(f_6), 快门(f_7), 并且该消费者给出的备选相机属性权重向量为 $w = (0.2, 0.1, 0.2, 0.1, 0.1, 0.1, 0.2)$.

为了支持该消费者做出合理的购买决策, 需要依据在线评论对以上 4 款备选相机进行排序.

首先, 以中关村在线(<http://www.zol.com.cn/>)提供的商品点评作为备选相机评论来源, 使用 Locoy Spider 软件采集备选相机在线评论, 得到备选相机在线评论 $R_i = \{R_{i1}, R_{i2}, \dots, R_{iq_i}\}$, 对 R_i 进行预处理得到每条评论的词集合 $WS_{ik} = \{W_{ik}^1, W_{ik}^2, \dots, W_{ik}^{q_{ik}}\}$, $i = 1, 2, 3, 4$, $k = 1, 2, \dots, q_i$, $q_1 = 402$, $q_2 = 201$, $q_3 = 220$, $q_4 = 350$. 以备选相机 A_1 为例, 对其评论进行预处理的结果如表 1 所示.

根据式(1), 对描述相机属性的同义词进行合并, 并识别每条评论中针对不同属性的评论信息, 即确定 WS_{ik}^j . 这里以备选相机 A_1 的第一条评论 R_{11} 为例进行说明, R_{11} 中仅包含画质(f_3)的评论语句, 对该条评论进行同义词合并, 提取包含描述属性 f_3 的词 W_{f_3} (画质)的两个相邻标点符号之间的形容词, 动词和副词可以得到该条评论中针对画质(W_{f_3})的评论信息为 $WS_{11}^3 = \{\text{非常/d, 好/a}\}$.

表 1 备选相机 A_1 的评论预处理结果

Table 1 The pre-processing results of the reviews concerning alternative camera A_1

评论序号	预处理结果
1	机身/n 轻巧/a, /wd 解除/v 腰/n 酸/a 背/n 痛/a(之) 苦/an. /w 高/a 感/vg 好/a, /w 画质/n 效果/n 非常/d 好/a. 佳能/nz(看) 应付/v 产品/n 迟迟/d 不/d 上市/v(避免) 尴尬/a ?/w
2	画质/n 赞/v, /w 高/a 感/vg 赞/v, /w 暗/a 部/n 对焦/n 能力/n 赞/v
⋮	⋮
402	操控/n 方便/a, /w 画质/n 漂亮/a, /w 套头/n 不/d(像) 网上/s 评价/v(那样), /w 挺/d 好/a. /w

然后, 根据式(2)~式(4), 建立相机领域正向情感词典 WS^+ 和负向情感词典 WS^- , 部分情感词见表 2.

表 2 部分情感词

Table 2 Partial sentiment words

WS^+	WS^-
满意、棒、好、赞、优秀、大气、精致、精细、高档 ……	差、惨、多、不行、离谱、毛糙、弱、寒酸 ……

根据情感强度分析算法, 计算备选相机针对属性的评论的情感强度值 s_{ik}^j , $i = 1, 2, 3, 4$, $k = 1, 2, \dots, q_i$, $q_1 = 402$, $q_2 = 201$, $q_3 = 220$, $q_4 = 350$. 这里以 $WS_{11}^3 = \{\text{非常/d, 好/a}\}$ 为例来进一步说明如何通过情感强度分析算法确定 s_{ik}^j 的值. 由于 WS_{11}^3 中仅存在正向情感词“好”和情感强度增强词“非常”, 即 $WS_{11}^3 \cap WS^+ \neq \emptyset$, $WS_{11}^3 \cap WS^- = \emptyset$, $WS_{11}^3 \cap WS_{\text{neg}} = \emptyset$ 和 $WS_{11}^3 \cap WS_d^+ \neq \emptyset$, 由情感强度分析算法的步骤 3, 步骤 4, 步骤 5 和步骤 6 可分别确定各指示变量的值, 即 $s_{11}^{3+} = 1$, $s_{11}^{3-} = 0$, $s_{11d}^3 = 1$ 和 $s_{11}^{3\text{neg}} = 0$. 在此基础上, 可以由情感强度分析算法的步骤 7 确定 s_{11}^3 的值, 即 $s_{11}^3 = 2 + 1 = 3$. 最终得到该条评论关于属性 f_3 的情感强度值为 3.

进一步地, 依据式(5)~式(8)和 s_{ik}^j , 确定相机 A_i 关于属性 f_3 的多粒度情感强度分布形式的属性值 $P_{ij}(x)$, $i = 1, 2, 3, 4$, $k = 1, 2, \dots, q_i$, $q_1 = 402$, $q_2 = 201$, $q_3 = 220$, $q_4 = 350$, 结果如表 3 所示.

依据表 3 和式(8), 可以得到针对 $P_{ij}(x)$ 的累积分布函数 $F_{ij}(x)$, $i = 1, 2, 3, 4$, $j = 1, 2, \dots, 7$. 为了节省篇幅, 这里以 $F_{11}(x)$ 为例, 简要说明其计算过程.

表 3 备选相机关于属性的多粒度情感强度分布形式的属性值

Table 3 Feature values in the form of distribution concerning multi-granularity sentiment strengths on alternative cameras

备选相机	属性	-3	-2	-1	0	1	2	3
A ₁	f_1	0.000 0	0.000 0	0.083 4	0.000 0	0.520 8	0.270 8	0.125 0
	f_2	0.024 4	0.073 2	0.414 6	0.000 0	0.463 4	0.024 4	0.000 0
	f_3	0.000 0	0.008 5	0.042 4	0.000 0	0.550 8	0.355 9	0.042 4
	f_4	0.000 0	0.045 5	0.090 9	0.136 4	0.545 5	0.181 8	0.000 0
	f_5	0.020 8	0.020 8	0.083 3	0.333 3	0.541 7	0.000 0	0.000 0
	f_6	0.011 9	0.059 5	0.494 0	0.029 8	0.339 3	0.059 5	0.006 0
	f_7	0.010 9	0.021 7	0.489 1	0.163 0	0.271 7	0.032 6	0.010 9
A ₂	f_1	0.000 0	0.042 6	0.446 8	0.000 0	0.319 1	0.106 4	0.085 1
	f_2	0.000 0	0.000 0	0.030 3	0.090 9	0.333 3	0.454 5	0.090 9
	f_3	0.000 0	0.063 5	0.476 2	0.095 2	0.333 3	0.015 9	0.015 9
	f_4	0.000 0	0.000 0	0.291 7	0.250 0	0.375 0	0.083 3	0.000 0
	f_5	0.019 6	0.039 2	0.392 2	0.411 8	0.117 6	0.019 6	0.000 0
	f_6	0.010 3	0.020 6	0.113 4	0.051 5	0.649 5	0.134 0	0.020 6
	f_7	0.000 0	0.021 3	0.148 9	0.042 6	0.723 4	0.042 6	0.021 3
A ₃	f_1	0.000 0	0.000 0	0.114 3	0.114 3	0.600 0	0.085 7	0.085 7
	f_2	0.000 0	0.000 0	0.263 2	0.105 3	0.631 6	0.000 0	0.000 0
	f_3	0.000 0	0.032 3	0.064 5	0.000 0	0.677 4	0.225 8	0.000 0
	f_4	0.000 0	0.058 8	0.117 6	0.352 9	0.235 3	0.176 5	0.058 8
	f_5	0.000 0	0.000 0	0.181 8	0.318 2	0.363 6	0.090 9	0.045 5
	f_6	0.031 3	0.062 5	0.593 8	0.125 0	0.156 3	0.031 3	0.000 0
	f_7	0.000 0	0.027 8	0.583 3	0.166 7	0.194 4	0.027 8	0.000 0
A ₄	f_1	0.000 0	0.000 0	0.164 4	0.191 8	0.424 7	0.109 6	0.109 6
	f_2	0.000 0	0.000 0	0.115 4	0.076 9	0.769 2	0.038 5	0.000 0
	f_3	0.000 0	0.037 0	0.074 1	0.074 1	0.777 8	0.037 0	0.000 0
	f_4	0.000 0	0.000 0	0.117 6	0.352 9	0.470 6	0.058 8	0.000 0
	f_5	0.000 0	0.013 7	0.164 4	0.342 5	0.383 6	0.082 2	0.013 7
	f_6	0.012 2	0.024 4	0.048 8	0.097 6	0.548 8	0.207 3	0.061 0
	f_7	0.012 0	0.024 1	0.289 2	0.228 9	0.325 3	0.072 3	0.048 2

表 4 正、负理想累积分布向量

Table 4 The probability distributions of the ideal and nadir solutions

	$(-\infty, -3)$	$[-3, -2)$	$[-2, -1)$	$[-1, 0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, +\infty)$
$F_1^+(x)$	0.000 0	0.000 0	0.000 0	0.083 3	0.083 3	0.604 2	0.875 0	1.000 0
$F_2^+(x)$	0.000 0	0.000 0	0.000 0	0.030 3	0.121 2	0.454 5	0.909 1	1.000 0
$F_3^+(x)$	0.000 0	0.000 0	0.008 5	0.050 8	0.050 8	0.601 7	0.957 6	1.000 0
$F_4^+(x)$	0.000 0	0.000 0	0.000 0	0.117 6	0.272 7	0.764 7	0.941 2	1.000 0
$F_5^+(x)$	0.000 0	0.000 0	0.000 0	0.125 0	0.458 3	0.863 6	0.954 5	1.000 0
$F_6^+(x)$	0.000 0	0.000 0	0.030 9	0.085 4	0.182 9	0.731 7	0.939 0	1.000 0
$F_7^+(x)$	0.000 0	0.000 0	0.021 3	0.170 2	0.212 8	0.789 5	0.951 8	1.000 0
$F_1^-(x)$	0.000 0	0.000 0	0.042 6	0.489 4	0.571 4	0.828 6	0.964 3	1.000 0
$F_2^-(x)$	0.000 0	0.024 4	0.133 3	0.512 2	0.533 3	1.000 0	1.000 0	1.000 0
$F_3^-(x)$	0.000 0	0.000 0	0.063 5	0.539 7	0.634 9	0.968 3	1.000 0	1.000 0
$F_4^-(x)$	0.000 0	0.043 5	0.058 8	0.695 7	0.913 0	1.000 0	1.000 0	1.000 0
$F_5^-(x)$	0.000 0	0.020 8	0.058 8	0.451 0	0.862 7	1.000 0	1.000 0	1.000 0
$F_6^-(x)$	0.000 0	0.031 3	0.093 8	0.687 5	0.812 5	0.968 8	1.000 0	1.000 0
$F_7^-(x)$	0.000 0	0.012 0	0.052 6	0.611 1	0.777 8	0.972 2	1.000 0	1.000 0

由表 3 可知, 备选相机 A₁ 关于属性 f_1 的多粒度情感强度分布形式的属性值为

$$P_{11}(-3) = 0.000\ 0, P_{11}(-2) = 0.000\ 0, P_{11}(-1) = 0.083\ 4, P_{11}(0) = 0.000\ 0, \\ P_{11}(1) = 0.520\ 8, P_{11}(2) = 0.270\ 8, P_{11}(3) = 0.125\ 0.$$

依据式(8), $P_{11}(x)$ 的相应累积分布函数 $F_{11}(x)$ 可以写为

$$F_{11}(x) = \begin{cases} 0.000\ 0, & x < -3 \\ 0.000\ 0, & -3 \leq x < -2 \\ 0.000\ 0, & -2 \leq x < -1 \\ 0.083\ 4, & -1 \leq x < 0 \\ 0.083\ 4, & 0 \leq x < 1 \\ 0.604\ 2, & 1 \leq x < 2 \\ 0.875\ 0, & 2 \leq x < 3 \\ 1.000\ 0, & x \geq 3. \end{cases}$$

在此基础上, 依据式(9)和式(10), 可构建正、负理想累积分布向量 $\mathbf{F}^+ = (F_1^+(x), F_2^+(x), \dots, F_7^+(x))$ 和 $\mathbf{F}^- = (F_1^-(x), F_2^-(x), \dots, F_7^-(x))$, 结果如表4所示.

根据式(11)~式(14), 计算备选相机 A_i 到正向和负向理想累积分布向量 $\mathbf{F}^+$ 和 $\mathbf{F}^-$ 的距离, 即 D_i^+ 和 D_i^- , $i = 1, 2, 3, 4$. 计算结果为 $D_1^+ = 0.533\ 1$, $D_2^+ = 0.744\ 9$, $D_3^+ = 0.724\ 1$, $D_4^+ = 0.497\ 9$, $D_1^- = 0.872\ 4$, $D_2^- = 0.660\ 5$, $D_3^- = 0.681\ 3$, $D_4^- = 0.907\ 5$. 最后, 依据式(15), 可计算备选相机的贴近度, 计算结果为 $C_1 = 0.620\ 7$, $C_2 = 0.470\ 0$, $C_3 = 0.484\ 7$, $C_4 = 0.645\ 7$. 通过比较4款备选相机的贴近度的值可得到4款相机排序结果为 $A_4 \succ A_1 \succ A_3 \succ A_2$, 即该消费者可以考虑购买备选相机 A_4 .

4 结束语

本文给出了一种依据商品在线评论的基于多粒度情感强度分析和随机 TOPSIS 的商品排序方法. 在该方法中, 首先, 采用爬虫软件和 ICTCLAS 对消费者关注的备选商品的在线评论信息进行获取和预处理. 然后, 依据预处理后的评论, 通过提出的多粒度情感强度分析算法可以确定每条评论针对消费者关注的商品属性的情感强度值. 进一步地, 通过对得到的情感强度值进行统计分析, 可以构建备选商品针对商品属性的多粒度情感强度分布形式的属性值. 在此基础上, 可以依据随机 TOPSIS 方法确定备选商品的排序. 该方法具有概念清晰、计算简单等特点, 有较强的可操作性和实用性, 为解决依据在线评论的商品排序问题提供了一种新的思路.

需要强调的是, 在本文研究中, 提出了一种多粒度情感强度分析算法. 使用该算法, 可以将在线评论的情感强度划分为七个情感粒度, 进而通过统计分析, 可以将海量在线评论中所蕴含的针对商品的情感强度转化为多粒度情感强度分布形式的属性值. 这种处理方式, 一方面避免了仅考虑正向和负向情感倾向所造成的大量信息损失, 另一方面方便借助已有的基于随机(频度)分布的信息处理和决策分析方法进行基于海量在线评论信息的信息处理与决策分析, 为进一步开展基于在线评论信息的决策分析奠定了良好的基础.

参考文献:

- [1] Chen H, Chiang R H L, Storey V C. Business intelligence and analytics From big data to big impact. *MIS Quarterly*, 2012, 36(4): 1165–1188.
- [2] 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 pricing strategies. *Journal of System Engineering*, 2014, 29(4): 560–570. (in Chinese)
- [3] Hennig-Thurau T, Gwinner K P, Walsh G, et al. Electronic word-of-mouth via consumer opinion platforms: What motivates consumers to articulate themselves on the internet. *Journal of Interactive Marketing*, 2004, 18(1): 38–52.
- [4] Liu Y, Bi J W, Fan Z P. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 2017, 36: 149–161.

- [5] Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 2004, 80(2): 159–169.
- [6] 张紫琼, 叶强, 李一军. 互联网商品评论情感分析研究综述. *管理科学学报*, 2010, 13(6): 84–96.
Zhang Z Q, Ye Q, Li Y J. Literature review on sentiment analysis of online product reviews. *Journal of Management Sciences in China*, 2010, 13(6): 84–96. (in Chinese)
- [7] Zhang W, Xu H, Wan W. Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications*, 2012, 39(11): 10283–10291.
- [8] Zhang K, Narayanan R, Choudhary A. Mining Online Customer Reviews for Ranking Products. Technical Report, EECS Department, Northwestern University, 2009.
- [9] Zhang K, Narayanan R, Choudhary A. Voice of the customers: Mining online customer reviews for product feature-based ranking // *Proceedings of the 3rd Conference on Online Social Networks*. 2010.
- [10] Zhang K, Cheng Y, Liao W, et al. Mining millions of reviews: A technique to rank products based on importance of reviews // *Proceedings of the 13th ACM International Conference on Electronic Commerce*. 2011.
- [11] Peng Y, Kou G, Li J. A fuzzy PROMETHEE approach for mining customer reviews in Chinese. *Arabian Journal for Science and Engineering*, 2014, 39(6): 5245–5252.
- [12] Chen K, Kou G, Shang J, et al. Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, And multi-dimensional scaling approaches. *Electronic Commerce Research And Applications*, 2015, 14(1): 58–74.
- [13] Najmi E, Hashmi K, Malik Z, et al. CAPRA: A comprehensive approach to product ranking using customer reviews. *Computing*, 2015, 97(8): 843–866.
- [14] Yang X, Yang G, Wu J. Integrating rich and heterogeneous information to design a ranking system for multiple products. *Decision Support Systems*, 2016, 84: 117–133.
- [15] Serrano-Guerrero J, Olivas J A, Romero F P, et al. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 2015, 311: 18–38.
- [16] Tang H, Tan S, Cheng X. A survey on sentiment detection of reviews. *Expert Systems with Applications*, 2009, 36(7): 10760–10773.
- [17] Liu Q, Li S. Word similarity computing based on How-Net // *Proceedings of the 3th Chinese Lexical Semantic Workshop*. 2002.
- [18] Huang S L, Cheng W C. Discovering Chinese sentence patterns for feature-based opinion summarization. *Electronic Commerce Research and Applications*, 2015, 14(6): 582–591.
- [19] Fan Z P, Zhang X, Liu Y, et al. A method for stochastic multiple attribute decision making based on concepts of ideal and anti-ideal points. *Applied Mathematics and Computation*, 2013, 219(24): 11438–11450.

作者简介:

毕建武(1988—), 男, 河北唐山人, 博士生, 研究方向: 大数据与管理决策分析, Email: jianwubi@126.com;

刘洋(1978—), 男, 辽宁沈阳人, 博士, 副教授, 博士生导师, 研究方向: 大数据管理决策分析, Email: liuy@mail.neu.edu.cn;

樊治平(1961—), 男, 江苏镇江人, 博士, 教授, 博士生导师, 研究方向: 运作管理与决策分析, Email: zpfan@mail.neu.edu.cn.