

2015 年全国大学生 统计建模大赛 参赛论文

论文题目 产品评价告诉了我们什么——基于国美电热水器
数据分析¹

作者姓名 吴梓栩、黄冬梅、黄德辉

指导老师 陈军才、林海明

参赛单位 广东财经大学

¹ 注:该论文获得由中国统计教育学会举办的“2015 年(第四届)全国大学生统计建模大赛”大数据统计建模类本科生组三等奖。

产品评价告诉了我们什么？

——基于国美电热水器数据分析

摘要

商品评论信息鱼龙混杂，不仅误导消费者的购买行为，也使厂商得不到有效的产品销售反馈，为此，本文采用“文本情感法”，根据一定的指标，用 Excel、Access 剔除通过数据爬虫得到的 49833 条评论中的水军，筛选出有效评论，并通过 SAS 聚类法、熵值法建立数学模型，以实证分析的方式，证明消费者评价的影响力，并从中挖掘消费者需求，获得产品的真实性质，为产商，销售商及潜在用户提供建议。

关键词：产品评价 产品三层次理论 文本情感分析 聚类 熵值法

一、引言

Web2.0 的蓬勃发展，改变了 Web1.0 时代用户被动接受门户网站编辑处理后的信息的情况，实现了网站与用户之间的互动，为用户获取信息、发表意见和情感提供了新的渠道。越来越多的用户在互联网上发表自己的观点，而基于 Web2.0 发展起来的 C2C（以淘宝、易趣为代表）、B2C（国美、京东为代表）等电子商务也呈现欣欣向荣的状态，广大消费者在各大电商平台上留下自己的用户体验，对潜在客户了解商品的服务和质量，作出是否购买决策产生影响，也帮助销售商和生产商获得消费者关于产品体验情况的反馈。但是，随着电商的发展，出现了越来越多的虚假评价，即参与评价的用户中存在部分水军，使得商品评论的作用大打折扣。他们的行为，不仅误导了消费者，也将对平台的信誉造成影响。如何剔除水军，在承载用户情感，关注度的评价中发现产品真实性质，了解用户需求显得更为重要。目前，网络购物评价的相关研究主要包括以下四个方面：

（一）电商平台的研究方向

由于电商发展至今历时较短，相关的研究并不充分，在目前已有的关于电商平台的研究中，主要分为两个方面，一方面是研究结合金融和互联网模式的以电商平台为核心的电商金融，分析其参与要素，金融模式（包括消费者信贷，中小微企业信贷）^[1]，剖析当下阿里巴巴、京东、腾讯等电商的发展情况存在的问题及面临的挑战^{[2][3]}；另一方面是通过当下电商的运营情况，分析电商的发展趋势，为电商运作者提供运营建议^[4]。近年来，我国网民规模快速扩大，网络购物市场飞速发展，网络购物的使用率已达到 33.8%，用户规模达到 1.42 亿^[5]，但作为基于 Web2.0 发展起来的电商的重要参与者——用户，其消费体验却鲜有研究，即使在网上购物消费者行为的早期研究中，大多数也都侧重研究消费者如何采用网络购物，以及网购意向与网购行为之间的关系。

（二）文本情感分析在网购评论中的应用

1. 情感分析法

情感分析是数据挖掘的一个新兴领域，具有重要的学术研究价值和实际应用价值，受到国外学术界和企业界的广泛关注。所谓情感分析，王洪伟等人^[6]认为是对用户发表在 Web 上的评论进行分析，从而识别出隐含在其中的情感信息，并发现用户情感演变的规律；赵妍妍等人^[7]认为情感分析又称作评论挖掘或意见挖掘，是对带有情感色彩的主观性文本进行提取、分析、处理、归纳和推理的过程。按照处理文本的类型不同，可以分为基于舆情的情感分析和基于

商品评论的情感分析。刘倩等人^[8]指出目前针对中文文本进行情感分析的研究较少，尤其是针对中文网络评论的情感分析研究更少。国外在这方面有较丰富的经验，如：康奈尔大学提供了关于影评数据集的资源库。

2.文本情感法在商品评论中的使用

奚金金等人^[9]提出了基于词典和统计语言（R-Gram）的情感评价词语的抽取、判别、分类、归纳的方法；钟将等人^[10]提出了基于矩阵投影（MP）和归一化向量（NLV）的文本算法实现对商品评价的情感分析，以实现在准确判断商品评价情感倾向的同时提高识别效率的目的。

（三）产品层次理论

1988 年著名的市场营销专家 P·科特勒，在其著作《市场管理：分析、计划、执行与控制》，提出了一个产品的三层结构理论。该理论认为，任何一种产品都可被分为三个层次：核心产品，即使用价值或效用；有形产品，包括式样、品牌、名称、包装等；附加产品，即附加服务或利益，并认为这三个层次是互相联系的有机整体。1994 年，其在该书的专注修订版中，将产品概念的内容由三层次结构说扩展为五层次结构说，即包括核心利益、一般产品、期望产品、扩大产品和潜在产品。三层次结构说是对产品的静态描述，侧重于企业和生产者对产品的主导作用，注重核心利益、形式产品和附加产品的分析，但缺乏面向客户的经营意识，忽视了消费者对产品需求和选购的根本决定权。而五层次结构理论从全新的角度去分析产品，强调消费者选购和使用对产品的决定性，具有面向市场和客户的意识，从产品的使用和消费过程进行分析，揭示了效用构成和价值实现的动态性。

（四）水军的识别

网络水军分两种，一种是自发的喜欢在论坛“灌水”的人或群体，一般称其为“灌水者”；另一种是出于商业或私人目的专门组织的网络群体，通过雇用大批人手在互联网上集体炒作，以达到宣传、推销或攻击某些人或商品的目的。这类为了商业利益以各种手法在网络上发帖、灌水的人群，被称作“网络水军”。“水军”制造的“水帖”不仅没有任何信息含量还污染了网络中有价值的信息，扰乱网络秩序。虽然得到学术界和工业界的广泛关注，但由于水军规模大，且行为逐渐趋向正常用户，该方面研究面临很大挑战，真正完善的用于识别和防范“水军”的体系也尚未建立。基于现状，南京师范大学的许敏^[11]提出了基于用户特征的虚假评价用户识别方法，分别根据卖家用户、买家用户各自之间的特征差异进行分析，构建用户相似度网络，通过已识别的水军用户和非水军用户在共同评价商品数量上的差别，识别整个集合中的水军用户和非

水军用户。华中科技大学新闻与信息传播学院的刘秋文^[12]从“主题相关性”、“排列格式”、“发帖频率”、“互动性”四个方面给出了识别“网络水军”的标准。

二、数据说明及预处理

（一）数据来源

采用数据爬虫的方法，通过“火车头数据采集器”，收集到国美平台上2013年1月~2014年12月关于5种品牌38种型号电热水器共49833条相关数据。每条数据包括产品名称、型号、评论、标签、评论时间、数据爬虫网址等信息。

（二）“文本情感法”下的数据预处理

随着互联网的普及率持续上升，网购用户的数量更是大幅增长。尽管网购可以给消费者带来方便快捷的购物体验，但是网购本身无地域限制的特点产生了购物质量依据匮乏、远距离鉴别困难、商品描述信息与实际不符等弊端，越来越多的消费者在网购时都选择先去了解已购顾客对该产品的评价和态度，以便作出可靠的决策。但是评论数量的日益增多，鱼龙混杂，使得消费者难以辨别其真假。为此，本文对所得数据做了如下处理：

1. 删除高相似度评论

在对商品的评论中，不同的用户对商品的体验可能不同，对商品的评价基本不会出现表述中相同的情况。这些相同的评论可能是网络水军为了宣传该产品的某种特点或者弥补在网络购物中的不足来吸引消费者的注意力而做出的评价。因而在对评论的预处理中，要把相似度高的评论数据剔除，同时清除用符号或者其它生僻词堆砌的评论，确保评论的有效性。

2. 结合语境，确定关键词，提取指标信息

在网络客户发表的评论中，有直接描述产品各种具体功能的词或短语，比如质量、价格、物流等。也有通过特殊的语境表达对产品的某一特征的感觉的词语，这部分词语的提取需要结合对语意的理解，提取难度大。

在国美电商平台上，首先利用电商平台对于热水器特点的相关宣传、商品评价栏上生成的推荐标签或特性和随机抽取的经过预处理的评论，形成对产品特点初步认识，并且对特征进行分类，分为安装、物流、性能、质量、价格、外观、服务、品牌八类。同时结合汉语词典、新华词典中的解释找到这些词的同义词、近义词或者能体现该特点的子词汇，如能体现服务的词语有服务、售后、耐心、信用、客服、细心等。基于样本的随机性和评论的多样性，为了保

证特征词的有代表性，我们从评论的数据集中随机抽取评论内容，并通过 Access 数据库查询功能，得出最后词频统计，同时选取符合本次评论出现的词语，最终筛选出如下的词语。

表 1 国美评论数据指标及特征词

类别	符号	特征词
安装	X ₁	师傅、安装
物流	X ₂	物流、快递、送货、发货、年、天、时间、日、月
性能	X ₃	容量、省电、用电、节能、性价、加热、水温、保温、功能、温度、出水量
质量	X ₄	实用、耐用、质量、安全、品质、货真价实
价格	X ₅	钱、元、费、实惠、价格、贵、便宜、优惠、划算、物超所值、物有所值
外观	X ₆	外表、外观、好看、外形、美观、漂亮、做工、省空间、美、粗糙、小巧
服务	X ₇	服务、售后、耐心、信用、客服、细心
品牌	X ₈	品牌、海尔、康宝、美的、万家乐、帅康

通过这些特征词选出来的评论个数，只能代表消费者对该次购物体验某种类别的关注情况，并不能实时反映消费者的喜好程度。因而，我们结合文本情感倾向对词语进行筛选，如在物流指标上，加上慢、很慢、很久、很快、1 天就到了、还没到、快速、差等词语进行同时筛选，就可以筛选出某种型号产品下的好评和差评。假设消费者有显厌恶或者喜欢的感觉时，才进行评论。在这样的假设下，除了明显带有差评词语的评价外，默认其他评价为好评。这种方法存在缺陷，忽略了对中性评论的筛选，但是由于筛选中性评论的标准界定不一，而且消费者如果是对产品的某些方面没有明显体验的话，很少会特地表达出来，所以这种方法具有可取性。

三、数据分析

（一）模型一：聚类分析

根据 38 种型号电热水器 8 个指标的有效评论占比的情况进行聚类从而分析消费者对产品的关注点，为商家提供改进建议。

表 2 产品编号及名称

编号 i	产品名称	编号 i	产品名称
1	海尔 ES40H-Q(E)电热水器	20	美的 F50-30W7 电热水器
2	海尔 ES50H-D5(E)电热水器	21	美的 F60-15WA1 电热水器
3	海尔 ES50H-Q(E)电热水器	22	美的 F60-21W6 (B) 电热水器
4	海尔 ES60H-D5(E)电热水器	23	美的 F80-15WA1 电热水器
5	海尔 ES60H-Q(E)电热水器	24	美的 F80-30W6 电热水器
6	海尔 ES80H-D5(E)电热水器	25	帅康 DSF-50DSQW 电热水器
7	康宝 CBD40-WA9 电热水器	26	帅康 DSF-50DWEL 电热水器
8	康宝 CBD50-WAF6 电热水器	27	帅康 DSF-50DYG 电热水器
9	康宝 CBD40-WAF7 电热水器	28	帅康 DSF-50JMF 电热水器
10	康宝 CBD50-WA9 电热水器	29	帅康 DSF-50JMG 电热水器
11	康宝 CBD50-WAF1 电热水器	30	帅康 DSF-60DYG 电热水器
12	康宝 CBD50-WAF7 电热水器	31	帅康 DSF-60JMW 电热水器
13	康宝 CBD60-WA9 电热水器	32	帅康 DSF-60JYG 电热水器
14	康宝 CBD60-WAF1 电热水器	33	万家乐 D40-GHF(B)电热水器
15	美的 F40-15WA1 电热水器	34	万家乐 D50-HG7C 电热水器
16	美的 F50-15WA1 电热水器	35	万家乐 D50-HV12F 电热水器
17	美的 F50-21W6 (B) 电热水器	36	万家乐 D60-GHF(B)电热水器
18	美的 F50-21WB1 电热水器	37	万家乐 D60-HG7C 电热水器
19	美的 F50-30W3 (B) 电热水器	38	万家乐 D60-HV12F 电热水器

我们采用最常用的八种聚类方法进行聚类，并结合以下两种处理方法^[13]选取最优聚类数：（1）根据分类问题本身的专业知识结合实际需要选择分类方法，并确定分类个数；（2）多用几种分类方法去做，把结果中的共性取出来，如果用几种方法的某些结果都一样，则说明这样的聚类确实反映了事物的本质，而将有争议的样品暂放一边或用其他方法如判别分析去归类。

表 3 八种聚类法所得聚类数

方法	聚类数
最短距离法 (<i>single</i>)	六类
最长距离法 (<i>complete</i>)	七类
类平均法 (<i>average</i>)	四类
重心法 (<i>centroid</i>)	四类
中位数法 (<i>median</i>)	四类
离差平方和法 (<i>ward</i>)	五类
可变类平均法 (<i>flexible</i>)	三类
可变法 (<i>mcquitty</i>)	五类

根据聚类过程表中立方聚类条件出现峰值所对应的分类数、伪 F 出现峰值所对应的分类数、伪 T 方出现峰值所对应的分类数^[13]，结合聚类分析图及八类聚类法的共性，得到初步聚类的结果。

表 4 八种聚类分析初步结果

类型	型号	类型	型号
第一类	美的 F50-15WA1 电热水器	第二类	万家乐 D50-HG7C 电热水器
	美的 F50-21W6 (B) 电热水器		万家乐 D60-GHF(B)电热水器
	美的 F50-30W3 (B) 电热水器		万家乐 D60-HG7C 电热水器
	美的 F50-30W7 电热水器		万家乐 D60-HV12F 电热水器
	美的 F50-21W6 (B) 电热水器	第三类	海尔 ES80H-D5(E)电热水器
第二类	康宝 CBD50-WAF6 电热水器		海尔 ES40H-Q(E)电热水器
	康宝 CBD40-WAF7 电热水器		海尔 ES50H-D5(E)电热水器
	康宝 CBD50-WA9 电热水器		海尔 ES50H-Q(E)电热水器
	康宝 CBD50-WAF1 电热水器		海尔 ES60H-D5(E)电热水器
	康宝 CBD60-WAF1 电热水器	第四类	海尔 ES60H-Q(E)电热水器
	帅康 DSF-50DSQW 电热水器	归类不清	康宝 CBD40-WA12 电热水器
	帅康 DSF-50DWEL 电热水器		康宝 CBD50-WAF10 电热水器
	帅康 DSF-50DYG 电热水器		美的 F40-15WA4 电热水器

	帅康 DSF-50JMF 电热水器		美的 F50-21WB4 电热水器
	帅康 DSF-50JMG 电热水器		美的 F60-15WA4 电热水器
	帅康 DSF-60DYG 电热水器		美的 F80-15WA4 电热水器
	帅康 DSF-60JMW 电热水器		美的 F80-30W9 电热水器
	帅康 DSF-60JYG 电热水器		万家乐 D50-HV15F 电热水器
	万家乐 D40-GHF(B)电热水器		康宝 CBD60-WA9 电热水器

通过“贝叶斯判别法”，根据已得四种类型的产品数据建立判别式，对归类不清的 9 种产品进行判别，结果如下：

表 5 原各组样品回判结果

从 type	1	2	3	4	合计
1	5	0	0	0	5
	100	0	0	0	100
2	0	18	0	0	18
	0	100	0	0	100
3	0	0	5	0	5
	0	0	100	0	100
4	0	0	0	1	1
	0	0	0	100	100
合计	5	18	5	1	29
	17.24	62.07	17.24	3.45	100
先验	0.25	0.25	0.25	0.25	

从表 5 得相应回判结果：原各组样品回判判对率为 100%，判别函数效果相当好。

表 6 新样品判别结果与后验概率

观测	产品型号	分为 type	1	2
7	康宝 CBD40-WA12 电热水器	1	1	0
11	康宝 CBD50-WAF10 电热水器	2	0	1

15	美的 F40-15WA4 电热水器	2	0.2911	0.7089
19	美的 F50-21WB4 电热水器	1	0.9979	0.0021
22	美的 F60-15WA4 电热水器	1	0.993	0.007
23	美的 F80-15WA4 电热水器	1	1	0
24	美的 F80-30W9 电热水器	2	0	1
35	万家乐 D50-HV15F 电热水器	2	0	1
13	康宝 CBD60-WA9 电热水器	2	0	1

判别法在聚类法基础上将归类不清的 9 种型号根据判别规则进行归类。

表 7 六类样品的中心均值数据

类别	\overline{X}_1	\overline{X}_2	\overline{X}_3	\overline{X}_4	\overline{X}_5	\overline{X}_6	\overline{X}_7	\overline{X}_8
第一类	0.1940	0.2002	0.3165	0.1051	0.2844	0.1950	0.1568	0.0965
第二类	0.0353	0.0368	0.0527	0.0298	0.0609	0.0377	0.0328	0.0137
第三类	0.6701	0.5860	0.7140	0.5581	0.6178	0.5585	0.4180	0.6359
第四类	0.4615	0.4118	0.0000	0.5714	0.6923	0.2941	0.5556	0.5000
总体平均	0.1773	0.1661	0.2125	0.1387	0.2123	0.1585	0.1322	0.1368

附注： \overline{X}_i 表示各个指标有效占比的均值

从表 7 的数据得出：

第一类产品：该类产品包含的 9 种型号中，8 种属于美的品牌，1 种是康宝。除了品牌指标的有效评论占比较低，各项指标的有效评论占都与总体均值接近，该类产品属于关注度一般的产品；

第二类产品：以康宝 CBD50-WAF6 电热水器、帅康 DSF-50DSQW 电热水器、万家乐 D40-GHF(B)电热水器、美的 F40-15WA4 电热水器为例，各项指标的有效评论比例均很低，大部分都小于 0.1，甚至出现 0 的情况，侧面反映出这部分产品的销售量比较低，关注度低；

第三类产品：该类产品均属于海尔品牌，其各项评论的占比均很高，可见该类产品的性价比高，受消费者的关注度高；

第四类产品：八种聚类法下，海尔 ES60H-Q(E)电热水器都自属一类，除了性能与外观这两项指标外，其余指标的有效评论占比均很高，可见其在这两项

指标上存在较为明显的不足，需要厂家改进，同时商家在销售时要注意销售策略。

(二) 模型二：熵值法

通过使用“熵值法”对 36 个型号热水器（原有 38 个样品，其中帅康 DSF-50DSQW 电热水器、万家乐 D60-GHF(B)电热水器只有差评，剔除这两条数据）8 个指标分别在本产品评论中的好评情况，算出综合得分，并根据得分项，对 36 个型号热水器进行排序。根据计算出的各项指标的权重，结合实际发掘消费者在国美平台上购买热水器时的关注点，并比较排名结果与实际产品的关注度（该产品的有效评论数/该平台上的有效评论数）的差异，分析消费者评论的影响力。

1. 假设

- 用于计算的数据是剔除水军虚假评论后的有效数据
- 每种型号热水器各个指标下的好评数能够反映消费者对该型号产品该指标的关注度

2. 变量的含义

表 8 变量及其定义

变量	含义
i	产品代码 ($i=1, 2, 3, 4, 5 \dots$)
x_j	第 j 项指标($j=1,2,3,4,5,6,7,8$)
x_{ij}	第 i 种型号产品的第 j 项指标的好评率 ($j=1,2,3,4,5,6,7$)
y_{ij}	第 j 项指标下第 i 年份指标值的比重
e_j	第 j 项指标的信息熵值
d_j	则第 j 项指标的信息效用值：
w_j	第 j 项指标的权重
U_i	第 i 种型号产品的综合得分
g_i	第 i 种产品的关注度

3. 具体实现步骤

首先，对数据的初始处理。由于各种产品的评论数差异较大，为此，我们剔除了没有好评记录的帅康 DSF-50DSQW 电热水器、万家乐 D60-GHF(B)电热水器两条数据，同时将好评数转为产品指标好评率，再进行离差标准化，完成对数据的初始处理。

$$x_{ij}(\text{产品指标好评率}) = \frac{\text{第 } i \text{ 种产品第 } j \text{ 个指标下有效好评数}}{\text{该产品的有效总评论数}} \quad \text{公式 (1)}$$

$$\text{数据离差标准化: } x'_{ij} = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)} \quad \text{公式 (2)}$$

接着，由于进行综合评价时不可避免会碰到一些极端值，如零或异常值，而本次数据标准化前后都有数据为 0 的情况，但是熵值法要求指标数据必须全部大于零，同时为了保证数据的完整性和评价的可信度，不能直接将数据为 0 的信息删去，因此可以通过对出现极端值或者负值的指标数据进行平移消除 0 值，然后再按照熵值法的步骤进行计算，从而对熵值法进行改进。对处理好的

$$\text{数据进行二次处理: } x''_{ij} = c + \frac{x'_{ij} - \min(x'_{ij})}{\max(x'_{ij}) - \min(x'_{ij})} \times d \quad \text{公式 (3)}$$

$$\text{其中, } d = 1 / \sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2}; c = \sum_{i=1}^n x_{ij} / \sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_{ij})^2} \quad \text{公式 (4)}$$

$$\text{数据的权重矩阵: } y_{ij} = \frac{x''_{ij}}{\sum_{i=1}^n x''_{ij}} \quad (\text{when } 0 \leq y_{ij} \leq 1) \quad \text{公式 (5)}$$

然后，利用熵值法对各项指标客观赋权，即先通过计算指标的信息熵，熵越小说明系统越有序，携带的信息越多，信息效用值越大，所赋予的权重越大。根据指标的相对变化程度对系统整体的影响来决定指标的权重，相对变化程度大的指标具有较大的权重。

$$\text{第 } j \text{ 项指标的信息熵值: } e_j = -\frac{1}{\ln n} \sum_{i=1}^n y_{ij} \ln y_{ij} \quad \text{公式 (6)}$$

$$\text{则第 } j \text{ 项指标的信息效用值: } d_j = 1 - e_j \quad \text{公式 (7)}$$

$$\text{则第 } j \text{ 项指标的权重: } w_j = \frac{d_j}{\sum_{i=1}^n d_j} \quad \text{公式 (8)}$$

$$\text{最后采取加权求和算出综合得分: } U_i = \sum_{j=1}^n y_{ij} w_j \quad \text{公式 (9)}$$

4. 通过“熵值法”进行计算的结果

表9 基于产品指标好评率的综合得分值及排名

编号 i	安装 x_1	物流 x_2	性能 x_3	质量 x_4	价格 x_5	外观 x_6	服务 x_7	品牌 x_8	最终结果总分	排名
28	0.1471	0.0294	0.0147	0.7059	0.0735	0.0147	0.0000	0.0000	0.1621	1
9	0.1500	0.1250	0.0750	0.1500	0.2000	0.2750	0.0250	0.0000	0.1598	2
29	0.1316	0.0760	0.0405	0.0911	0.4076	0.1139	0.0481	0.0354	0.1577	3
35	0.1416	0.1030	0.2146	0.2403	0.1931	0.0558	0.0000	0.0258	0.1515	4
22	0.1550	0.1628	0.0155	0.1499	0.2817	0.0853	0.0568	0.0233	0.1512	5
6	0.0644	0.0552	0.0437	0.2736	0.2759	0.1816	0.0391	0.0506	0.1512	6
13	0.0919	0.0595	0.0811	0.3135	0.2378	0.1081	0.0541	0.0108	0.1494	7
30	0.0526	0.1053	0.1579	0.3158	0.2105	0.1053	0.0000	0.0000	0.1480	8
7	0.1127	0.0706	0.1180	0.2718	0.2212	0.0914	0.0535	0.0314	0.1477	9
19	0.1225	0.1359	0.0417	0.2140	0.2692	0.0538	0.0808	0.0242	0.1469	10
2	0.0720	0.1259	0.0448	0.3176	0.2414	0.0494	0.0435	0.0678	0.1443	11
23	0.1325	0.1325	0.0241	0.1416	0.2952	0.0753	0.0843	0.0301	0.1442	12
15	0.0927	0.0944	0.0725	0.2192	0.2985	0.0438	0.1046	0.0270	0.1442	13
21	0.1590	0.1778	0.0272	0.1736	0.2155	0.0669	0.0649	0.0481	0.1437	14
18	0.1016	0.1332	0.0271	0.2054	0.2777	0.0587	0.1174	0.0248	0.1425	15
17	0.0714	0.0789	0.0138	0.3195	0.2456	0.0952	0.0877	0.0288	0.1425	16
16	0.1078	0.1226	0.0335	0.2074	0.2581	0.0762	0.1216	0.0441	0.1424	17
20	0.1073	0.0678	0.0226	0.1977	0.2994	0.0904	0.1299	0.0113	0.1422	18
12	0.0648	0.0602	0.1157	0.2963	0.2037	0.1250	0.0972	0.0278	0.1416	19
14	0.0887	0.1070	0.1651	0.2202	0.1927	0.0795	0.1254	0.0183	0.1388	20
11	0.0865	0.0692	0.1419	0.2976	0.1834	0.0623	0.1142	0.0173	0.1379	21
10	0.1062	0.0619	0.0796	0.2500	0.2013	0.0951	0.1173	0.0199	0.1357	22
8	0.0811	0.0541	0.1486	0.2838	0.1622	0.1081	0.0811	0.0676	0.1355	23
26	0.0833	0.0000	0.0833	0.3333	0.1667	0.1667	0.0000	0.0000	0.1348	24
32	0.0400	0.2800	0.1200	0.1600	0.1600	0.1200	0.0000	0.0400	0.1337	25

4	0.0790	0.1180	0.0419	0.1627	0.2518	0.1066	0.1088	0.1028	0.1334	26
24	0.1129	0.1290	0.0323	0.1452	0.2097	0.1290	0.0000	0.0484	0.1290	27
31	0.0714	0.0714	0.0000	0.4286	0.0714	0.1429	0.0714	0.0714	0.1283	28
1	0.1154	0.1282	0.0833	0.2115	0.1474	0.0577	0.0577	0.1859	0.1268	29
3	0.0828	0.1954	0.0960	0.2252	0.1093	0.0530	0.0464	0.1854	0.1235	30
38	0.0000	0.0000	0.0000	0.4167	0.1667	0.0833	0.0000	0.0000	0.1091	31
27	0.0000	0.0000	0.0000	0.7143	0.0000	0.0000	0.0000	0.0000	0.1090	32
5	0.0737	0.1158	0.0421	0.1684	0.1263	0.0526	0.0526	0.2947	0.1022	33
34	0.0405	0.0135	0.0135	0.4054	0.0405	0.0135	0.4595	0.0135	0.0968	34
37	0.0000	0.0000	0.0000	0.0000	0.3333	0.0000	0.0000	0.0000	0.0678	35
33	0.0000	0.0000	0.0000	0.2105	0.0000	0.0000	0.0000	0.0000	0.0321	36

表 10 计算所得各指标权重

指标	性能 x_3	质量 x_4	价格 x_5	外观 x_6	物流 x_2	安装 x_1	服务 x_7	品牌 x_8
权值	0.1058	0.1527	0.2033	0.1389	0.1352	0.2166	0.0272	0.0203

表 11 根据关注度选出的排名前 5 种型号热水器

产品名称	有效评论数	排名
海尔 ES60H-D5(E)电热水器	3153	1
康宝 CBD40-WA9 电热水器	2450	2
海尔 ES50H-D5(E)电热水器	2390	3
美的 F50-15WA1 电热水器	2088	4
美的 F50-21W6 (B) 电热水器	798	5

5. 结论和建议

(1) 根据表 2 的综合得分值及排名, 排在前 5 的分别是帅康 DSF-50JMF 电热水器、康宝 CBD40-WAF7 电热水器、帅康 DSF-50JMG 电热水器、万家乐 D50-HV12F 电热水器、美的 F60-21W6 (B) 电热水器, 这与根据有效评论数衡量的关注度筛选出的前 5 种产品差别较大, 可见, 产品评论数量的影响力非常大。以

上计算所用数据已经剔除“水帖”，尚且结果差异大，在未剔除水军虚假评论的情况下，可能存在更大的差异，产品评价也将失去其原有参考价值。所以，作为电商平台，要建立完善的识别和剔除水军的系统，保护消费者的利益，才能保证平台的长期发展。

(2) 根据表 3. 各指标权重，消费者对于产品关注点按权重大小依次是安装、价格、质量、物流、外观、性能、服务、品牌。安装的好坏关系到日后的产品使用，安装费用的高低也直接影响消费者的购买成本，所以其所占权重排在第一位；从消费者的消费心理出发，不难发现，价格与质量是其关注的核心点，对应所得结果第二第三位置也是合理的；产品的性能，作为评价产品好坏的核心指标之一，排在第六位，对比原始数据及结合实际，得到的解释是：产品性能的评价体现的是消费者在使用产品达到一段时间后对产品功能的感受，而电商平台上的评论一般是在消费者收到产品或者即刻试用后的直观感受，所以，性能在评论中所占的重要程度比较低；对于潜在消费者而言，在通过评价做购买决策时，应该多关注评价中的产品安装，质量，价格，这三个方面在产品的评价中所含信息量最大，有较高的参考价值。

(3) 根据熵值法确定在不同指标下每个产品的综合得分，看出消费者对产品的关注点有所侧重，根据产品供求平衡理论，厂家需要根据消费者的偏好和需求调整生产。接下来分析当前电商平台的产品定位是否能满足消费者的需求、厂家对于产品的定位需要怎么的转变才能满足消费者。

根据 P·科特勒的产品的三层结构理论，把热水器的产品结构分为核心利益、附加产品、附加服务或利益三层^[14]，依次对应的产品定位分别为高核心价值型、高外在价值型、高服务价值型。然后计算每个产品在三个层次的各自得分，通过比较各层次的得分确定产品定位。

(4) 依据上文的各指标的权重，对每个层次的指标的权重指标进行加总，得到该层次下指标的权重，如表 8 所示。

表 12 各产品层次对应的指标权重

产品层次	核心产品层 A			外在产品层 B		附加产品层 C		
产品定位	高核心价值型			高外在价值型		高服务价值型		
指标	性能 x_3	质量 x_4	价格 x_5	外观 x_6	品牌 x_8	服务 x_7	物流 x_2	安装 x_1
层次权值	0.4617			0.1592		0.379		

接着，再对国美 36 个型号热水器计算每个产品层次下的得分，并判断得出

该产品的定位。

表 13 产品层次定位

编号 i	核心产品层 A	外在产品层 B	附加产品层 C	最大得分	所属层次
1	0.4422	0.2436	0.3013	0.4422	A
2	0.6038	0.1172	0.2414	0.6038	A
3	0.4305	0.2384	0.3246	0.4305	A
4	0.4564	0.2094	0.3058	0.4564	A
5	0.3368	0.3473	0.2421	0.3473	B
6	0.5932	0.2322	0.1587	0.5932	A
7	0.611	0.1228	0.2368	0.611	A
8	0.5946	0.1757	0.2163	0.5946	A
9	0.425	0.275	0.3	0.425	A
10	0.5309	0.115	0.2854	0.5309	A
11	0.6229	0.0796	0.2699	0.6229	A
12	0.6157	0.1528	0.2222	0.6157	A
13	0.6324	0.1189	0.2055	0.6324	A
14	0.578	0.0978	0.3211	0.578	A
15	0.5902	0.0708	0.2917	0.5902	A
16	0.499	0.1203	0.352	0.499	A
17	0.5789	0.124	0.238	0.5789	A
18	0.5102	0.0835	0.3522	0.5102	A
19	0.5249	0.078	0.3392	0.5249	A
20	0.5197	0.1017	0.305	0.5197	A
21	0.4163	0.115	0.4017	0.4163	A
22	0.4471	0.1086	0.3746	0.4471	A
23	0.4609	0.1054	0.3493	0.4609	A
24	0.3872	0.1774	0.2419	0.3872	A
26	0.5833	0.1667	0.0833	0.5833	A
27	0.7143	0	0	0.7143	A
28	0.7941	0.0147	0.1765	0.7941	A

29	0.5392	0.1493	0.2557	0.5392	A
30	0.6842	0.1053	0.1579	0.6842	A
31	0.5	0.2143	0.2142	0.5	A
32	0.44	0.16	0.32	0.44	A
33	0.2105	0	0	0.2105	A
34	0.4594	0.027	0.5135	0.5135	C
35	0.648	0.0816	0.2446	0.648	A
37	0.3333	0	0	0.3333	A
38	0.5834	0.0833	0	0.5834	A

上表结果显示，除了 5 号海尔 ES60H-Q(E)电热水器处在定位在 B 高外在价值产品，34 号万家乐 D50-HG7C 电热水器定位在 C 高服务价值产品外，其它的产品均定位在 A 高核心价值产品。出现这种结果的原因，一方面是产品层次对应的权重差异大，权值分别为 0.4617、0.1592、0.379，其中核心产品层的权数最大。消费最看重产品的核心价值，关注产品的实用性。另一方面，由于网购跨区域，消费者无法售前接触实物，同时，作为耐用品的电热水器价格较其他商品贵，运输过程中容易受损，安装需要专业人员等原因，消费者比较重视热水器的售后服务。

结合“产品三层次理论”，商家应该保障产品的质量和性能，利用新技术不断提高产品质量。同时，重视产品的销售服务。安排由专业的销售人员为客户解答疑问，同时选择荣誉比较好的快递公司，同时及时安排相应的安装人员负责产品的售后安装，进而提高服务质量。虽然产品外在层所占权重低，但是从品牌的推广和产品的营销上来看，应该要提升产品的品牌影响力，在市场上形成良好的品牌效应，这样可以带动同一品牌其它产品的销售。

四、本文的创新点和不足

（一）创新点

在现有的研究中，几乎没有结合文本情感法和实证分析对网购评论类数据进行一个全面的信息挖掘，大部分都停留在理论层面，也极少有完全站在一个“旁观者”的角度，通过评价类文本信息对进入电商中的三个角色——平台运作者、商家、消费者的行为进行分析。而本文通过文本情感分析法从国美平台上 49833 条电热水器评论中提取 8 个评价指标，并结合卡勒特的“产品三层次理论”将 8 个指标归类到产品的三个层次中，通过建立“水军”的识别体系，筛选有效评论，并针对有效评论实证分析，通过熵值法、聚类分析法建立模

型，最后根据实证分析结果，结合被剔除的虚假评论特点，为电商平台、商家、消费者（特别是潜在消费）提供建议。

（二）不足

- 指标中关键词的筛选存在局限性，由于中文含义的多样性，表达意思的含蓄性，无法完全的提取评论中的关键信息。
- 在分析产品的服务、物流、安装时，没有对不同的厂家的数据进行分类处理，仅根据同类产品的评论的数据进行分析，这样不能有针对性地对不同厂家进的服务进行分析。
- 熵值法评价模型缺少与其它模型的对比分析，同时缺少在其它平台热水器产品的实证检验，未能充分考虑对其它电商平台如京东、淘宝天猫的适用性。

五、结论

根据中国电子商务研究中心发布的《100 位 CEO 看 2010 年的中国电子商务》^[15]调查报告显示，在被调查的 100 位 CEO 中，有 66 位 CEO 表示未来 B2C 将代替 C2C 成为网购主流。大部分 CEO 认为，相比 C2C 的个人卖家，B2C 才是专业的，而提供专业的服务才是将来的发展趋势。C2C（淘宝、易趣等）的网购模式要避免走向“黄昏”，就应该提高“平台进入标准”，拒绝低信誉严格管理进入平台的商家，提高平台的信誉，而 B2C 模式下的电商平台，例如国美，要利用自己 28 年的供应链实力，继续保持高服务标准，让客户在体验中形成用户粘性。国美总裁王俊洲曾说：“电商的竞争力之一是价格竞争；二是物流竞争；三是用户体验”。而用户体验是决定电商存亡的关键因素，所以，互联网时代的真正对手是“用户”。

参考文献

- [1]范舒婷.互联网金融中电商平台的研究[J].东方企业文化,2014,20:223.
- [2]方敬尧.基于以电商平台为核心的互联网金融研究[J].经营管理者,2015,03:17-18.
- [3]刘倩,陈曦,于洋.以电商平台为核心的互联网金融市场状况研究[J].经营管理者,2015,10:267.
- [4]张燕.以电商平台为核心的互联网金融探讨[J].现代经济信息,2015,01:345.
- [5]吴兴杰.阿里巴巴的第二次春天在哪里?——2020 年后工业 4.0 时代的“智能制造”将会颠覆电商平台[J].商业文化,2014,31:6-27.
- [6]王洪伟,刘勰,尹裴,廖雅国.Web 文本情感分类研究综述[J].情报学报,2010,(第 5 期).
- [7]赵妍妍,秦兵,刘挺.文本情感分析[J].软件学报,2010,08:1834-1848.
- [8]刘倩.网络文本情感分析综述——从中文和英文的缝隙中看到契机[J].新闻传播,2013,08:19-20+22.
- [9]奚金金,霍欢,徐亚.文本情感分析在网购评论中的应用前景[J].信息技术,2013,12:71-74.
- [10]钟将,杨思源,孙启干.基于文本分类的商品评价情感分析[J].计算机应用,2014,08:2317-2321.
- [11]许敏.C2C 电子商务中虚假评价用户的识别方法研究[D].南京师范大学,2014.
- [12]刘秋文.网络水军的识别和防范[J].新闻前哨,2012,06:37-38.
- [13]任雪松,于秀林著.多元统计分析.北京:中国统计出版社,2011.03.
- [14]曹三省,赵巍.未来新媒体产品的发展趋势——从新媒体产品发展的三层次理论出发[14].新闻与写作,2015,01:5-8.
- [15]中国电子商务研究中心.100 位 CEO 看 2010 年的中国电子商务[N].中国经济导报,2010-01-02B03.
- [16]宁连举,万志超.基于团购商品评论的网络意见领袖识别[J].情报杂志,2013,08:204-207.

附录

(一) SAS 程序

1. 聚类分析

Data pgm33b;

Input x1-x8;

cards;

0.5263	0.3448	0.6667	0.7059	0.6000	0.5294	0.3600	0.7857
0.6727	0.5960	0.7701	0.4779	0.5933	0.4902	0.4658	0.6338
0.7027	0.6458	0.5000	0.4444	0.5333	0.6154	0.2857	0.6842
0.7615	0.7181	0.8333	0.5970	0.7259	0.6813	0.5537	0.6212
0.4615	0.4118	0.0000	0.5714	0.6923	0.2941	0.5556	0.5000
0.6875	0.6250	0.8000	0.5652	0.6364	0.4762	0.4250	0.4545
0.2055	0.1425	0.1833	0.1567	0.1888	0.1664	0.1066	0.1009
0.0261	0.0202	0.0233	0.0294	0.0424	0.0435	0.0171	0.0000
0.0350	0.0365	0.0130	0.0182	0.0548	0.0185	0.0118	0.0000
0.1054	0.0604	0.0444	0.0425	0.1434	0.0791	0.0767	0.0426
0.0353	0.0467	0.0513	0.0479	0.0662	0.0217	0.0370	0.0877
0.0935	0.0857	0.2069	0.1136	0.1833	0.1860	0.1358	0.0000
0.0720	0.0463	0.0678	0.0339	0.1185	0.0661	0.0463	0.0526
0.0674	0.0876	0.0969	0.0464	0.1039	0.0806	0.0879	0.0000
0.0737	0.0820	0.2000	0.0823	0.1523	0.0729	0.0432	0.0067
0.2418	0.2523	0.3358	0.1691	0.3184	0.2533	0.2062	0.0366
0.2604	0.3503	0.3750	0.1163	0.4028	0.3313	0.2454	0.3134
0.1320	0.1123	0.1964	0.0336	0.2277	0.1200	0.0677	0.0380
0.2190	0.2680	0.5088	0.1111	0.3212	0.2367	0.1571	0.0351
0.3538	0.2391	0.5000	0.0870	0.5294	0.2439	0.3871	0.1333
0.1372	0.1598	0.2500	0.1165	0.2500	0.1346	0.1189	0.0182
0.1157	0.1606	0.2432	0.0538	0.1778	0.1569	0.0640	0.1406
0.0808	0.1165	0.2558	0.1020	0.1439	0.1118	0.0585	0.0526
0.1067	0.1475	0.1429	0.0571	0.1515	0.0952	0.1351	0.0000
0.0000	0.0000	0.0000	0.0000	0.0217	0.0182	0.0000	0.0000
0.0000	0.0000	0.0030	0.0038	0.0000	0.0000	0.0000	0.0000

```

0.0000    0.0000  0.0030  0.0000  0.0000  0.0000  0.0000  0.0000
0.0042    0.0030  0.0000  0.0338  0.0027  0.0000  0.0019  0.0000
0.0000    0.0853  0.0539  0.0000  0.2061  0.1095  0.0623  0.0444
0.1203    0.0049  0.0051  0.0000  0.0046  0.0000  0.0066  0.0000
0.0024    0.0026  0.0000  0.0000  0.0000  0.0114  0.0000  0.0000
0.0034    0.0245  0.0234  0.0000  0.0000  0.0265  0.0000  0.0000
0.0078    0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
0.0547    0.0032  0.0263  0.0127  0.0000  0.0049  0.0000  0.0476
0.0030    0.1094  0.2500  0.1630  0.0222  0.0329  0.0917  0.0120
0.0000    0.0000  0.0000  0.0000  0.1280  0.0000  0.0000  0.0215
0.0000    0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
0.0000    0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000

```

```
;
```

```
Proc cluster standard method=single nonorm
```

```
nosquare ccc pseudo out=tree;
```

附注：single（最短距离法），complete（最长距离法），average（类平均法），centroid（重心法），median（中位数法），ward（离差平方和法），flexible（可变类平均法），mcquitty（可变法）

2.判别分析

```
data discdat1;
```

```
input no x1 x2 x3 x4 x5 x6 x7 x8 type;
```

```
label x1='A'
```

```
x2='B'
```

```
x3='C'
```

```
x4='D'
```

```
x5='E'
```

```
x6='F'
```

```
x7='G'
```

```
x8='H';
```

```
cards;
```

```

16 0.2418 0.2523 0.3358 0.1691 0.3184 0.2533 0.2062 0.0366 1
17 0.2604 0.3503 0.3750 0.1163 0.4028 0.3313 0.2454 0.3134 1
20 0.2190 0.2680 0.5088 0.1111 0.3212 0.2367 0.1571 0.0351 1

```

```

21 0.3538 0.2391 0.5000 0.0870 0.5294 0.2439 0.3871 0.1333 1
18 0.1157 0.1606 0.2432 0.0538 0.1778 0.1569 0.0640 0.1406 1
8 0.0261 0.0202 0.0233 0.0294 0.0424 0.0435 0.0171 0.0000 2
9 0.0350 0.0365 0.0130 0.0182 0.0548 0.0185 0.0118 0.0000 2
10 0.1054 0.0604 0.0444 0.0425 0.1434 0.0791 0.0767 0.0426 2
12 0.0353 0.0467 0.0513 0.0479 0.0662 0.0217 0.0370 0.0877 2
14 0.0674 0.0876 0.0969 0.0464 0.1039 0.0806 0.0879 0.0000 2
25 0.0000 0.0000 0.0000 0.0000 0.0217 0.0182 0.0000 0.0000 2
26 0.0000 0.0000 0.0030 0.0038 0.0000 0.0000 0.0000 0.0000 2
27 0.0000 0.0000 0.0030 0.0000 0.0000 0.0000 0.0000 0.0000 2
28 0.0042 0.0030 0.0000 0.0338 0.0027 0.0000 0.0019 0.0000 2
29 0.0000 0.0853 0.0539 0.0000 0.2061 0.1095 0.0623 0.0444 2
30 0.1203 0.0049 0.0051 0.0000 0.0046 0.0000 0.0066 0.0000 2
31 0.0024 0.0026 0.0000 0.0000 0.0000 0.0114 0.0000 0.0000 2
32 0.0034 0.0245 0.0234 0.0000 0.0000 0.0265 0.0000 0.0000 2
33 0.0078 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 2
34 0.0547 0.0032 0.0263 0.0127 0.0000 0.0049 0.0000 0.0476 2
36 0.0000 0.0000 0.0000 0.0000 0.1280 0.0000 0.0000 0.0215 2
37 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 2
38 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 2
6 0.6875 0.6250 0.8000 0.5652 0.6364 0.4762 0.4250 0.4545 3
1 0.5263 0.3448 0.6667 0.7059 0.6000 0.5294 0.3600 0.7857 3
2 0.6727 0.5960 0.7701 0.4779 0.5933 0.4902 0.4658 0.6338 3
3 0.7027 0.6458 0.5000 0.4444 0.5333 0.6154 0.2857 0.6842 3
4 0.7615 0.7181 0.8333 0.5970 0.7259 0.6813 0.5537 0.6212 3
5 0.4615 0.4118 0.0000 0.5714 0.6923 0.2941 0.5556 0.5000 4

```

```
;
```

```
run;
```

```
data discdat2;
```

```
input no x1 x2 x3 x4 x5 x6 x7 x8;
```

```
label x1='A'
```

```
x2='B'
```

```
x3='C'
```

```

x4='D'
x5='E'
x6='F'
x7='G'
x8='H';

cards;
7  0.2055  0.1425  0.1833  0.1567  0.1888  0.1664  0.1066  0.1009
11 0.0935  0.0857  0.2069  0.1136  0.1833  0.1860  0.1358  0.0000
15 0.0737  0.0820  0.2000  0.0823  0.1523  0.0729  0.0432  0.0067
19 0.1320  0.1123  0.1964  0.0336  0.2277  0.1200  0.0677  0.0380
22 0.1372  0.1598  0.2500  0.1165  0.2500  0.1346  0.1189  0.0182
23 0.0808  0.1165  0.2558  0.1020  0.1439  0.1118  0.0585  0.0526
24 0.1067  0.1475  0.1429  0.0571  0.1515  0.0952  0.1351  0.0000
35 0.0030  0.1094  0.2500  0.1630  0.0222  0.0329  0.0917  0.0120
13 0.0720  0.0463  0.0678  0.0339  0.1185  0.0661  0.0463  0.0526
;

run;

proc discrim data=discdatl testdata=discdat2

    crosslisterr testlist;

    class type;

    var x1-x8;

run;

```

（二）程序输出结果——聚类分析表及图

以最短距离法为例，其余见 excel 附表

附表 1 最短距离法聚类过程

聚类数	连接聚类		频数	半偏 R 方	R 方	近似期望 R 方	立方聚条件类	伪 F 统计量	伪 t 方	MinDist	结值
37	OB37	OB38	2	0	1	0	
36	OB27	CL37	3	0	1	.	.	1.80E+05	.	0.012	

35	OB26	CL36	4	0	1	.	.	6.20E+04	6.8	0.0192	
34	CL35	OB33	5	0	1	.	.	2.50E+04	7.1	0.0347	
33	CL34	OB31	6	0	1	.	.	9762	9.2	0.0634	
32	OB25	CL33	7	0	1	.	.	2963	15.4	0.1045	
31	CL32	OB28	8	0.0001	1	.	.	1534	8	0.1552	
30	CL31	OB32	9	0.0001	1	.	.	1037	5.2	0.1648	
29	OB8	OB9	2	0.0001	1	.	.	989	.	0.1869	
28	CL29	CL30	11	0.0006	0.999	.	.	393	17.1	0.2955	
27	OB10	OB13	2	0.0002	0.999	.	.	386	.	0.3012	
26	CL28	OB34	12	0.0003	0.999	.	.	341	3.3	0.3223	
25	CL26	OB30	13	0.0007	0.998	.	.	251	6.6	0.3787	
24	CL27	OB14	3	0.0003	0.998	.	.	243	2.3	0.4033	
23	CL24	OB11	4	0.0008	0.997	.	.	208	3.1	0.4169	
22	OB15	OB23	2	0.0003	0.996	.	.	212	.	0.4306	
21	CL25	OB36	14	0.0009	0.995	.	.	188	5.6	0.4447	
20	CL21	CL23	18	0.0064	0.989	.	.	85.5	24.6	0.4763	
19	CL20	OB24	19	0.0056	0.983	.	.	62.6	9	0.5372	
18	CL19	CL22	21	0.0083	0.975	.	.	46.2	9.5	0.5733	
17	CL18	OB18	22	0.0052	0.97	.	.	42.4	4.2	0.6062	
16	CL17	OB22	23	0.0069	0.963	.	.	38.3	4.8	0.6068	
15	OB12	OB21	2	0.0006	0.962	.	.	42.1	.	0.6188	
14	CL16	OB29	24	0.0022	0.96	.	.	44.6	1.3	0.6226	
13	CL14	CL15	26	0.013	0.947	.	.	37.4	7.8	0.6315	
12	OB7	CL13	27	0.0071	0.94	.	.	37.2	3.4	0.6748	
11	OB16	OB19	2	0.0011	0.939	.	.	41.6	.	0.8251	
10	CL12	OB35	28	0.004	0.935	.	.	44.8	1.7	0.8721	
9	OB2	OB6	2	0.0016	0.933	.	.	50.8	.	0.9856	
8	CL10	CL11	30	0.0416	0.892	.	.	35.3	17.9	1.1828	
7	CL8	OB17	31	0.0353	0.857	0.925	-4.4	30.8	9.6	1.49	

6	CL9	OB4	3	0.0057	0.851	0.913	-3.7	36.5	3.5	1.6226	
5	CL7	OB20	32	0.0429	0.808	0.898	-4.4	34.7	9.1	1.6329	
4	CL6	OB3	4	0.0096	0.798	0.874	-2.4	44.8	2.6	1.7473	
3	OB1	CL4	5	0.0128	0.785	0.827	-1.2	64.1	2.3	2.0937	
2	CL3	OB5	6	0.0365	0.749	0.695	1.31	107	4.9	3.5477	
1	CL2	CL5	38	0.7489	0	0	0	.	107	3.7947	

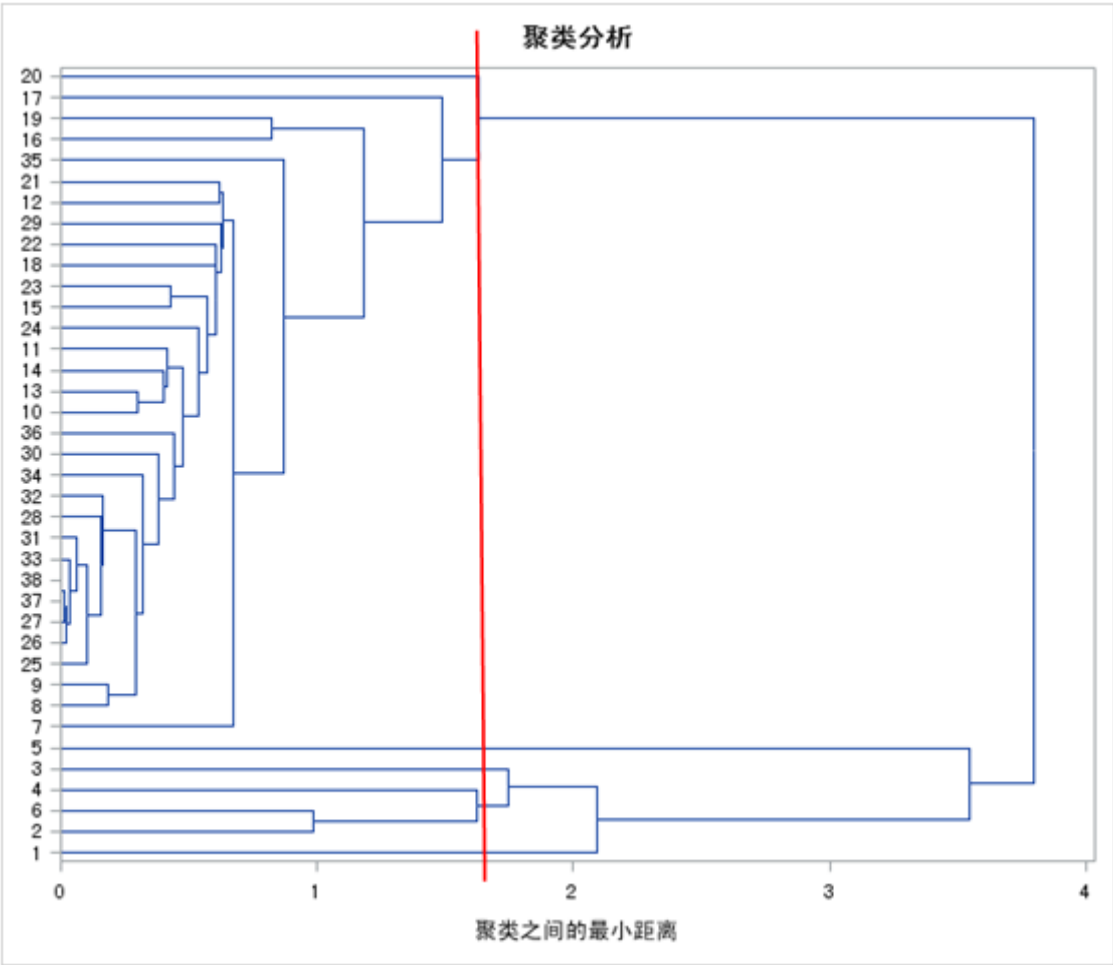


图 1 最短距离法聚类图

（三）数据处理表格

附表 2 国美平台下各产品各个指标下的水军统计

产品	安装	物流	性能	质量	价格	外观	服务	品牌
海尔 ES40H-Q(E)电热水器	18	19	1	5	10	8	16	3
海尔 ES50H-D5(E)电热水器	91	101	20	71	85	104	117	26

海尔 ES50H-Q(E)电热水器	11	17	7	15	14	10	35	6
海尔 ES60H-D5(E)电热水器	83	84	24	54	71	80	108	25
海尔 ES60H-Q(E)电热水器	14	10	2	3	4	12	4	6
海尔 ES80H-D5(E)电热水器	15	15	5	10	16	22	23	6
康宝 CBD40-WA9 电热水器	990	1083	517	479	636	496	813	98
康宝 CBD50-WAF6 电热水器	224	242	84	99	113	110	172	23
康宝 CBD40-WAF7 电热水器	138	132	76	54	69	53	84	14
康宝 CBD50-WA9 电热水器	484	467	237	203	233	233	361	45
康宝 CBD50-WAF1 电热水器	464	429	222	179	254	225	364	52
康宝 CBD50-WAF7 电热水器	126	128	46	39	49	35	70	20
康宝 CBD60-WA9 电热水器	232	247	110	114	119	113	206	18
康宝 CBD60-WAF1 电热水器	360	354	177	144	207	171	280	39
美的 F40-15WA1 电热水器	628	448	56	223	256	305	399	149
美的 F50-15WA1 电热水器	759	560	89	285	334	398	516	184
美的 F50-21W6 (B) 电热水器	196	128	25	76	86	107	123	46
美的 F50-21WB1 电热水器	309	245	45	115	173	176	234	76
美的 F50-30W3 (B) 电热水器	239	183	28	88	112	129	161	55
美的 F50-30W7 电热水器	42	35	4	21	16	31	19	13
美的 F60-15WA1 电热水器	239	184	33	91	120	135	163	54
美的 F60-21W6 (B) 电热水器	237	183	28	88	111	129	161	55
美的 F80-15WA1 电热水器	239	182	32	88	119	135	161	54
美的 F80-30W6 电热水器	67	52	6	33	28	38	32	17
帅康 DSF-50DSQW 电热水器	144	95	9	46	45	54	83	32
帅康 DSF-50DWEL 电热水器	720	676	328	262	366	309	532	73
帅康 DSF-50DYG 电热水器	719	667	333	254	364	306	522	69
帅康 DSF-50JMF 电热水器	718	669	324	257	365	306	527	74

帅康 DSF-50JMG 电热水器	416	397	193	148	208	179	301	43
帅康 DSF-60DYG 电热水器	424	407	197	154	217	181	303	42
帅康 DSF-60JMW 电热水器	419	387	178	148	205	173	319	44
帅康 DSF-60JYG 电热水器	296	279	125	97	134	110	198	33
万家乐 D40-GHF(B)电热水器	380	250	22	123	113	166	236	95
万家乐 D50-HG7C 电热水器	467	308	37	156	147	205	292	120
万家乐 D50-HV12F 电热水器	334	228	54	154	132	206	208	82
万家乐 D60-GHF(B)电热水器	326	223	26	129	109	203	183	91
万家乐 D60-HG7C 电热水器	328	218	29	132	107	207	184	90
万家乐 D60-HV12F 电热水器	418	282	35	162	139	259	224	111
总计	12314	10614	3764	4799	5886	6119	8734	2083

附表3 国美平台下的已有的产品价格、好评度和评论总数

型号	价格	好评度	总评论数
海尔 ES40H-Q(E)电热水器	849.00	100%	193
海尔 ES50H-D5(E)电热水器	1368.00	99%	1,412
海尔 ES50H-Q(E)电热水器	949.00	99%	297
海尔 ES60H-D5(E)电热水器	1599.00	98%	1,631
海尔 ES60H-Q(E)电热水器	1099.00	99%	104
海尔 ES80H-D5(E)电热水器	1699.00	100%	260
康宝 CBD40-WA9 电热水器	599.00	100%	5,104
康宝 CBD50-WAF6 电热水器	699.00	100%	1,034
康宝 CBD40-WAF7 电热水器	799.00	99%	449
康宝 CBD50-WA9 电热水器	699.00	100%	1,967
康宝 CBD50-WAF1 电热水器	749.00	99%	1,665
康宝 CBD50-WAF7 电热水器	799.00	98%	471
康宝 CBD60-WA9 电热水器	799.00	100%	1,042
康宝 CBD60-WAF1 电热水器	849.00	99%	1,367

美的 F40-15WA1 电热水器	699.00	99%	1,862
美的 F50-15WA1 电热水器	799.00	99%	2,876
美的 F50-21W6 (B) 电热水器	999.00	98%	830
美的 F50-21WB1 电热水器	1199.00	99%	1,137
美的 F50-30W3 (B) 电热水器	1199.00	97%	976
美的 F50-30W7 电热水器	1499.00	98%	207
美的 F60-15WA1 电热水器	949.00	98%	902
美的 F60-21W6 (B) 电热水器	1249.00	98%	849
美的 F80-15WA1 电热水器	1149.00	99%	820
美的 F80-30W6 电热水器	2499.00	99%	206
帅康 DSF-50DSQW 电热水器	1799.00	0.00	0.00
帅康 DSF-50DWEL 电热水器	1949.00	99%	2,386
帅康 DSF-50DYG 电热水器	1399.00	0.00	0.00
帅康 DSF-50JMF 电热水器	1099.00	99%	2,365
帅康 DSF-50JMG 电热水器	648.00	99%	1,710
帅康 DSF-60DYG 电热水器	2899.00	99%	1,000
帅康 DSF-60JMW 电热水器	949.00	0.00	0.00
帅康 DSF-60JYG 电热水器	849.00	0.00	0.00
万家乐 D40-GHF(B)电热水器	545.00	0.00	0.00
万家乐 D50-HG7C 电热水器	1598.00	100%	1,161
万家乐 D50-HV12F 电热水器	2798.00	100%	1,105
万家乐 D60-GHF(B)电热水器	763.00	0.00	0.00
万家乐 D60-HG7C 电热水器	1698.00	100%	921
万家乐 D60-HV12F 电热水器	2898.00	0	0

附表 4 有效评论中的好评数

	安装	物流	性能	质量	价格	外观	服务	品牌
海尔 ES40H-Q(E)电热水器	18	20	13	33	23	9	9	29

海尔 ES50H-D5(E)电热水器	172	301	107	759	577	118	104	162
海尔 ES50H-Q(E)电热水器	25	59	29	68	33	16	14	56
海尔 ES60H-D5(E)电热水器	249	372	132	513	794	336	343	324
海尔 ES60H-Q(E)电热水器	7	11	4	16	12	5	5	28
海尔 ES80H-D5(E)电热水器	28	24	19	119	120	79	17	22
康宝 CBD40-WA9 电热水器	276	173	289	666	542	224	131	77
康宝 CBD50-WAF6 电热水器	6	4	11	21	12	8	6	5
康宝 CBD40-WAF7 电热水器	6	5	3	6	8	11	1	0
康宝 CBD50-WA9 电热水器	48	28	36	113	91	43	53	9
康宝 CBD50-WAF1 电热水器	25	20	41	86	53	18	33	5
康宝 CBD50-WAF7 电热水器	14	13	25	64	44	27	21	6
康宝 CBD60-WA9 电热水器	17	11	15	58	44	20	10	2
康宝 CBD60-WAF1 电热水器	29	35	54	72	63	26	41	6
美的 F40-15WA1 电热水器	55	56	43	130	177	26	62	16
美的 F50-15WA1 电热水器	225	256	70	433	539	159	254	92
美的 F50-21W6 (B) 电热水器	57	63	11	255	196	76	70	23
美的 F50-21WB1 电热水器	45	59	12	91	123	26	52	11
美的 F50-30W3 (B) 电热水器	91	101	31	159	200	40	60	18
美的 F50-30W7 电热水器	19	12	4	35	53	16	23	2
美的 F60-15WA1 电热水器	76	85	13	83	103	32	31	23
美的 F60-21W6 (B) 电热水器	60	63	6	58	109	33	22	9
美的 F80-15WA1 电热水器	44	44	8	47	98	25	28	10
美的 F80-30W6 电热水器	7	8	2	9	13	8	0	3
帅康 DSF-50DSQW 电热水器	0	0	0	0	0	0	0	0
帅康 DSF-50DWEL 电热水器	1	0	1	4	2	2	0	0
帅康 DSF-50DYG 电热水器	0	0	0	10	0	0	0	0
帅康 DSF-50JMF 电热水器	10	2	1	48	5	1	0	0
帅康 DSF-50JMG 电热水器	52	30	16	36	161	45	19	14

帅康 DSF-60DYG 电热水器	1	2	3	6	4	2	0	0
帅康 DSF-60JMW 电热水器	1	1	0	6	1	2	1	1

附表 5 有效评论中各产品的差评条数

	安装	物流	性能	质量	价格	外观	服务	品牌
海尔 ES40H-Q(E)电热水器	2	0	0	0	0	0	0	0
海尔 ES50H-D5(E)电热水器	16	63	6	5	0	0	0	0
海尔 ES50H-Q(E)电热水器	1	0	1	0	0	0	0	0
海尔 ES60H-D5(E)电热水器	16	66	7	1	0	0	0	0
海尔 ES60H-Q(E)电热水器	5	1	1	0	0	0	0	0
海尔 ES80H-D5(E)电热水器	5	1	1	0	0	0	0	0
康宝 CBD40-WA9 电热水器	34	25	10	1	2	0	0	0
康宝 CBD50-WAF6 电热水器	0	1	0	0	0	0	0	0
康宝 CBD40-WAF7 电热水器	0	0	0	0	0	0	0	0
康宝 CBD50-WA9 电热水器	25	3	2	0	1	0	0	0
康宝 CBD50-WAF1 电热水器	0	2	0	0	1	0	5	0
康宝 CBD50-WAF7 电热水器	2	0	0	0	0	0	0	0
康宝 CBD60-WA9 电热水器	2	3	2	1	0	0	0	0
康宝 CBD60-WAF1 电热水器	0	1	0	0	0	0	0	0
美的 F40-15WA1 电热水器	16	7	1	0	4	0	0	0
美的 F50-15WA1 电热水器	27	15	10	2	6	0	0	0
美的 F50-21W6 (B) 电热水器	13	17	6	4	0	0	7	0
美的 F50-21WB1 电热水器	9	8	3	2	0	0	2	0
美的 F50-30W3 (B) 电热水器	17	14	4	2	3	0	3	0
美的 F50-30W7 电热水器	5	4	0	0	1	0	3	0
美的 F60-15WA1 电热水器	5	10	2	1	2	0	12	0

美的 F60-21W6 (B) 电热水器	3	12	3	1	1	0	7	0
美的 F80-15WA1 电热水器	4	9	3	1	0	0	11	0
美的 F80-30W6 电热水器	1	1	0	0	0	0	10	0
帅康 DSF-50DSQW 电热水器	0	0	0	0	0	0	3	0
帅康 DSF-50DWEL 电热水器	0	0	1	0	0	0	1	0
帅康 DSF-50DYG 电热水器	0	0	0	0	0	0	4	0
帅康 DSF-50JMF 电热水器	0	0	0	0	0	0	1	0
帅康 DSF-50JMG 电热水器	6	7	3	1	4	0	1	0
帅康 DSF-60DYG 电热水器	0	0	0	0	0	0	1	0
帅康 DSF-60JMW 电热水器	0	0	0	0	0	0	1	0
帅康 DSF-60JYG 电热水器	0	0	0	1	0	0	1	0
万家乐 D40-GHF(B)电热水器	0	0	0	0	0	0	15	0
万家乐 D50-HG7C 电热水器	0	0	0	0	0	0	0	0
万家乐 D50-HV12F 电热水器	2	4	0	0	0	0	0	0
万家乐 D60-GHF(B)电热水器	0	0	0	0	0	0	4	0
万家乐 D60-HG7C 电热水器	1	1	0	0	0	0	0	0
万家乐 D60-HV12F 电热水器	0	0	0	0	0	0	4	0
总计	217	275	66	23	25	0	96	0