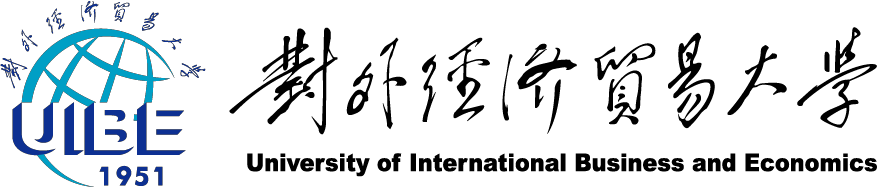
**学校代码：10036**



**博士学位论文**

**B2C 购物网站商品评价的效应研究**

**培养单位：国际经济贸易学院专业名称：产业经济学**

**研究方向：电子商务与网络金融作** 者：孙妮

**指导教师：陈进**

**完成日期：二〇一四年五月**

**Research on Effect of Online Reviews of B2C Shopping Website**

**学位论文原创性声明**

**本人郑重声明：所呈交的学位论文，是本人在导师的指导下， 独立进行研究工作所取得的成果。除文中已经注明引用的内容外， 本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文所涉及的研究工作做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律责任由本人承担。**

**特此声明**

**学位论文作者签名：** 年 月 日

**学位论文版权使用授权书**

**本人完全了解对外经济贸易大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保存学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或部分的阅览服务；学校有权按照有关规定向国家有关部门或者机构送交论文; 学校可以采用影印、缩印或者其它方式合理使用学位论文，或将学位论文的内容编入相关数据库供检索；保密的学位论文在解密后遵守此规定。**

**学位论文作者签名：** 年 月 日

**导师签名：** 年 月 日

摘 要

随着互联网和通信技术的发展与应用的不断深化，以互联网为平台的信息传播凸显出其对经济和社会巨大的影响力。网络购物用户购买商品以后对商品发表的在线评价信息作为在线口碑的一种表现形式，正在深刻改变着潜在网络购物用户的购买等经济行为模式、企业的销售能力和盈利性能以及其市场策略和行业竞争力。尽管我国电子商务环境的发展十分快速，学术界对在线评价信息这种新兴的信息媒体如何反应网络购物用户的网络购物行为规律以及其怎样影响商品的销售量等问题的研究还处于初始阶段。因此，研究在线评价信息对网络购物用户参考已经存在的在线评价信息提高购买决策的效率、电子商务销售企业提高对在线评价信息的管理、提高网络销售企业对网络购物用户行为活动的掌握从而制定更具竞争力的管理策略和营销策略，最终推动电子商务行业实现更大的盈利和健康发展具有现实意义与价值。

本论文以B2C购物网站的网络购物行业为研究背景，以京东商城为研究对象，借助网页信息抓取工具历经一个多月从京东商城抓取了30款手机的213516 条

在线评价信息23482条咨询信息，通过对这些信息进行处理和量化以后形成了不含时间的横截面数据和包含时间因素的面板数据。研究中以行为导向为研究范式，结合定性和定量分析。本研究的主要内容从以下三个方面进行：首先从网络购物用户的咨询、购买及评价行为出发分析了网络购物用户的在线活动的规律；其次通过因子分析找出网络购物用户购买商品之前对购物商场和商品的关注焦点以及购买商品之后发表在线评价时对商品的关注焦点；最后使用回归分析在线评价信息对商品销售数量的影响。

本研究使用的主要方法包括：比较分析法、关联规则挖掘法、因子分析法、最小二乘法、固定效应模型法和广义矩估计法。使用的主要工具分为数据抓取处理工具和数据分析工具两类。数据抓取处理数据为MetaSeeker、谷歌Xml文件合并工具、中科院的分词工具和牛津大学出版社的词频统计工具。数据分析工具为Excel函数、Spss16.0和Stata.

本研究的主要结论如下：

1. 网络购物用户的数量和消费能力与网络购物消费者所在省（市）的经济发展水平有正向相关关系。

2. 每周“星期二”和“星期三”是网络购物用户进行网络购物相关活动的高峰日期，每周“星期六”和“星期日”是网络购物用户进行网络购物相关活动的

低峰日期。

3. 每天的“下午”和“晚上”时间段是网络购物用户进行网络购物相关活动的高峰时间段，“凌晨”时间段是网络购物用户进行网络购物活动的低峰时间段。

4. 网络购物用户咨询与购买行为之间没有必然联系，进行咨询活动的用户不一定购买商品，购买了商品的用户不一定在购买之前进行咨询活动。

5.网络购物用户购买与评价行为之间的日期差与消费者本身的特征和行为习惯有关。

6. 手机购物用户购物之前进行咨询时主要关注商品价格、商城的支付、物流、售后服务、商品性能和产品规格，购物以后进行在线评价活动时主要关注手机的基本功能、是否是国产品牌以及手机信号的稳定性。

7. 京东商城手机销售数量与商家制定的手机销售价格存在显著负相关关系，与在线评价信息的有用数和评价效价存在显著正相关关系。手机销售数量与手机的上市时间长短没有明显关系，与在线咨询数量没有显著关系。

8. 国产品牌手机消费者受口碑效应的影响比非国产品牌手机消费者明显，非国产品牌手机的消费者受从众效应的影响比国产品牌手机消费者明显。

本文立足于网络购物用户群体层面，分析了网络购物用户群体的网络购物行为规律，分析销售数量的影响因素，给网络商城和销售企业提供了消费者的行为规律和销售量的影响因素，并提出了合理的建议，有利于提高网络商城和企业的竞争力，有着比较重要的现实意义。

关键词：网络购物； 在线评价； 购物行为； 在线咨询； 电子商务

**Abstract**

With the development and application of Internet and communication technology, Internet information dissemination convex platform shows its great influence on the economic and society. The online review information published by the online consummers as a form of online word-of-mouth, is profoundly changing the potential of online consummers' buy and other economic behavior, and its ability of sales and profit performance and its market strategies and industry competitiveness. Although the development of China's e-commerce environment is very fast, research on how review information as the newly information media reaction of online shopping behavior of online consummers and how to influence the sales, is still in the initial stage. Therefore, online review information as online consumers' reference already exists to improve the efficiency of their purchasing decision, improve the online enterprise's review information management, improve the network marketing enterprises to grasp the online consumers' activity rules and make more competitive strategy and marketing strategy, and ultimately to promote the electronic commerce industry to achieve greater profitability and healthy development.

In this thesis, online shopping industry of B2C shopping website as the research background, the Jingdong Mall as the research object, by means of Webpage information capture tool after more than a month from Jingdong mall we grabbed 213516 online review information and 23482 consult information of 30 mobile phones, the formation of cross section data does not contain time factor and panel data includes a time factor through after the information is processed and quantized. Take action oriented research paradigm in the research, combining with qualitative and quantitative analysis. The main content of this study from the following three aspects: firstly, evaluation of behavior of consummers' online activity regulations analysis from the online consummers' consulting and review information; secondly, finding out the online consummers' foucus before buying and after buying activity; finally, anylyze the factors that affect the sales with the method of regression analysis.

The main research methods used in this thesis includes: comparative analysis, association rule mining method, factor analysis method, least square method, the fixed effect model and generalized method of moments estimators. The main tools used includes tow kinds: data capture & processing tools and data analysis tools. Data capture

& processing tools includes: MetaSeeker, Google Xml file merge tool, the Chinese Academy of Sciences segmentation tool and University of Oxford press tool. Data analysis tools used in the research includes Excel, Spss16.0 and Stata.

The major findings are as follows:

1. There is positive correlation between the number & consumption ability of online consumers and the level of economic development of the consumers' province (city).

2." Tuesday" and" Wednesday" is the high peak date for the online shopping activities," Saturday" and" Sunday" is the low peak date of the online shopping activities.

3. Every" afternoon" and" Nights" time period is the peak period of online shopping related activities," dawn" time period is the low peak period of online shopping activities.

4. There is no necessary connection between the consultation and purchase behaviors, the users with consulting activity doesn't have to buy goods, users who buy goods are not necessarily to consult before buying.

5. Purchasing behavior and reviewing date difference is related to the habits and behavior of the online consumers.

6. Mobile phone consumers consult for advice before shopping mainly focus on commodity prices, payment, logistics, customer service center service, product performance and product specifications; when review after perchasing they mainly focus on stability of basic function, domestic brands or not and the mobile phone signal.

7. There is a significant negative correlation between the number of mobile phone Jingdong Mall sales and mobile phone prices, there is a significant positive correlation between the number of sales and online reviews and assessment and useful information. No obvious relationship between the mobile phone sales and the length of time of mobile phone published, and no significant relationship with the number of online consulting.

8. Domestic brand mobile phone consumers are more affected by word-of-mouth effect than the none domestic mobile phone brand consumers clearly, the effects of non domestic brand mobile phone consumers are more affected by conformity effect than domestic brand mobile phone consumers.

This paper is based on the online consumer groups, analyzes the online shopping

Behavior of the online consumer groups, analysis of the factors affecting the quantity of sales, provides the factors affecting the behavior of consumers and sales to Internet shopping and sales enterprise, and give the forward reasonable suggestion, help to improve the competitiveness of enterprises, has very important practical significance.

**Keywords: Online Shopping; Online Review; Purchasing Behavior; Electronic Commerce**

目 录

[摘 要](#_Toc686941665) 3

**[Abstract](#_Toc686941666)** 3

[目录](#_Toc686941667) 4

[第](#_Toc686941668)**[1](#_Toc686941668)**[章 绪论](#_Toc686941668) 5

[1.1 前言](#_Toc686941669) 5

[1.1.1 研究背景](#_Toc686941670) 5

[1.1.2 研究目的与意义](#_Toc686941671) 6

[1.2 研究内容](#_Toc686941672) 6

[1.2.1 本文研究分析思路与框架](#_Toc686941673) 6

[1.2.2 论文采用的研究方法](#_Toc686941674) 6

[1.2.3 本文的组织结构](#_Toc686941675) 7

[第2章 ，我国B2C购物网站的概况及在线评价的介绍。](#_Toc686941676) 7

[第3章 ，B2C购物网站商品评价的效应研究基本理论体系与文献综述。（需要修改）](#_Toc686941677) 7

[第4章 ，B2C购物网站用户的商品评价信息的采集与要素量化。](#_Toc686941678) 7

[第5章 B2C网站的在线信息数据的初步分析。](#_Toc686941679) 7

[第6章 B2C网络购物用户商品评价效应的相关实证研究](#_Toc686941680) 7

[第7章 B2C网络购物用户商品评价的效应研究的总结及前景预测](#_Toc686941681) 7

[第 2 章我国](#_Toc686941682)**[B2C](#_Toc686941682)**[购物网站的概况及在线评价的介绍](#_Toc686941682) 7

[2.1 我国B2C购物网站的概况](#_Toc686941683) 7

[2.1.1 电子商务以及B2C的定义](#_Toc686941684) 7

[2.1.2 我国B2C购物网站的经营现状与市场格局](#_Toc686941685) 8

[2.1.3 本文B2C购物网站的研究对象](#_Toc686941686) 8

[2.2 B2C购物网站用户在线商品评价的状况](#_Toc686941687) 9

[2.2.1 在线商品评价的定义](#_Toc686941688) 9

[2.2.2 商品评价的分类与结构](#_Toc686941689) 9

[2.2.3 研究B2C购物网站商品评价的范围及意义](#_Toc686941690) 11

[第3 章](#_Toc686941691) **[B2C](#_Toc686941691)**[购物网站商品评价的效应研究的理论基础与文](#_Toc686941691) 12

[3.1 商品评价效应的基本定义](#_Toc686941692) 12

[3.2 网络购物信息抓取技术的发展及文献综述](#_Toc686941693) 12

[3.2.1 中文网络信息抓取技术的发展历程](#_Toc686941694) 12

[3.2.2 中文网络信息抓取技术的相关文献综述](#_Toc686941695) 12

[3.3 在线口碑效应理论与从众效应理论](#_Toc686941696) 12

[3.3.1 在线口碑理论](#_Toc686941697) 12

[3.3.2 从众效应理论与综述](#_Toc686941698) 13

[3.4 网络购物在线评价信息的效应研究理论及文献综述](#_Toc686941699) 13

[3.2 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价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价需要很长时间来获得消费者的累计“有用性](#_Toc686941700) 13

[3.4.1 购物网站用户商品评价研究的特征选取](#_Toc686941701) 13

[3.4.2 购物网站用户在线商品评价研究的技术和方法](#_Toc686941702) 13

[3.4.3 本论文所用的研究方法与技术](#_Toc686941703) 14

[3.4.4 购物网站用户在线商品评价效应研究的文献综述](#_Toc686941704) 14

[第 4 章](#_Toc686941705) **[B2C](#_Toc686941705)**[购物网站数据信息的采集与要素量化](#_Toc686941705) 14

[4.1 网络信息抓取技术和工具的发展](#_Toc686941706) 14

[4.1.1 常用的网络信息抓取工具](#_Toc686941707) 14

[4.1.2 网络信息抓取工具的选取（MetaSeeker）](#_Toc686941708) 17

[4.2 B2C购物网站在线商品评价信息和咨询信息的抓取](#_Toc686941709) 17

[4.2.1 B2C购物网站在线商品评价信息的结构](#_Toc686941710) 17

[4.2.2 B2C购物网站咨询信息的结构](#_Toc686941711) 17

[4.2.3 B2C购物网站在线商品评价信息和咨询信息的抓取过程](#_Toc686941712) 17

[4.3 B2C购物网站在线商品评价信息和咨询信息的要素量化](#_Toc686941713) 18

[4.3.1 京东商城在线商品评价信息的要素量化](#_Toc686941714) 18

[4.3.2 京东商城咨询信息的要素量化](#_Toc686941715) 19

[4.4 本章小结](#_Toc686941716) 19

[第 5 章](#_Toc686941717) **[B2C](#_Toc686941717)**[购物网站在线信息的初步分析](#_Toc686941717) 24

[5.1 网络购物用户的区域分布、会员等级以及消费品牌的研究](#_Toc686941718) 24

[5.1.1 网络购物用户会员等级以及消费能力省份分布分析](#_Toc686941719) 24

[5.1.2 不同等级网络购物会员用户对手机品牌的购买情况](#_Toc686941720) 37

[5.2 网络购物用户的咨询、购买行为以及对商品进行评论等行为规律研究](#_Toc686941721) 54

[5.2.1 网络购物用户咨询、购买、评价时间特征](#_Toc686941722) 54

[5.2.2 网络购物用户咨询与购买行为关系分析](#_Toc686941723) 63

[5.2.3 网络购物用户购买行为与在线发表评价行为的关系分析](#_Toc686941724) 65

[5.2.4 在线评价信息购买与评价时间差的多元线性回归分析](#_Toc686941725) 68

[5.3 本章小结](#_Toc686941726) 72

[结论：](#_Toc686941727) 72

[第 6 章](#_Toc686941728) **[B2C](#_Toc686941728)**[网络购物用户商品评价效应的相关实证研究](#_Toc686941728) 72

[6.1 网络购物用户在线评价和咨询信息的关注焦点研究](#_Toc686941729) 72

[6.1.1 因子分析理论介绍](#_Toc686941730) 72

[6.1.2 在线评价信息的因子分析](#_Toc686941731) 74

[6.1.3 咨询信息的因子分析](#_Toc686941732) 83

[6.2 在线评价信息对销售量的影响](#_Toc686941733) 97

[6.2.1 基本变量的定义与描述以及相关的基本假设](#_Toc686941734) 97

[6.2.2 在线评价信息对销售量的影响](#_Toc686941735) 98

[6.2.3 本节小结](#_Toc686941736) 106

[6.3 本章小结](#_Toc686941737) 107

[6.25 所示。](#_Toc686941738) 108

[6.26 所示。](#_Toc686941739) 109

[第7 章](#_Toc686941740) **[B2C](#_Toc686941740)**[网络购物用户商品评价的效应研究的总结及前](#_Toc686941740) 111

[7.1 本文的研究结论](#_Toc686941741) 111

[7.2 根据研究结果给网络商家提出合理的建议](#_Toc686941742) 111

[7.3 本文的创新之处](#_Toc686941743) 112

[7.4 本文的不足之处以及进一步研究的方向](#_Toc686941744) 112

[参考文献](#_Toc686941745) 112

[附录 A](#_Toc686941746) 115

[个人简历在读期间发表的学术论文与研究成果](#_Toc686941747) 129

# 第**1**章 绪论

## 1.1 前言

### 1.1.1 研究背景

一．经济背景

据中国互联网络信息中心（CNNIC）第32次中国互联网络发展状况统计报告，

截至2013年6月底，我国网民规模达5.91亿，网络购物用户规模达到2.71亿人，我国网络购物使用率提升至45.9%，比2012年12月底相比半年增长率为11.9%。在网民增长速度逐步放缓的背景下，我国网络购物应用依然呈现迅猛的增长势头，

2012全年用户绝对增长量超出2011年，增长率高出2011年同期4个百分点。根据

工信部在2012年前三季度工业通信业发展情况新闻发布会上提供的数据显示，前三季度电子商务整体市场规模达到了5万亿元，占前三季度整体GDP的14.1%，同比增长13.7%，其中网络零售市场交易规模为7609亿元，增幅达到34.5%。网络经济已经成为中国经济的重要组成部分。在传统经济低迷的同时，网络经济的价值在不断凸显。

据普华永道2013年的调查报告显示，有58%的中国受访者表示每周至少进行网络购物一次，这一比例在他们调查所涉及的所有国家中位居第一。作为对比，美国只有42%的受访者每周至少进行网络购物一次，英国为41%，德国为29%。法国的比例最低，仅为13%。中国消费者使用智能手机和平板电脑进行网络购物的比例也远高于其他国家和地区，超过三分之一的中国网络购物用户使用智能手机和平板电脑等设备，大约是全球平均水平的两倍。普华永道亚太区域零售与消费者事务负责人余凯莉表示，五年前，网络购物在中国几乎还不存在，但如今却已经成为零售商的重要销售渠道，2012年的合并销售额约为1.3万亿元人民币（约合2110亿美元）1。

在中国网民通过网络购物规模和频率都处于领先地位的今天，网络购物用户的问题识别、信息搜集、商品选择、决策购买、购后评价等消费者行为将主要在互联网上完成，而这些商品评价隐含的商业价值对于网络购物用户的决策制定和商家竞争力等影响也越来越大，网络购物商品评价的研究越来越受到重视，并成为学者研究的重要领域。

1 数据来源：[http: //tech. sina. com. cn/i/2013-05-08/14298318014. shtml](http://tech.sina.com.cn/i/2013-05-08/14298318014.shtml)

二．信息化和大数据时代背景

互联网时代，一个大规模生产、分享和应用数据的时代已经开始，大量信息和数据的真实价值就像漂浮在海洋中的冰ft，人们往往只能看到冰ft的一角，而冰ft的大部分都隐藏在表面之下。目前，电子商务、社交网络和移动通信把我们人类社会带入了以“PB”2为单位的结构与非结构数据信息的新时代。在云计算出现之前，传统的计算机无法处理如此大量且不规则的“非结构数据”。以云计算为基础的信息存储、分享和挖掘手段可以便宜、有效的将这些大量、高速、多变化的终端数据存储下来，并随时进行分析与计算。

我国将近6亿人的庞大的网络用户人群和应用市场，复杂性高、充满变化，使得我国成为世界上最复杂的大数据国家。解决由大规模数据和信息引发的问题，探索以大数据为基础的解决方案，是我国产业升级、提高效率的重要手段。数据挖掘不仅能够能为公司竞争力的来源，也将成为国家竞争力的表现之一。目前，我国现代化所面临的各种问题以及国内消费动力不足、教育、交通、医疗保健等各方面挑战，通过大数据这种创新方式来解决问题，创建新的产业群，实现中国制造到中国创造的改变，意义更大。

数据信息将逐渐成为现代社会基础设施的一部分，就像公路、铁路、港口、水电和通信网络一样不可或缺。但是大数据的价值特征而言，它不会像基础设施一样因为人们的使用而折旧和贬值。大数据的出现，使得通过数据分析获得的知识、商机和社会服务能力从以往局限于少数象牙塔之中的学术精英圈子扩大了普通的机构、企业和政府部门，比如我国央行等银行组织近几年开始成立大数据研究中心，研究银行业数据的潜在价值。

大数据发展的障碍和困难在于数据的流动变化性快和可获取性差。美国政府创建了Data. gov网站，为数据和信息研究敞开了大门；英国和印度也有“数据公开”运动。我国要赶上一场数据信息变革，学术界以及商业各界应该首先尝试公开数据、方法和方式。就像工业革命要开放物质交易、流动一样，开放、流通的数据是时代趋势的要求。获取、挖掘、处理和分析数据信息成了各界努力和尝试的新起点，随着思维的转变，数据和信息就可以用来发现现在产品和服务存在的问题，激发新产品和新型服务。

2 信息存储单位，1PB=1024TB，1TB=1024GB。

### 1.1.2 研究目的与意义

宏观方面，我国网络零售规模以万亿计，微观方面，网络购物网民数量2.71亿人这样的庞大的群体，对网络购物用户的行为研究是有现实研究价值的。为了对

2013年上半年的2.71亿并且还在不断增长的网络购物群体网上购买行为进行研究，本文以京东商城上的网购用户为研究对象，通过网络信息抓取工具MetaSeeker对网络购物用户的购买前咨询信息、购买商品后的在线评价信息进行信息挖掘，找出网络购物用户购物前对商品咨询的关注信息、购物后发表评论以及咨询行为与购买行为之间的关系。

现实中，我们周围的同学和朋友都是网上购物的群体，但具体该怎样获取作为网络口碑的商品评论中的有用信息、怎样通过商品评论来影响其他人的购买、怎样有效的发表商品评论等现实问题却没有明确的指导和提示、网络卖家对于网络购物用户购买商品和服务过程产生的巨量的咨询和评论等信息怎样挖掘利用，本文的研究将对这些问题有比较明确的回答。

根据工信部在2012年前三季度工业通信业发展情况新闻发布会上提供的数据显示，2012年前三季度我国网络零售市场交易规模为7609亿元，在传统经济低迷的情况下，网络经济的优势和价值在不断凸显，通过网络购物用户的商品评论对网络用户的购物行为和购买决策等研究对我国经济的发展有巨大帮助和实际应用价值。综上所述，在线购物网站用户的在线咨询和评论研究是一个比较新的领域，尤其是针对中国电子商务领域网络购物用户的中文商品咨询和评价研究具有巨大潜力和现实应用价值。

## 1.2 研究内容

### 1.2.1 本文研究分析思路与框架

本文的分析思路框架建立如下：

首先，综述我国B2C网络购物的基本概况，阐述网络购物在我国经济中的重要地位及对我国经济的重要影响。阐述国内外学者对商品评论的研究方法和研究现状。

其次，由于我国2013年上半年网络购物用户规模已经达到2.42亿人，网络购物使用率提升至42.9%，2012年比2011年增长率为24.8%，2012年前三季度网络

零售市场规模7609亿元规模，在这种背景下，选取我国主流B2C网络购物网站，本文综合对比了京东商城、亚马逊、苏宁、国美和易迅等网站的在线咨询和评论信息，从信息包含用户信息、评论数量和质量等方面综合考虑最终选取京东商城网站为研究对象，利用网络信息抓取工具MetaSeeker获取其13个品牌的30款主流手机的23482条网络购物咨询信息和213516条在线商品评价信息，进行数据收集、处理、筛选。

再次，利用中科院的分词工具和牛津大学出版社的词频统计工具对抓取的网络购物消费者的商品咨询信息和在线评价数据进行量化，得到咨询信息特征和在线评价信息特征的定量数据。

然后，根据对网络购物用户对商品咨询和在线评价信息具有的定量特征进行网络购物用户网络购物行为规律进行分析，利用因子分子方法找到影响用户网络购物买前关注的因素和用户进行商品和服务购买后商品评价的关注因素，根据对网络购物用户咨询、购买和评价行为分析得出的行为规律结果和因子分子分析得到的结果，分析销售量的影响因素。

最后，对分析结论进行总结，根据分析结果给网络商城以及卖家提出合理的有价值的建议。

### 1.2.2 论文采用的研究方法

本文采用定性与定量相结合的研究方法，首先B2C网络购物的经济和技术环境背景，选取主流B2C购网站，利用网络信息提取工具Metaseeker对用户的买前咨询和买后在线评论进行定性和定量的方法，找出影响用户网络购物的因素和用户进行买后商品评价的有效影响因素。具体来说，通过如下几个方法来实现本文的研究：

#### （1）文献法

随着我国网络购物的快速发展，关于网络购物商品评价的研究逐渐出现，本研究是在前人对中文商品评价研究的基础上展开的，通过学校图书馆以及网络电子数据库等多种途径对现有的国内外相关文献资料进行搜集、整理和分析，为论文的理论分析和实证研究打下了坚实基础。通过文献检索、阅读、综合分析发现网络购物在线咨询以及在线商品评价的焦点及难点所在，建立研究的相关视点和焦点，以及此研究领域存在的薄弱环节，选择有利于研究的理论方法，明确深化进一步的研究方向。

在论文研究与写作过程中，对外经济贸易大学丰富的学术资源为本论文的文献研究提供了良好的资源。在整个研究过程中，作者长期追踪国内外的多个电子学术

资源数据库，这些数据库包括图书馆EBSCO Econlit with Full Text经济学数据库、

Elsevier 数据库（Science Direct）、Emerald 全文期刊数据库、EMIS（新兴市场商业资讯）数据库、JSTOR西文过刊数据库以及LexisNexis Academic学术大全数据库、OCLC WorldCat硕博论文数据库、同方知网数据库、万方数字资源系统、维普中文期刊全文数据库等。

#### （2）比较分析法

通过比较可以发现问题，找出规律。不同购物网站的商品评价内容和形式存在很大差异，同一购物网站类似产品的商品评价关注点也可能会存在不同，而且同一购物网站的相同产品的在线评价信息由于网站系统升级等原因也会发生形式变化。例如京东商品和亚马逊购物网站的商品评价形式和特征有所不同，而京东商城同一品牌商品用户购买之前关注点和购买之后进行评论的关注点可能不同，京东商城上不同品牌手机用户的商品评价关注点可能不同，通过对比可以找出其共同点和不同点，由此可以商品评价中重复最多的问题，提出用户关注焦点的关键所在。对比了网络购物用户购物过程中不同阶段不同的关注点，这样能够更好得找出网络购物用户在不同购物阶段做不同决策的影响因素。

#### （3）关联规则挖掘法

关联规则挖掘是数据挖掘和分析的一个重要方法，主要用于发现数据库中项与项之间的相关性和某种联系。而这些相关性或联系在数据中没有直接显示。关联规则挖掘已经被广泛的应用于各个方面，如：超市货架的布置、存货管理、商品的营销策略、银行、电信、移动的数据分析及其保险、医疗等其他行业。面对日益膨胀的海量数据，传统的关联规则挖掘方式已经满足不了人们的需求，因此大数据集的关联规则挖掘研究显得尤为重要。针对上述问题，本文利用Excel中的Vlookup函数从基于抽样的关联规则挖掘方面着手进行京东商城抓取的大数据集的挖掘和分析问题。

#### （4）预测法

通过分析具体的商品在线评价信息，本研究预测某些商品销售影响因素的结果，为我国的B2C网站及商家对网络购物用户的销售方法提供可参考的方案及决策，并预测了未来我国网络购物用户购物习惯可能的发展方向。

#### （5）因子分析法

通过因子分子法研究在线评价信息和咨询信息包含的众多变量之间的内部依赖关系，探求在线评价和咨询信息数据中的基本结构，并用少数几个独立的不可观测的假想变量来表示在线评价和咨询信息的基本数据结构。这几个不可观测的假想变

量能够反映原来众多变量的主要信息，达到分析网络购物用户在线评价以及咨询时关注的焦点。

#### （6）实证分析法

通过对采集到的京东商城的用户买前咨询信息和买后对商品的在线评价信息的进行采集、筛选、量化，然后通过信息挖掘、横截面数据模型、面板数据模型回归分析，得出对销售数量影响的定量结果。

### 1.2.3 本文的组织结构

本文共分七章进行阐述，文章的具体研究内容如下：第1章，绪论。

首先介绍本文的研究背景、研究目的以及意义，然后阐述本文的主要研究思路和研究框架，最后介绍本文的研究思路和研究方法以及组织结构。

# 第2章 ，我国B2C购物网站的概况及在线评价的介绍。

本章阐述了研究的电子商务环境背景，首先阐述了B2C购物网站的定义，分析我国B2C主流购物网站的经营现状和市场格局，然后介绍不同购物网站的咨询和在线商品评价信息的分类与结构，介绍购物网站咨询和在线商品评价的研究范围和意义。

# 第3章 ，B2C购物网站商品评价的效应研究基本理论体系与文献综述。（需要修改）

本章介绍了在线商品评价的基本定义，信息抓取技术的发展及文献综述，研究在线商品评价与在线口碑理论体系的具体内容和关系，以及网络购物网站商品评价与从众效应和行为决策理论。

# 第4章 ，B2C购物网站用户的商品评价信息的采集与要素量化。

本章详细介绍了论文研究数据的来源，以及使用Metaseeker等网页信息抓取工具抓取京东商城在线咨询和在线评价信息的过程。介绍了研究所获取数据的结构，绍了买前咨询信息化和买后在线评价信息的要素构成，以及采用分词和词频统计工具对要素进行量化的过程和技术。

# 第5章 B2C网站的在线信息数据的初步分析。

在第4章获取的信息数据进行处理和量化的基础上，利用Excel工具自带的分析函数对数据进行初步分析，分析不同省份、不同等级网络购物用户的网络行为规律，以及不同等级网络购物用户对不同品牌手机的消费情况，分析了网络购物用户咨询、购买与评价行为之间的关系。

# 第6章 B2C网络购物用户商品评价效应的相关实证研究

本章首先利用因子分子方法对在线评价信息和在线咨询信息进行分析，找出网络购物用户购物之前进行咨询活动时关注焦点和购物以后评价时的关注焦点。然后，利用横截面数据模型和面板数据模型通过最小二乘（OLS）、固定效应模型（Fixed Effect Model）和广义矩估计（GMM）对评价信息进行了回归分析，找出影响销售数量的影响因素。

# 第7章 B2C网络购物用户商品评价的效应研究的总结及前景预测

本章对论文的分析结果进行总结，根据研究结果给网络商城卖家提出合理的建议。然后总结本文的创新点，找出不足与进一步研究需要做的工作，总结进一步研究的方向，给未来的相关学术研究给出合理建议与预期。

本文的研究结构图如下：

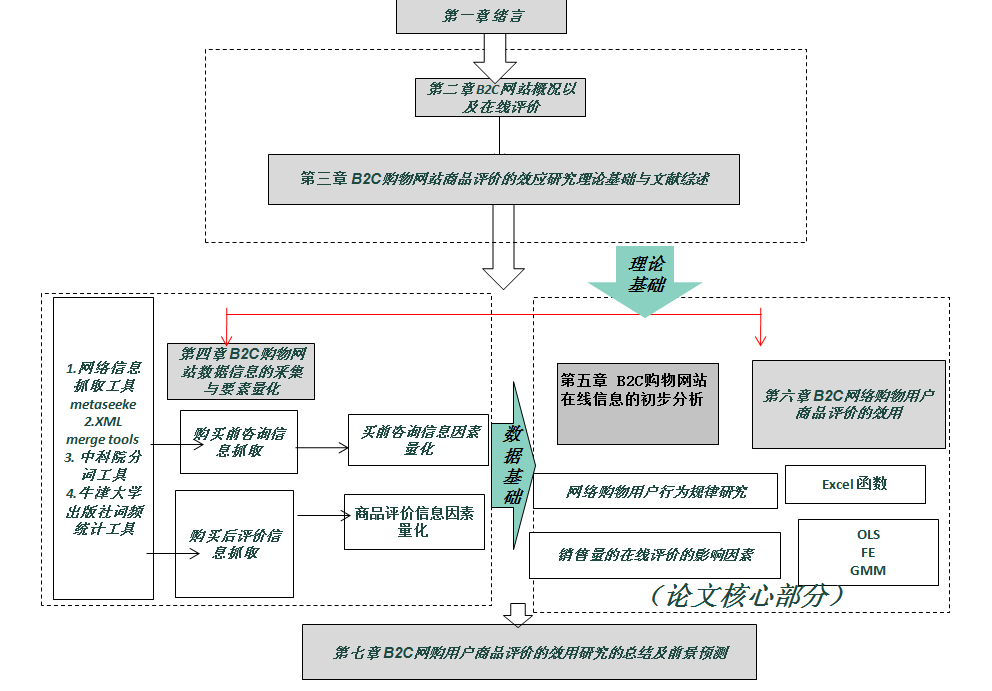


图 1.1 本文研究结构图

# 第 2 章我国**B2C**购物网站的概况及在线评价的介绍

## 2.1 我国B2C购物网站的概况

### 2.1.1 电子商务以及B2C的定义

1.基本定义

电子商务是一个新型学科，也是一个新行业，目前正日新月异地发展。电子商务至今没有统一的定义，这也是电子商务概念很容易引起混乱的原因之一。国内外不同的书籍、机构等对于电子商务的定义都有差异，电子商务研究者从不同角度给出了众多电子商务的定义。

1997年11月6日-7日，国际商会在法国首都巴黎举行了世界电子商务会议（The World Business Agenda for Electronic Commerce）。全世界商业、信息技术、法律等领域的专家和政府部门的代表，共同探讨了电子商务的概念问题。这是迄今为止电子商务比较有权威的概念阐述：电子商务（Electronic Commerce），是指实现整个贸易活动的电子化。从涵盖范围方面可以定义为：交易各方以电子交易方式而不是通过当面交换或直接面谈方式进行的任何形式的商业交易；从技术方面可以定义为：电子商务是一种多技术的集合体，包括交换数据（如电子数据交换、电子邮件）、获得数据（如共享数据库、电子公告牌）以及自动捕获数据（如条形码）等。

艾瑞咨询网给出的广义的电子商务的定义是：电子商务源于Electronic business的翻译，主要是指通过电子的手段来进行商业的活动，具体就是通过网络使企业，供应商，客户和用户之间实现企业与企业，企业与用户的业务流程的电子化，从而提高企业的生成、库存、流通和资金等各个环节的效率。

由埃弗瑞姆.特伯恩（Efrain Turban）和戴维.金（David King）著作的《电子商务导论》（《Introduction of Electronic Business》）一书中从六个视角给出了电子商务的不同概念：（1）通信的观点：电子商务是借助电话、计算机网络或任何其他电子媒介进行信息、产品或服务传递以及支付的过程。（2）业务流程观点：电子商务是将技术应用于与营销活动相关的过程和工作流以实现自动化的过程。（3）服务的观点：电子商务是传达公司、消费者和管理层的需求，从而降低服务费用、提高产品质量和服务速度的工具。（4）在线的观点：电子商务提供了通过internet购买和销售产品、信息的能力，并提供了其他在线服务的可能。（5）合作的观点：电子商务是在组织间和组织内部进行合作的框架。（6）社区的观点：电子商务为社区成员提

供了一个学习、集会和合作的集会场所。然后再给出的电子商务的统一定义是“通过计算机网络进行买入、卖出和交换商品、服务和信息的过程。”本文更赞成此书中对电子商务的定义。

由埃弗瑞姆.特伯恩（Efrain Turban）和戴维.金（David King）著作的《电子商务导论》（《Introduction of Electronic Business》）一书把电子商务的主体分为政府

（Government）、企业（Business）和个人（Client）三种类型，根据交易双方的主体不同，他们把电子商务分为B2B、B2C、C2C，B2G，C2G等电子商务模式，其中

B代表企业，C代表个人，G代表政府。B2C模式是指企业对个人模式，是指企业把商品、服务和信息卖给个人。本文的研究范畴就是我国的B2C电子商务模式。

2.我国B2B、C2C和B2C电子商务模式的基本概况

#### （1）B2B模式方面

根据中国电子商务研究中心的数据，如图2.1所示，我国B2B商业模式主要包括在线上外贸服务为主综合B2B模式、在线下内贸服务为主的综合B2B模式、以行业门户+联盟为主的综合B2B模式、以小宗外贸服务为主的综合B2B模式、以供求商机信息服务为主的行业B2B模式、以行业资讯服务为主的行业B2B模式、以招商加盟为主的行业B2B模式、以在线交易服务为主的行业B2B模式、以技术社区服务为主的行业B2B模式和以项目外包服务为主的行业B2B模式十大模式。



图2.1十大主流B2B商业模式及典型企业，来源自中国电子商务研究中心根据中国电子商务研究中心的数据，截止到2013年6月，我国B2B电子商务

服务企业达11400家，同比增长4%，2013年上半年中国B2B电子商务服务商的营

收规模为93.8亿元，同比增长25.9%。关于B2B电子商务重要构成部分的大宗商品交易市场方面，至2012年，我国有各类大宗商品电子交易市场396家，年交易额超

过12万亿元，目前处于营业状态的有357家，处于无交易或停业状态的有39家。

#### （2）C2C模式方面

淘宝、易趣和拍拍网目前是中国最有影响力的三大C2C平台。百度本来在2008年进军电子商务领域也建立了C2C 电子商务平台有啊（youa. com），但是由于发展等各方面的不顺利，以及支付工具百付宝没有获得牌照等各种原因，最终在2011

年3月31日宣布关闭。根据中国电子商务研究中心相对旧一些的数据来看，如图

2.2所示，C2C服务商的主要情况来看，盈利情况并不乐观。由于假货泛滥，管理成本高等各方面原因C2C电子商务的发展并不顺利，尤其是百度有啊的关闭被视为

C2C模式的衰败。



图 2.2 中国C2C市场主要服务商，来源自中国电子商务研究中心

（1）B2C模式方面

与C2C模式发展相比，我国B2C电子商务的发展较为顺利。根据经营范围分类B2C商城主要分为百货类、家电类、旅游类、服装类、母婴类和钻石类，各分类具体代表网络商城如图2.3所示。



图 2.3 中国B2C市场主要分类，来源自中国电子商务研究中心

据中国电子商务研究中心监测数据显示，截至2012年12月底中国B2C网络零售市场（包括平台式与自主销售式）上，排名第一的是天猫商城，占52.1%份额；京东商城名列第二，占据了22.3%份额；排名第三位的是苏宁易购开始后来居上，达到了3.6%份额。后续第4-10位排名依次为：腾讯B2C(3.3%)、凡客诚品(2.7%)、亚马逊中国(2.3%)、库巴网(1.4%)、当当网(1.2%)、易迅网(0.6%)、新蛋中国(0.3%)。

### 2.1.2 我国B2C购物网站的经营现状与市场格局

根据艾瑞咨询发布的2013年第一季度中国网络购物市场数据，2013年第一季度中国网络购物市场交易规模为3520.8亿元，较去年同期增长36.6%；从网络购物市场结构来看，B2C占比超过30%，与去年同期相比提高了10个百分点；从网购市场份额来看，天猫占比51.3%，继续领跑B2C市场，京东商城则以43.4%的份额稳居自主销售为主B2C市场首位。

从B2C市场整体来看，如图2.4市场份额位居前五位的分别是天猫、京东商城、腾讯电商（含易迅和QQ网购）、苏宁易购和当当网。其中，天猫以51.3%的占比稳居行业之首，领先优势难以超越；经过平台整合后的腾讯电商，尤其是易迅网发展较快，市场份额亦有所提升，Q1占比达6.8%. B2C市场份额集中效应进一步加剧，大型B2C电商企业的平台优势越来越明显，TOP10企业合计占比达91.1%，与2012年同期相比提升了近3个百分点。



图 2.4 市场份额分布图（数据来源：艾瑞咨询）

从图2.5自主销售为主B2C市场来看，排名居榜首的仍是京东商城，占比为

43.4%, Q1交易额同比增长100.8%；位居次席的苏宁易购占比11.2%，其一季度的同比增速亦超过100%。位于B2C领域第二集团的重要参与者还包括当当、易迅、唯品会、亚马逊中国及一号店（含一号商城）等，如上企业市场份额均超过4%，且占比均有所提升。



图 2.5 自主销售为主B2C市场份额（数据来源：艾瑞咨询）

根据艾瑞咨询iEcTracker的监测数据如图2.6显示，2013年1-3月B2C平台数码相机单品交易额分布中，京东商城稳居行业首位，连续三个月占据B2C显示数码相机交易50%左右的销量份额；而从市场整体来看，淘宝系（天猫+淘宝集市）与京东商城合计占到相机网络零售渠道市场份额的80%。



图 2.6 数码相机B2C网络零售市场销量占比（数据来源：艾瑞咨询）

### 2.1.3 本文B2C购物网站的研究对象

由于本文选取的研究对象是用手机为样本，手机样本数据归类为家电类，所以选取我国最大的家电网络商城——京东商城。京东商城自2004年1月涉足电子商务领域，目前是我国B2C市场最大的3C3网络购物专业平台。

本研究共选取了京东商城销售的十三个品牌的三十个型号的手机进行研究，这十三个手机品牌三十个型号的选取是在华为终端有限公司的三个手机方面工程师的指导下参考京东商城上的手机销售情况进行选取的，这些手机是当前现实社会中使用比较多，比较流行的经典手机款式。

## 2.2 B2C购物网站用户在线商品评价的状况

### 2.2.1 在线商品评价的定义

目前对在线商品评价的相关研究很多，也有研究把商品评价称为商品评论，但

3 3C是计算机（Computer）、通讯（Communication）和消费电子产品（ConsumerElectronic）三类电子产品的简称。

是没有学者对商品评价给出统一或权威的定义。

当当网在其网站上把评价分为服务评论和商品评论两种评论，并分别进行了定义：服务评论是“对订单，价格，送货及包装，服务等非商品本身的评论”。商品评论是指“购买过此商品的顾客针对该商品的使用过程发表的使用感受”。当当网认为在选购商品时，网络消费者可以看看其他用户发表的评论，这些评论对其它网络消费者具有一定的参考价值。京东商城把评价分为“服务评分”和“商品评价”，并进行了定义：服务评分“用来对本次购物过程中的京东服务进行评价”，服务评价内容包含“商品描述相符”、“卖家服务态度”、“物流发货速度”和“配送人员态度”四个分项，最好的给5星，最差的给1星。京东商城的商品评价定义为“用来对所购

商品的质量进行评价“，网络消费者可以针对被评价的商品给星，最好的给5星，最

差的给1星。百度百科上给出的商品评价的定义是，商品评价是指“生产厂家、商家或者消费者根据具体商品的性能、规格、材质、使用寿命、外观等商品的内在价值设定一个可量化或定性的评价体系，由消费者对商品使用价值进行评价的过程。”

综上所述，本文把研究对象在线商品评价定义为“网络购物用户通过网络购买商品后，针对所购买的商品服务和使用过程的感受给出的综合评价的过程。”本文的研究对象是商品评价，不包括针对网络商城平台的服务评价。

### 2.2.2 商品评价的分类与结构

本文的研究数据是抓取自京东商城的商品评价，这些商品评价是针对13个品牌

的30款经典主流手机的23482条网络购物咨询信息和213516条在线商品评价信息，为什么选择京东商城的商品评价为研究对象？

由于2013 年第一季度自主销售B2C 网络销售份额中，京东商城的销售份额

43.4%远远超过苏宁易购、当当、亚马逊等其它B2C网站，京东商城的网络购物用户数量很大，满足学术研究取样样本容量的要求，所以对京东商城用户的研究更具有典型代表性。

从我们能够获取的公开的商品评价信息的内容来看，京东商城上的用户商品评价信息相对完整。如图2.7所示，我们能够获取到用户等级、用户名称、用户省份、购买日期、评价日期、买前咨询信息、用户评分、商品购买和使用心得、评价有用数量、评价回复数量、优点、不足等信息，如表2.1所示，而苏宁、当当网、易迅网、亚马逊中国、国美在线等B2C购物网站的商品评价信息却只有用户名称、评价内容、评价日期、是否购买过所评价产品几项内容，对于数据挖掘获取更准确信息来说只有京东商城的商品评价信息包含信息内容更丰富所以更适合做学术研究。而

且，京东商城上只有购买过商品的用户才能对所购商品进行评价，用户评价以后可以获取积分，积分又可以最终折抵用购物现金，所以京东商城上的用户只要购买过相关产品的用户就有动力对相关商品进行评价，用户购物评价与所购买的商品的相关度很高，所以最终用户评价数量可以大致用来估算商品销售数量，这对于难以获取商品销售数量的学者来说是非常有利的。



图 2.7 京东商城用户评价截图

表 2.1 京东商城、苏宁、当当网等网站商品评价信息内容总结

|  | 用户  等级 | 用户  名称 | 用户  省份 | 购买  日期 | 评价  日期 | 用户  评分 | 使用  心得 | 评价有  用数量 | 评价回  复数量 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 京东商城 | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 苏宁易  购 |  | √ |  | √ | √ | √ | √ | √ | √ |
| 当当网 | √ | √ |  |  | √ | √ | √ | √ | √ |
| 易迅网 | √ | √ |  |  | √ | √ | √ |  |  |
| 亚马逊  中国 |  | √ |  |  | √ | √ | √ | √ | √ |
| 国美  在线 | √ | √ |  |  | √ | √ | √ |  |  |

### 2.2.3 研究B2C购物网站商品评价的范围及意义

本文根据获取到的商品评价信息，对不同省份、不同等级用户的消费品牌进行了分析，根据京东商城的会员等级规则进一步分析了不同省份用户的消费能力，网络购物用户的咨询、购买以及买后的商品评价等行为进行分析，得出网络购物用户的各种网络行为规律。通过对购买前的咨询信息和买后评价信息，可以分析出用户

进行网络购物各个过程的关注焦点。

B2C购物网站用户的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度。作为一种用户反馈机制，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电子网络口碑，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服务。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。

# **第3 章** **B2C**购物网站商品评价的效应研究的理论基础与文

**献综述**

本章主要分析论文研究相关的理论基础和文献综述。首先阐述了商品在线评价效应的基本定义，并介绍在线信息的抓取技术。然后介绍与在线评价信息相关的口碑效应理论与从众效应理论，最后介绍在线评价信息效应研究的相关理论和文献综述。

## 3.1 商品评价效应的基本定义

由于本研究主要做了网络购物在线评价相关信息的研究，所以应该先介绍网络购物用户的购物过程，然后阐述在线商品评价的效应定义。如图3.1所示，用户的网络购物过程主要经过用户注册、查找商品、对商品进行咨询、将商品放入购物车、提交订单、查看订单状态、收货、评价等网络行为。网络购物用户的每个网络行为会产生相应的信息。我们通过工具对这些信息进行抓取并进行分析和研究。



图 3.1 网络购物用户购物流程图4

效应（Effect），是指在有限环境下，一些因素和一些结果构成的一种因果现象，多用于对一种自然现象和社会现象的描述。效应一词的使用泛围比较广，并不一定是严格的科学定理、定律中的因果关系。例子如温室效应、蝴蝶效应、毛毛虫效应、

4 图片来源：根据京东商城网络的购物提示过程以及作者在上多年的购物经验总结得出这一购物流程图。

从众效应、音叉效应、木桶效应、完形崩溃效应等等。

经济学中的效用（Utility）与效应（Effect）不同，经济学中的效用是指某种被研究的商品或者行为满足人的欲望的能力评价，或者说效用是指消费者消费某种商品时感受到的满足程度。效用是分析经济学上分析消费者行为的基础，按照对其度量，又可分为基数效用理论和序数效用理论两种理论，两者区别在于效应是否可以加总。

在线商品评价的效应目前在科学研究领域还没有明确定义，根据研究的内容本文给出了在线商品评价效应的定义。在线商品评价的效应是指在电子商务大环境中，进行网络购物的消费者在网络购物过程中产生的在线评论所具有的信息和经济价值。比如，这些在线评价信息反映出网络购物用户怎样的网络行为规律？这些在线评价信息对商品销售量有没有影响？有什么样的影响？本文针对这几个问题分别进行了研究。

## 3.2 网络购物信息抓取技术的发展及文献综述

### 3.2.1 中文网络信息抓取技术的发展历程

为了对网络购物用户购物过程中产生的信息进行研究，应该先批量获取或抓取这些信息。由于电子商务发展时间相对来说不是很长，所以对于网络在线信息的研究相对比较新，抓取网络信息的相关技术发展时间不是很长，但是由于商业行业发展研究以及科学研究需要等原因发展比较快。

目前抓取网络信息的主要技术是网络爬虫技术，网络爬虫又被称为网页蜘蛛或者网络机器人，是一种按照一定的规则，自动的抓取网络信息的程序或者脚本。网络爬虫是相对普通搜索引擎而来的，通用搜索引擎如百度和Google等，作为辅助人们检索信息的工具成为用户访问网络的入口和指南。但是，这些通用搜索引擎存在着一定的局限性，比如只能基于关键字的检索，难以支持根据语义信息提出的搜索。因此，定向抓取相关网页资源的主题网络爬虫( focused crawler or topical crawler)应运而生（Yu 、Sheikoleslami& ZhangA, 2002）。主题网络爬虫是一个自动下载网页的程序，它根据既定的抓取目标，有选择的访问万维网上的网页与相关的链接，获取所需要的信息。与通用搜索引擎不同，主题网络爬虫并不追求大的覆盖，而将目标定为抓取与某一特定主题内容相关的网页，为面向主题的用户查询准备数据资源。

### 3.2.2 中文网络信息抓取技术的相关文献综述

在抓取技术研究方面，罗敖生（2002）提出利用ASP技术调用ActiveX控件实现网页信息的自动下载，但是只能抓取某一个URL网页中的信息，如果要针对其他的URL网页，只能借助ASP开发工具或者“记事本”工具才能修改抓取程序。

朱晓静、方磊坤和李太君（2007）提出了基于jxta协议的检索算法并采用P2P方式的信息检索的网络信息抓取系统。顾潇华和郭军城（2007）使用Delphi集成开发环境提供的TIdhttp组件抓取网页文本，并利用正则表达式和Mshtml组件从中提取超链信息保存到指定数据库，并能对网页文本进行自动的统计分类。最近几年，对于主题网络爬虫的研究相对较多。张博和蔡皖东（2009）采用向量空间模型来计算网页的主题相关度，使用改进的Shark- Search 网页搜索策略来决定待抓取链接的访问次序，提高了主题网页采集的精度。袁浩和黄烟波（2009）针对主题爬虫所采用的基于关键词和基于语义概念的两种基本抓取网页的方式效率比较低的问题提出了通过网页标题分析对主题爬虫的改进方案。白鹤，汤迪斌和王劲林（2009）提出采用基于分类标注的多主题策略，解决了同一爬虫系统内多主题自适应兼容的问题，并且提出基于Trie树的URL存储策略的改进方法，能够高效地支持URL查询、插入和重复性检测。

以上是科研学者从理论方面对网络信息抓取技术的研究，这些研究在现实应用中被没有进行商业化或者公开使用。目前进行科学研究的很多通用工具都是企业提供的网络抓取工具，比如深圳市天据电子商务有限公司GooSeeker推出的免费网络信息抓取工具Metaseeker；以众包的形式使用大家的力量，为高校学生和科研人员抓取所需要的互联网数据的中国爬萌等工具，具体工具的特点在第4章中会详细介绍。

## 3.3 在线口碑效应理论与从众效应理论

### 3.3.1 在线口碑理论

在线评价信息作为网络购物用户获取参考信息决定购买策略的重要来源，是在线口碑的一种表现形式，不同于传统的口碑效应（Word-of-Mouth），传统口碑通常采用口头口口相传的传播方式，口碑信息往往不容易量化，只能把传统口头口碑做定性分析。在线评价信息以文本形式呈现，可以长久保存，尤其是目前的网络购物

网站还提供了评价者对商品的评价分数、评价者的购买时间、购买的商品的特征、消费者的特征等各种信息，这些特征为准确测量和研究在线评价信息的效应提供了数据基础。

许多学者对传统口碑给出了明确的定义，Emanuel Rosen等学者（2000）认为，口碑是关于品牌的所有的评述，是关于某个特定产品、服务或公司里的人的所有的口头交流的总和。Blackwell等人（2001）认为除了营销人员之外的人与人之间非正式的传送思想、评价、意见形成口碑。国内学者也对口碑概念进行了定义。郭会斌

（2005）认为，口碑指的是关于某一组织的信用和可信赖度、经营方法、服务等方面的信息，它可以通过社会公众口口相传，从一个顾客传播到另一顾客，形成的对某一产品或服务长期的、统一的、稳定的看法。学者阙克儒（2004）认为口碑就是顾客愿意主动且积极的，以非商业的目的，散布自己或他人对于产品的使用经验。黄英和朱顺德（2003）认为，口碑是由生产者以外的个人，通过明示或暗示的方式，不经过第三方处理加工，传递关于某一特定或某一种类的产品、品牌、服务、厂商、销售者，以及能够使人联想到上述对象的任何信息，从而导致受众获得信息、改变态度、甚至影响购买行为的一种双向互动的传播行为。

把传统口碑的应用环境换成网络环境便成了在线口碑(Online Word-of-Mouth)，在线口碑也称为网络口碑或者电子口碑（Electronic Word-of-Mouth），指的是在网络传播的口碑信息。在线口碑的表现形式很多，比如网络商城（Online Shopping Mall）、虚拟社区（Virtual Community）、讨论组（Disscussion Group）等网络空间中的口碑。随着电子商务的发展，在线口碑的定义也有变化，定义越来越符合目前电子商务经济的环境。在电子商务发展初期，Christiansen和Tax（2000）把在线口碑定义为，网络用户（口碑发送者）在网络上张贴文章的行为，等同于人与人之间的口碑沟通行为，只是信息呈现形式由声音（auditory）转换成书写（wiritten）的形式。到2003年前后电子商务环境逐渐成熟，Dellarocas（2003）认为在线口碑就是在线的客户反馈系统，又称声誉系统，即利用网络的双向沟通特征，个人就公司、产品、服务或某件事在网络上分享自己的经验和意见。Henning-Thurau等人（2004）认为潜在、目前或者先前的消费者对产品或者企业做出的大众可以通过互联网获得的正向或者负向的评述构成在线口碑。Datta等人（2005）认为大量潜在的、目前或者以前的消费者就使用经验和情境进行在线的沟通形成在线口碑。本文对于在线口碑定义采用哈尔滨工业大学郝媛媛（2010）博士论文中的定义，把在线口碑定义为大众可以通过互联网获得的，潜在、实际或者先前的消费者主要以文本形式做出的，对商品、服务、企业等特质的评定或对产品或服务使用体验的表述，可以是正面的也可以是

负面的。

在线口碑作为传统口碑的一种新形势的发展，内容仍然是对商品、服务、产品和服务供应商特征、评价等信息描述形式，是不以商业推广为目的的非正式的沟通，产生的目的与传统口碑相同，就是消费者通过网络对产品或服务使用体验的表述。在线口碑与传统口碑具有很多共同点，章晶晶（2007）的两个总结是：（1）在线口碑和传统口碑都具有较高的可信度。（2）在线和线下口碑都具有很强的交互性。不管是在线的还是传统口头传播的口碑，都来源于独立的第三方，传播者可能是某产品领域的专家或个人消费者，他们能客观地描述有关产品的事实情况，因此在受众心目中是比商业广告更为可靠的信息源。尤其是网络给全世界的消费者提供了信息交换和共享平台，消费者可以通过网络搜索到许多来自不同地区且具有不同背景的消费者对于同一产品的评论，然后综合多个使用者的独立意见可以得到比较公正、无偏的评价。

在线评价信息作为在线口碑的表现形式，在网络上传播，发布者和接受者都是网络购物消费者，与销售企业的经济利益没有直接关系，所以这种在线评价信息比其它营销渠道得到的广告宣传信息更客观真实可靠。通过在线评价信息这种在线口碑形式能够潜意识地影响其他网络购物用户对商品的认知，有效的增加销售数量，从而增加企业的竞争力。这就是在线评价信息与在线口碑理论之间的关系。

在线口碑能够潜在影响网络购物用户的消费决策，从而影响消费者的购买行为并最终影响某种商品的销售数量。Katz & Lazarsfeld（1964）研究发现口碑效应促使消费者转换消费品牌的效果比期刊广告、销售人员推销、和电台广告的效果好很多。Bristor学者（1990）通过研究也得到了类似结论，认为口碑效应对消费者对消费品牌的转换影响力是广告影响力的2倍。国内也有学者对口碑效应对商品销售的影响进行了研究。卢向华和冯越（2009）通过大众点评网对餐馆点评的实证分析研究发现在线口碑的数量、评分以及负面点评率对产品的销售收入有显著的影响。

### 3.3.2 从众效应理论与综述

从众效应（Bandwagon Influence）是普遍存在的一种社会现象，消费者的从众行为是购买行为中非常重要的一种。Leibenstein（1950）给出从众效应的定义，他认为从众效应是一种纯粹的心理效应，即当顾客与众人的行为一致时，自我感觉会比较好。Lascu和Zinkhan（1999）认为消费者的从众行为是指消费者在接受到他人的产品评价、购买意愿或购买行为的信息后，改变了自己的产品评价、购买意愿或购买行为，努力与他人保持一致。

消费者受从众效应的影响主要有以下特征：（1）消费者从众行为的本质是一种模仿行为。Banerjee（1992）认为人们往往喜欢模仿别人，跟随别人做其他人所做的事。在现实生活中，潜在消费者周围好朋友或者同事做出的一些消费行为成为示范性的消费行为，潜在消费者受周围人的消费行为影响去做同样的购买行为，比如网络上报道的有学生为了模仿同学购买苹果手机的行为但由于受经济条件限制而做出违法的事情。当某种消费行为被认可，评价较高时，其他消费者就会效仿，也去购买同一品牌的同一产品，从而引发了一种流行趋势，这就是消费者的从众行为。（2）消费者的从众行为与不确定性和有限理性有关。通常情况下，由于信息不对称，消费者不能获得做出购买决策所需要的所欲信息，消费者是有限理性的，消费者为了降低由于信息不足导致的不确定性就容易发生从众行为。

在线评价信息对网络购物用户形成一种信息影响，由于本文研究网站的对象京东商城只有购买了商品才能对商品做出一次在线评价，所以在线评价数量的多少某种程度上反映了销售数量的多少，当消费者看到某种产品的在线评价信息很多，某一种商品的在线评价信息甚至多达几万条时就知道至少已经有几万人购买了该商品，就很容易受从众效应影响去购买相应的产品。本文根据抓取到的在线评价信息对国产品牌手机和非国产品牌手机销售数量的影响进行了研究。

## 3.4 网络购物在线评价信息的效应研究理论及文献综述

不论买家还是卖家，作为在线商品评价的读者都希望从中获得有助于制定决策的信息。然而，针对商品的在线商品评价往往数量众多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价上，进而迅速制定购买决策。许多B2C电子商务网站，如京东、亚马逊、当当，提供在线商品评论的效应评价功能（如图

## 3.2 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价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价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，在这种人工效应评价机制下，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。在线商品评价的效应分析是指应用数据挖掘的方法研究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包含的有价值的信息要素，研究这些评价信息反映的网络消费者

行为规律，以及对商品销售数量的影响。



图 3.2 京东商城华为D2手机评论截图

在线商品评论往往具有较强的主观性，消费者需要揣测每条评论中的观点是否合理，权衡哪些观点对自己的购买决策有用。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（如图3.3所示）。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），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（Zhang Z, 2008）。例如，图3.3京东商城上某件茶具商品的平均得分是4.1分，只能表明购买该件商品的所有评论者对这件购买商品服务的平均评分是积极的，而不能反映消费者对网上购买这件商品所涉及享受服务的具体意见，到底积极性体现在哪些具体细节潜在消费者都不能直接观察。因此，B2C电子商务网站所呈现的某件商品的平均得分可能有误导性。



图 3.3 京东商城某品牌茶具商家服务综合评分4.1 分

为了解决当前B2C电子商务购物网站对在线商品评论效应评价的局限性，国外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（2011）等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（2007）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应。Otterbacher（2009）却认为评论的内容质量(message quality)是评价在线商品评论效应的重要方面。可以看出，以往研究对于影响在线商品评论效应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，采集不同网站的数据集，尤其是不同国度的消费者的评价语言和习惯不同，用不同的算法评价指标会得出不同的结论。

### 3.4.1 购物网站用户商品评价研究的特征选取

国外学者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、情感分析和评价效用三方面。主要有以下原因：首先，在线商品评论的内容富含观点与情感相关的内容（Ye Q., Zhang Z. Q. & Law R, 2009）；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（Zhang Z.,2008）；再次，购物网站上用户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（Abbasi A, Chen H C, Salem A, 2008）。而且研究文本主要是英语文本，对于中国商品评价的研究需要进行中文文本分析。

### 3.4.2 购物网站用户在线商品评价研究的技术和方法

在线商品评价的研究技术可以分为基于机器学习和基于相似度得分等方法。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(Abbasi A, Chen H C, Thoms S, et a1,2008)，评论的效用(Zhang Z.,2008; Liu Y, Huang X J, An A J, et a1.,2008; Kim S M, Pantel P, Chklovski T, et a1.,2006)。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（2007）利用支持向量机（SVM）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(3-1)计算，

*F* (*x*)*w*T *x* *b*

（3-1）,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（2011） 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(3-2)表示，

*U*(*x*)*F*( *f*1,*f*2,... *fi*,... *f p* )

（3-2），其中

*f*1 (*x*),... *fi* (*x*),... *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∈[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(Liu Y, Huang X J, An A J, et a1.,2008)。

### 3.4.3 本论文所用的研究方法与技术

本论文中使用的技术和方法主要是描述性分析、因子分析以及回归分析。

描述性分析是把在线评价信息和咨询信息处理和要素量化以后，在第5章中根据需要利用Excel的相关形式转换函数，转换成需要的格式，然后利用统计函数进

行占比分析、关联分析，通过分析得出网络购物用户的网络购物行为习惯。

对消费者进行网络购物之前对商品的咨询信息的关注焦点以及购买商品之后对商城和所购商品的使用体验做出的在线评价信息的分析使用因子分析的方法。因子分析是指在分析处理多变量问题时，由于众变量之间往往存在一定的相关性，使得观测数据所反映的信息存在重叠现象。因此为了尽量避免信息重叠和减轻工作量，人们就往往希望能找出少数几个互不相关的综合变量来尽可能地反映原来数据所含有的绝大部分信息，因子分析正是为解决此类问题而产生的多元统计分析方法。

研究中使用了最小二乘法回归（OLS），固定效应模型回归和广义矩估计（GMM）回归。对总体样本数据取对数后转换成线性关系，然后利用最小二乘法对总体样本横截面数据进行回归分析，找出价格、咨询数量、评价效价、上市时间以及评价有用数等因素对销售数量的影响。然后考虑时间因素，把数据分期形成面板数据，使用固定效应模型进行研究，最后考虑到前一期销售数量对本期销售数量的影响使用广义矩估计回归进行研究，得到口碑效应和从众效应对国产品牌手机和非国产手机品牌销售数量的影响。

### 3.4.4 购物网站用户在线商品评价效应研究的文献综述

最近几年，关于商品评论信息的效用研究主要从商品评价的有用性、对购买决策的影响、与商品销量之间的关系、这些商品评价的可信度以及顾客的满意度五大方面来研究。

1.商品评价的有用性相关研究

学者殷国鹏（2012）从社会性因素的影响效用研究什么样的在线评论更有用，Soo-Min Kim等（2006）研究用机器学习的方法采用SVM回归方式自动预测在线评论的有用性，姜巍，张莉，戴翼，蒋竞和王刚（2013）从面向用户需求方面利用评论间的语义关联提出基于复杂网络的评论有用性的分析方法，杨爽，徐畅（2013）从虚拟社区角度对在线评论的信息质量和评论者的社区地位方面来判断评论有用性。郝媛媛等（2009）以影评数据的在线评论为研究对象，结合文本挖掘技术和实证研究方法，从文本特征出发探索影响在线评论有用性的因素。

2. 在线评论对购买决策的影响相关研究

Yuanyuan Hao等（2010）从正面评论和负面评论对搜索型商品与体验型商品的不同影响进行了研究，发现正面评论对搜索型商品的影响比体验型商品的影响大，而负面评论对搜索型商品与体验型商品的影响没有明显差异。Do-Hyung Park和Sara Kim（2008）认为在线评论的认知契约对购物老手的购物决策影响比新手大，而评

论数量对新手的购买决策影响比老手的大。Jumin Lee等（2008）对负面评论对消费者的研究发现卷入度高的网购消费者关注负面评论的质量，而卷入度低的消费者更关注负面评论的角度。国内学者李宏等（2011）从负面在线评论的质量、消费者的卷入程度以及性别三者对消费者网络购买决策的影响，发现负面在线评论质量的高低对消费者对产品的满意度以及购买选择具有显著影响。

3. 在线评论的对顾客的满意度影响的相关研究

消费者满意度研究主要开始于20世纪70年代的美国和日本。美国AT&T公司为了在市场竞争中处于优势地位而开始研究顾客对企业所提供服务的满意情况。日本丰田汽车公司通过研究顾客满意度来掌握外界对企业自身的认识情况。90年代中期我国开始应用顾客满意度调查，从而获得顾客信息、竞争优势和员工评价等。对于网络购物的满意度调查主要是从2000年以后网络通信和电子商务快速发展以后才开始的。

Xia Liu等学者（2008）尝试从在线购物体验的整体流程找出那些因素会影响中国网络消费者的满意度，通过问卷调查方式研究发现信息质量（information quality）、购物网站设计(website design)、商品属性(merchandise attributes)、交易能力(transaction capability)、隐私安全性（security/privacy）、支付（payment）、物流（delivery）和客户服务（customer service）八个方面对网络购物用户的满意度有很大影响。我国学者潘勇和赵军民（2008）从顾客满意度的视角对我国B2C电子商务网站进行评价，构建了以顾客满意为基础的评价指标体系，采用层次分析法与灰色评价相结合的方法，避免了层次分析法在生成各项评价指标权重时主观性太强的特点，采用了问卷调查的方式获取了126份有效基础数据进行实证数据研究。李思曼，王宇航和李亚平（2008）在技术接受模型、交易成本理论和顾客满意度理论的基础上，提出了基于顾客满意度消费者网上购物决策模型，并通过结构方程模型方法对所建模型进行实证研究，结果证明顾客满意度、感知风险、感知信任、感知利得和感知易用性共同影响着消费者网上购买意愿，感知利得和感知易用性同时还通过影响顾客满意度来影响消费者购买行为，感知信任和感知风险之间也存在着影响关系。这两篇中文文章的不足之处是只从外部因素分析影响消费者网上购物行为，但是没有考虑到内部因素，即网络消费者个人的特征。不同的消费群体具有不同的消费体验，比如，有的消费者是风险偏好型，这时感知风险这一变量可能对他的影响就不明显，而不同地区的消费者消费习惯可能也有不同。

4.在线评论的可信度研究

孟美任等（2013）通过问卷调查，选取内容完整性、情感平衡性、评论时效性

以及发布者身份明确性４类特征，采用CRFs模型进行评论可信度４级分类，并进行特征组合实验，得到判断在线评论的最佳特征组合，准确率能达到75%。李念武等（2009）也通过问卷调查和主成分因子分析法，发现正面在线评论的可信度影响网络用户对商家的信任，负面在线评论的可信度对网络用户对商家的信任和购买意向都有显著影响。施国良等（2011）通过文本挖掘的理论，提出不同购物网站商品评论对比分析的方法，对不同购物网站同一商品评论是否一致进行研究，发现不同购物网站对同一商品的评论并不完全一致，这种不一致主要体现在商品特征方面。

5. 在线评论与商品销量关系的研究

近几年对于在线评论与销量之间关系的研究相对比较多，Wenjing Duan 等

（2008）发现在线评论的正面口碑效应对电影票房有影响，Garrett Sonnier等（2011）进一步研究发现正面评论、中性评论和负面评论对电影票房的影响有着显著影响。Judith Chevalier 等（2006）通过提高在线评论的正面口碑效应对Amazon. com 和

Barnesandnoble. com网站的图书销售量有提升作用，而且一星评论的影响比5星评论的影响大。Qiang Ye等（2009）通过与Judith Chevalier（2006）相同的方法研究发现在线评论对中国旅馆的网络订购数量有显著关系。

在文献综述与深度分析基础上，本文认为目前已有研究大多集中在在线评论的内容特征，对评论者的视角因素关注不够，几乎没有考虑网络购物用户购买意愿发起时的咨询信息，也没有学者对网络购物用户的行为规律进行深入挖掘研究。已有研究的数据要么是采取问卷调查获取，主观性太大，要么就是几百顶多数千条在线评论，数据和样本量严重不足，容易产生较大偏差。针对这些不足，本文在大数据时代背景中以大数据思想为基础以网络购物用户购物过程中所产生的海量真实的咨询和评论信息为基础，详细阐述了数据处理过程，并对数据进行挖掘，分析出网络购物用户的网络购物行为规律以及在线评价信息与咨询信息对销售数量的影响。

# **第 4 章** **B2C**购物网站数据信息的采集与要素量化

## 4.1 网络信息抓取技术和工具的发展

在第三章中提到过，目前网络信息抓取的主要技术是网络爬虫技术，又被称为网页蜘蛛或者网络机器人，是一种按照一定的规则，自动的抓取网络信息的程序或者脚本。网络爬虫是相对普通搜索引擎而来的，由于通用搜索引擎存在着一定的局限性，所以定向抓取相关网页资源的主题网络爬虫( focused crawler or topical crawler)才应运而生，并且在网络爬虫的基础上很多公司创造了很多面向网络用户个体或者企业的集成工具，我们能够直接使用这些网页信息抓取工具进行网络信息的抓取。

### 4.1.1 常用的网络信息抓取工具

Web网络是一个巨大的资源宝库，目前页面数目已超过800亿5，每小时还以惊人的速度增长，Web网页里面有大量有价值的信息，例如潜在客户的列表与联系信息，竞争产品的价格列表，实时金融新闻，供求信息，论文摘要等等。但是要获取特定网页中的特定信息就需要专门的网页信息抓取工具。常用的网页信息抓取工具有Metaseeker、中国爬萌、乐思等工具，这些信息抓取工具各有特色。

1．网页信息抓取工具Metaseeker

MetaSeeker是由深圳市天据电子商务有限公司GooSeeker研发的一个Web网页抓取/数据抽取/页面信息提取工具包，能够按照用户的指令从Web页面上筛选出用户需要的信息，并输出含有语义结构的提取结果文件（XML文件）。这个工具的用户有两种个人用户和企业用户，针对个人用户的工具是免费的。

Web 页面显示的信息是供网络用户阅读的，对于机器来说，是无结构的，

MetaSeeker解决了一个关键问题就是将无结构的Web页面信息转换成有结构的适于机器阅读和处理的信息，这种服务可以应用于专业搜索、Mashup服务和Web数据挖掘领域，所以MetaSeeker的适用范围比较广泛，几乎可以抓取Web网页上的所有信息。

MetaSeeker工具包主要是供垂直搜索（或称为专业搜索）服务、信息汇聚和门户服务、Mashup服务、企业网信息汇聚、商业情报采集、论坛或者博客迁移、智能

5数据来源于[http: //www. knowlesys. cn/product\_webdataminer\_index. html](http://www.knowlesys.cn/product_webdataminer_index.html)。

信息代理、个人信息检索和信息挖掘的企业和个人使用。常用的案例就是抓取当当网、京东商城网站的商品价格信息等、以及批量定时抓取旅游网站的机票价格等信息。

MetaSeeker网络抓取工具的工具包主要有MetaStudio、DataScraper、MetaCamp和Datastore四个组件组成，我们主要使用MetaStudio和DataScraper两个组件来设置网页信息抓取规则和进行网页信息的抓取。

#### （1）MetaStudio：是Web页面信息结构描述工具包，使用MetaStudio自动生成Web网页信息提取规则文件，供网页信息提取工具包DataScraper使用。MetaStudio提供友好的GUI（图形用户界面）界面，作为Firefox浏览器扩展（Firefox extension）发行，不同的操作系统平台和Firefox版本有不同的发行包。Gooseeker公司推荐与

MetaCamp和DataStore配套使用，这样信息结构描述文件和各种信息提取指令文件就可以上载到MetaCamp和DataStore服务器，以拥有协同描述页面信息结构和分享信息提取成果的能力。

#### （2）DataScraper：是从Web 页面进行信息提取(网页抓取/抽取)工具，利用

MetaStudio生成的各种信息提取指令文件，对特定页面的信息进行连续提取，并将信息存储在DataStore服务器中。提供GUI（图形用户界面）界面，作为火狐浏览器Firefox扩展发行，技术核心是一个GooSeeker公司自己研发的工作流引擎，由信息提取工作流指令文件驱动。

使用时值得注意的是DataScraper是Web页面信息提取工具，内含基于工作流框架的定题网络爬虫（也叫聚焦网络爬虫），主要提取网页文字、页面超链接URL、

HTML网页元素的属性值等等，并不直接下载图片、视频或者其它文件，而是提取文件的URL地址。通过DataScraper提取的网页信息内容根据信息结构描述格式化成XML文件，方便集成到各种网络应用中。DataScraper作为Firefox扩展（Firefox

extension）发行，V4.10.0版本以前，DataScraper可以跨平台部署，所有操作系统和

Firefox都可以部署同一个datascraper. xpi；到V4.10.0版本，DataScraper分成高性能版和普通版，高性能版不能跨平台部署，不同的操作系统平台和Firefox版本有不同的发行包，但普通版仍然能够跨平台部署。

#### （3）MetaCamp：是存储和管理信息结构描述文件的服务器。作为一个应用

（application）部署在Tomcat等Servlet容器中。

#### （4）DataStore：是存储和管理信息提取线索、各种信息提取指令文件和信息提取结果文件的服务器，集成Lucene v2.3.2技术，能够为结果文件建立索引。作为一个应用（application）部署在Tomcat等Servlet容器中。

2．中国爬萌

中国爬萌以众包的形式使用大家的力量，为高校学生和科研人员抓取所需要的互联网数据。目前中国爬萌的抓取对象是以微博数据为主，包括博主基本信息、关注关系、微博消息数据，每一份资源都需要5个积分来下载，所以最好能以固定群体共同使用一个爬萌账号来积攒积分便于获取信息。

使用爬萌工具抓取微博信息需要信息抓取者首先要有一个微博账号，用自己的账号和密码登陆进微博以后才能进行微博信息抓取。由于每家微博服务供应商对自己的微博用户都有一定的隐私保护措施，当意识到自己的用户微博信息被网络机器人大量复制以后，微博服务供应商开始采取反网络机器人措施以阻止微博信息被无限制的复制下载，所以使用爬萌抓取微博信息目前能获取到的信息量也有一定限制，只能采取众包方式才能获得足够多的微博信息。爬萌获取微博的信息内容如表4.1所示，由于我们不是针对微博信息进行研究，具体使用方式就不再赘述。

表 4.1 爬萌抓取微博信息的信息内容

| 字段名 | 完整名称 | 描述 |
| --- | --- | --- |
| \_id | 消息原始 ID |  |
| mt | Message content | 微博内容 |
| ui | User id | 用户的原始 ID |
| un | User name | 用户名 |
| sn | Screen name | 用户屏幕名 |
| iu | Image url | 用户头像 URL |
| ri | Retweet msg id | 转发消息 ID |
| mu | Msg url | 微博消息 URL |
| srn | Msg source name | 微博消息来源 |
| pu | Msg picture url | 微博消息含图片 URL |
| au | Msg audio url | 微博消息含音频 URL |
| vu | Msg video url | 微博消息含视频 URL |
| rc | Retweet count | 转发数 |
| cc | Comment count | 评论数 |
| ct | Crawler time | 采集时间，整型时间戳 |
| nc | Name card | 消息中@到的用户 |

3．乐思网页信息抓取工具

乐思网络数据抽取系统是深圳乐思软件研发的，专门用来批量抽取新闻，论坛，网页等的各种网络数据，直接保存到数据库或发布到网络的一种信息化系统工具。它可以根据用户设定的规则自动采集原网页，获取格式网页中需要的内容，现在发展成也可以对数据进行处理的工具/系统。

乐思网络信息采集系统的主要功能是根据用户自定义的任务配置，批量而精确地抽取因特网目标网页中的半结构化与非结构化数据，然后把这些数据转化为结构化的记录，保存到本地数据库中，用于内部使用或外网发布，快速实现外部信息的获取。乐思网络信息采集系统主要用于舆情监测，品牌监测，价格监测，门户网站新闻采集，行业资讯采集，竞争情报获取，商业数据整合，市场研究，数据库营销等领域。

乐思网络信息采集系统分为标准版、专业版和企业版三个版本，但是都是需要购买才能使用。

### 4.1.2 网络信息抓取工具的选取（MetaSeeker）

由于研究的网站对象是京东商城上的咨询和在线评价信息，这些信息是公开的任何网络用户在任何时候都可以随时浏览的信息，考虑到购买软件成本等各方面因素，由于MetaSeeker在线版可以免费下载和使用，并且其源代码可以阅读也不需要用户自己编码，使用比较方便比较简单，所以本研究选择了免费的网页信息抓取工具——MetaSeeker。

## 4.2 B2C购物网站在线商品评价信息和咨询信息的抓取

### 4.2.1 B2C购物网站在线商品评价信息的结构

通过京东商城网站与当当网、苏宁易购网、易迅网、亚马逊和国美在线网的商品评价对比以后，本文选取京东商城的在线商品评价信息为主要研究内容。由于购物网站的更新和升级等各种因素，在线评价的结构和外在表现形式并不是一成不变。以京东商城上iPhone4手机的在线评价为例，2013年5月3日以及此日期之前，京东商城的在线评价包括“用户名称”、“用户等级”、“省份”、“购买日期”、“评价日期和时间”、“有用数”、“回复数”、“星级”、“（商品）优点”、“（商品）不足”和“（使用）心得”，如图4.1所示2013年5月3日开始，京东商城在线评价的内容形式进行了第一次明显改进升级，把“（商品）优点”、“（商品）不足”和“（使用）心得”

三项改成“（商品）颜色”、“商品型号”和“（使用）心得”三项。如图4.2所示2013

年7月14日开始，京东商城在线评价的内容形式进行了第二次改进升级，就是在第一次改进的基础上添加了商品评价“标签”项。



图 4.1 京东商城在线评价第一次内容形式变化



图 4.2 京东商城在线评价第二次内容形式变化

京东商城怎样鼓励用户进行合理的评论？如图4.3所示，京东商城为了鼓励网络购物用户对商品进行有效评论，对评论者进行积分奖励，京东商城的规定是只有购买了某种商品才能对该商品进行评价，而且只能评价一次，有效评价奖励对应的积分，积分可以兑换成购物券用来抵现金，兑换的购物券为电子券，有效期为一年，自成功兑换之日起计算。京东购物券仅限兑换券的用户使用，不能折算为现金、也不能再次兑换为积分。10元京券需要铜牌以上会员级别兑换，20元京券需要银牌以上会员级别兑换，50元京券需要金牌以上会员级别兑换，100元京券需要钻石以上会员级别兑换。每日至多换取三张优惠券。积分可直接换购电子书，积分和人民币比例为10: 1。用户如果用积分兑换电子书，需全额用积分支付，不支持与其他支付方式混合使用。从我个人作为京东商城用户获取的积分经验来看，我每次进行在线

评价至少获取10个积分，所以，根据京东商城积分兑换现金或者购物券的比例，我个人作为网络购物者进行一次在线评论每次可以“获利”1.0元以上。从价值角度判断，京东商城一条有效的在线评论的价值不低于1.0元。从利润最大化角度来看，网络购物者有动力进行合理的评论。

需要注意的是，在线评价信息还有一种特殊形式，即“晒单”形式，这种形式除了包括“用户名称”、“用户等级”、“省份”、“购买日期”、“评价日期和时间”、“有用数”、“回复数”、“星级”、“（商品）优点”、“（商品）不足”和“（使用）心得”等特征外还增加了评价者自己拍摄的商品的图片，这种“晒单”形式其实就是在普通评价基础上增加了网络购物用户所购买的商品图片的展示。



图 4.3 京东商城在线评价换取积分

京东商城为什么要对这些在线评价的形式内容进行多次改进升级呢？从在线评价的形式来看，第一次升级前，京东商城的在线评价包括（商品）优点、（商品）不足和（使用）心得，但并没有对评论者给出明确的评价定义。通过仔细的研究，我们发现很多评价的（商品）优点和（使用）心得内容可能是重复的，或者（商品）不足和（使用）心得的内容是重复的，评论者没有那么大的耐心仔细去琢磨（商品）优点、（商品）不足和（使用）心得，由于同一个评价者的评价重复的内容，对于京东商城的数据库是一种资源浪费，对评价者也是一种时间浪费，所以，京东商城在

2013年5月3日把线商品评价的“（商品）优点”、“（商品）不足”和“（使用）心得”三项改成“（商品）颜色”、“商品型号”和“（使用）心得”三项。这样浏览在线评价的潜在消费者就可以从评价中看到评价者购买商品的商品特征以及购买者的使用心得。由于，有些评价者的评价内容比较多，重点不够突出，商品评价浏览者需要仔细阅读商品评价，对于比较内容比较长的在线评价可能有些潜在用户就不去仔细看了，造成了占存储空间大的在线评价实际价值的缩水。所以，京东商城在2013

年7月14日又对在线评价进行了第二次升级改进，即增加了在线评价的“标签”项，

这些标签项就是让在线评价者把自己对商品的“使用心得”用关键的几个词浓缩出来，在线评价这根据自己的需要可以随意添加标签项以概况自己对所购买商品的印象，给浏览者提供了言简意赅的参考词语。

这些升级与改进行为对我们获取数据有什么影响呢？由于我们所使用网页抓取工具Metaseeker需要使用到抓取规则MetaStudio组件和数据抓取DataScraper组件，设置抓取规则时需要根据信息结构的不同对MetaStudio组件进行不同的设置，由于京东商城的在线评价改进升级了两次，所以在线评价信息的内容形式至少有三种，要抓取到这三种形式就需要对MetaStudio 组件进行三次不同的设置，由于

MetaSeeker是B/S（Browser/Server）结构，即通过客户端的浏览器（火狐Browser）对服务器（Server）下达指令，然后再执行，所以不同的结构对应的MetaStudio组件的不同设置不能在服务器（Server）上保存。这种情况下，对网页信息抓取者的网络环境要求很高，网络速度要求也很高，而且对不同形式的在线评价就要重新进行规则设置，网页信息抓取的工作量很大。

### 4.2.2 B2C购物网站咨询信息的结构

京东商城用户咨询信息的结构比较稳定，没有明显的结构性变化。如图4.4所示，京东商城在线咨询信息包括咨询用户名称、咨询日期、咨询时间、咨询内容、京东回复内容、京东回复日期和时间。



图 4.4 京东商城用户咨询样式图

由于在线咨询信息的结构比较稳定，并且比在线评价信息的数量要少很多，在后续的信息抓取过程中，用户咨询信息的抓取比较快，相对在线评价信息的抓取比较容易。

### 4.2.3 B2C购物网站在线商品评价信息和咨询信息的抓取过程

由于网页信息抓取工具MetaSeeker是在线工具，使用它抓取京东商城上的商品评价信息必须需要先通过网络在服务网站6上注册MetaSeeker服务账号、选择合适

6MetaSeeker注册服务的网站地址[http: //www. metacamp. cn/goomaster/home. htm](http://www.metacamp.cn/goomaster/home.htm) 。

的MetaSeeker版本进行下载安装、通过MetaStudio程序包进行网页信息抓取规则的运行与配置，然后通过DataScraper组件抓取网页信息的过程。

1. 注册网页抓取服务账号

打开GooSeeker网页进行注册成为GooSeeker的网站用户，成为网站用户后打开信息抓取的账号注册网站7，如图4.5所示，网页打开后填写用户名称和邮箱，点击提交进行注册，通过填写的邮箱对账号进行激活。激活以后会收到MetaSeeker运营支持系统发出的email，邮件中包含用户使用MetaSeeker帐号和口令。这里需要注意的是MetaSeeker服务的注册用户和GooSeeker社区的注册用户是不同的。



图 4.5 MetaSeeker注册信息

2. 选择适合自己电脑系统的软件包进行下载和安装

点击服务网站中的“下载链接”，如图4.6所示，在下载链接网页中输入邮件中的账户和密码，就可以选择相应的火狐和对应的网页下载工具版本进行下载和安装了。



图 4.6 下载MetaSeeker工具包需要先登录

工具包下载完成后，首先要运行Firefox 浏览器，然后后将metastudio 和

datascraper的程序包文件（分别是metastudio\_xxx. xpi和datascraper. xpi）拖到Firefox

上即可启动安装过程。安装完成后，如果在Firefox 浏览器状态栏的右下角显示

7 信息抓取服务账号注册网站地址为[http: //www. metacamp. cn/goomaster/register. htm? language=zh](http://www.metacamp.cn/goomaster/register.htm?language=zh) 。

“MetaStudio Installed"和”DataScraper Installed"，表示两者安装成功。如果安装成功，如图4.7所示，在Firefox的“工具”菜单上会出现“MetaStudio"和”DataScraper"两个菜单项。



图 4.7 安装成功以后在火狐浏览器工具栏中出现“MetaStudio”和

“DataScraper”

3．通过MetaStudio设置网页信息抓取规则

打开已经安装好网页抓取工具的火狐浏览器，在MetaStudio的状态栏的右下角可以看到MetaCamp和DataStore服务器的连接状态，在DataScraper上，连接状态显示在右下栏的Status Monitor操作区上表示连表示未连通。在连通状态 的前提下网页抓取工具才能使用，如图4.8所示，根据要抓取的网页信息进行Meta

-Studio的设置。首先在地址栏中输入要抓取信息的京东商城网址，点击“回车”键以后会在浏览器的左下角“浏览器”项中出现我们在地址栏中输入的网站页面，然后根据我们需要获取的信息来定义基本的抓取规则。



图 4.8 设置抓取规则

由于所有要抓取的信息不是同一个网页上，需要让抓取器自动翻页抓取，所以定义完抓取规则以后，如图4.9所示，还要在“Clue Editor”页签中定义自动翻页规则。



图 4.9 数据抓取自动翻页设置

4．使用DataScraper工具包进行网页信息的抓取

在第三步设置完成的情况下，而且在保持网络连接稳定的前提下，就要进行京东商城咨询和在线评价信息的抓取。点击火狐浏览器工具栏中的DataScraper选项，进行数据抓取工作。如图4.10所示，本研究抓取到的数据的原始状态是是XML格式文件，可以用浏览器形式打开查看文件中包含的抓取内容，需要经过进一步的处理与量化才能进行进一步的研究，具体的量化过程在下一节中阐述。



图 4.10 网页信息抓取后的原始状态

本研究一共选取了京东商城上30款手机的数据进行抓取，每款手机的咨询信息和在线评价信息需要用不同的设置规则进行抓取，每款手机咨询和在线评价信息平均所用抓取时间为1.5小时。抓取时间从2013年5月21日开始，但是由于网络稳

定性、工具熟练程度等各种原因，正式开始抓取从2013年9月20日开始一直持续

到10月13日，共获取23482条有效的网络购物咨询信息和213516条有效的在线商品评价信息。

## 4.3 B2C购物网站在线商品评价信息和咨询信息的要素量化

通过MetaSeeker工具抓取的在线评价信息和咨询信息的原始状态是XML格式文件，我们需要通过专业的工具进行处理和量化才能做进一步的研究。

### 4.3.1 京东商城在线商品评价信息的要素量化

由于京东商城在线评论信息中有标签特征，因此有利于简化评论信息的处理过程，具体信息量化过程如图4.11所示：



图 4.11 在线评价信息的处理步骤

（1）评价信息抓取

京东商城上用户的评价信息如图4.12所示，评论信息汇总过程中通过使用网页抓取工具Metaseeker抓取了包含用户名称、会员等级、省份、评论星级、评论日期和时间、购买日期、标签、有用数和回复数用户购买信息，生成为后缀为xml格式的文件。详细的抓取步骤在4.1章中已经阐述，这里就不再赘述。



图 4.12 京东商城购物网站标签类在线评论信息示例

（2）合并xml文件并转化成Excel文件。合并xml文件的方式有两种：手工合并或者使用工具合并。手工合并xml文件的方式很简单，就是打开每个xml文件，把文件内容拷贝到一个xml文件中。使用手工合并xml文件的前提是xml文件数量少，如果是数十个甚至成百上千的xml文件的合并，则必须使用xml专业合并工具。本研究使用谷歌公司提供的xml 专业的合并工具Mergex，如图4.13 所示，通过

Windows系统的命令框调用Mergex工具的文件合并命令，把MetaSeeker抓取工具抓取到的所有的多达数万个xml形式的咨询信息文件批量合并成一个。

研究中使用专业的合并工具的好处是比手工操作效率上高出很多，谷歌公司的

Mergex工具可以在瞬间合并数万个xml文件。将所有xml文件合并以后，打开新形成的xml文件，把文件另存为Excel文件，那么所有从网络上抓取到的在线评价信息就可以以excel文件形式打开并进行研究。



图 4.13 使用Mergex工具批量合并xml文件

（3）统计标签数量，并按照标签描述的特征进行量化，总结出网络购物用户购买手机商品后的在线评论主要关注“外观漂亮”、“性价比高”、“系统流畅”、“功能齐全”、“照相分辨率高”、“通话质量好”、“反应快”、“屏幕大”、“支持国产机”、“信号稳定”这十大关注点。

“外观漂亮”、“性价比高”、“系统流畅”、“功能齐全”、“照相分辨率高”、“通话质量好”、“反应快”、“屏幕大”、“支持国产机”、“信号稳定”这十大关注点不是我们主观想象出来的词语，而是现实用户在实际购买手机并进行使用以后对手机评价时使用的描述词语，如图4.14所示，这些词来源于京东商城在线评价用户做出评价时选择已经存在的标签或者用标签“自定义”功能自己创建出来的描述词汇。

在数据量化过程中，我们对30款手机的所有标签都进行了抓取，对于“后盖很漂亮”、“比较一般”、“铃声大”、“字体大”、“电池耐用”、“待机时间长”、“软件也不错”七个标签，由于他们仅仅在1款或者两到三款手机中出现，对于这种没有普遍性的这七个标签我们进行了舍弃处理，在研究中不再涉及。



图 4.14 京东商城在线评价信息标签来源图

### 4.3.2 京东商城咨询信息的要素量化

由于京东商城在线咨询信息中没有标签特征，因此相对于在线评价信息的量化处理过程在线咨询信息的处理过程相对复杂，需要借助更多的工具进行处理，具体信息量化过程如图4.15所示：



图 4.15 在线咨询信息的处理步骤

（1）京东商城上用户的咨询信息，使用网页抓取工具Metaseeker的抓取原理是把抓取对象以结构化单元抓取下来形成含有用户名称、咨询内容、咨询日期和时间、答复内容、答复日期和时间的xml文件。

（2）合并xml文件并转化成Excel文件格式。使用谷歌的xml文件合并工具Mergex，把所有xml形式的咨询信息文件合并成一个xml文件，然后另存为Excel文件，具体操作步骤与在线评价信息的合并类似，只要修改合并的文件路径即可操

作。

（3）把Excel中的咨询内容转化成Unicode形式的TXT文本，为分词做准备。

（4）使用中科院分词软件ICTCLAS把TXT文本进行分词，生成带空格的TXT

文本文件，为词频的统计做准备。

（5）本研究根据中科院分词软件ICTCLAS分词以后的TXT文本，手工选择出现的能够反映咨询者关注焦点的词创建了一个包含203个用户关注特征词的语料库8，然后使用牛津大学出版社的Wordsmith工具根据我们创建的特征词语料库进行词频统计，形成词频统计结果并以Excel文件保存。

（6）根据词频统计结果进行要素量化。本文根据词频统计的结果，把用户咨询内容分成“价格”、“支付”、“物流”、“售后服务”和“手机特征”五大关注点，根据其中的咨询情况，又把“手机特征”细分成“颜色等外部特征”、“手机卡支持”、

“功能”、“配件”、“软件操作系统”、“使用区域”和“质量”七个子关注点。根据词频统计结果把这些关注点进行了量化，为数据挖掘和数据分析做准备。

## 4.4 本章小结

本章详细阐述了研究所需要网页信息抓取工具的功能特征、数据的来源、在线评价信息和咨询信息抓取过程以及量化过程。如表4.2所示，所有研究数据来自京东商城“苹果”、“三星”、“华为”、“中兴”、“诺基亚”、“联想”、“HTC”、“索尼”、

“小米”、“酷派”、“魅族”、“步步高”和“天语”13个品牌30个型号手机，从2013

年9月20日到10月13日历经23天时间抓取到的23482条有效的网络购物咨询信

息和213516条有效的在线商品评价信息。

表 4.2 本研究涉及的13个品牌的30个型号的手机名单

|  | 品牌 | 型号 | 总评价数 | 晒单数 | 咨询数 |
| --- | --- | --- | --- | --- | --- |
| 1 | 苹果 | iPhone4 | 41312 | 1086 | 3050 |
| 2 | 苹果 | iPhone4S | 15748 | 304 | 1790 |
| 3 | 苹果 | iPhone5 | 28587 | 843 | 1440 |
| 4 | 三星 | Galaxy\_Note | 5553 | 143 | 720 |
| 5 | 三星 | Galaxy\_Note\_II | 4604 | 134 | 344 |
| 6 | 三星 | Galaxy\_SIII | 18154 | 498 | 1795 |

8 语料库详细内容请参考附录一。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 7 | 三星 | Galaxy\_S4 | 5905 | 158 | 997 |
| 8 | 华为 | Huawei\_D2 | 1414 | 51 | 303 |
| 9 | 华为 | Huawei\_P6 | 1126 | 48 | 309 |
| 10 | 华为 | Huawei\_Mate | 1213 | 60 | 115 |
| 11 | 华为 | Huawei\_A199 | 2981 | 81 | 476 |
| 12 | 中兴 | ZTE\_Z5 | 8053 | 1227 | 1903 |
| 13 | 中兴 | ZTE\_N5 | 1034 | 128 | 228 |
| 14 | 中兴 | ZTE\_V956 | 3916 | 119 | 200 |
| 15 | 诺基亚 | Nokia\_520 | 5345 | 135 | 443 |
| 16 | 诺基亚 | Nokia\_920T | 2229 | 90 | 239 |
| 17 | 联想 | Lenovo\_K900 | 6901 | 777 | 961 |
| 18 | 联想 | Lenovo\_A820T | 4215 | 108 | 222 |
| 19 | HTC | HTC\_One | 2467 | 118 | 474 |
| 20 | HTC | HTC\_329W | 2690 | 76 | 150 |
| 21 | HTC | HTC\_T528t | 1970 | 35 | 361 |
| 22 | 索尼 | SONY\_Z\_L36h | 2486 | 132 | 602 |
| 23 | 索尼 | SONY\_LT30p | 1190 | 54 | 141 |
| 24 | 小米 | MI\_2S | 15332 | 430 | 2161 |
| 25 | 小米 | MI\_2A | 14761 | 477 | 1510 |
| 26 | 酷派 | CoolPad\_5890 | 1969 | 76 | 219 |
| 27 | 酷派 | CoolPad\_8720 | 1588 | 41 | 166 |
| 28 | 魅族 | MX2 | 5417 | 161 | 542 |
| 29 | 步步高 | vivo\_X510w | 3269 | 448 | 1413 |
| 30 | 天语 | K-Touch\_T619 | 2087 | 54 | 208 |
| 总计 |  |  | 213516 | 8092 | 23482 |

在抓取到基本的在线评价和咨询信息以后再借助中科院分词工具和牛津大学的词频统计工具对抓取数据进行分词统计以后，对在线评价信息和咨询信息进行要素量化，便于进一步的数据挖掘和分析研究。

# **第 5 章** **B2C**购物网站在线信息的初步分析

本章根据第四章抓取到的213516条在线评价和23482条咨询信息，使用Excel的透视表以及Excel自带的各种函数功能对网络购物用户的省份、会员等级、消费品牌、咨询和评论时间及关注焦点等网络行为进行分析和研究。

## 5.1 网络购物用户的区域分布、会员等级以及消费品牌的研究

### 5.1.1 网络购物用户会员等级以及消费能力省份分布分析

由于我们抓取数据的时间是2013年9月20日至10月13日，所以我们的数据

都是2013年10月13日之前的在线评价和咨询信息，而京东商城的网络购物用户在

2013年10月15日之前是根据年消费金额来划分等级，消费金额达到规定数量以后

自动升级，如表5.1所示，网络购物用户共分为十个等级。而2013年10月15日京东商城对用户等级进行了调整，把原来的用户等级从十个等级调整为五个等级。调整规则为：调整日之前一年内有购物的“铁牌会员”调整为“铜牌会员”，原“铜牌会员”调整为“银牌会员”，原来“一钻会员”、“双钻会员”、“三钻会员”、“四钻会员”和“五钻会员”统一调整为“钻石会员”，其它级别享有的会员级别不变。所以在已有数据的基础上，不管是在用户等级调整前还是调整后我们都可以根据不同省份的不同会员等级，分析各省网络用户的消费能力和水平。根据京东会员调整规则，新形成的数据表5.2所示。本研究除了分析高消费水平的会员所在的省市区域分布研究以外根据研究需要所进行的分析都是以调整后的会员等级数据为分析基础，如果使用了旧的会员等级会做相应说明。

表 5.1 京东商城会员等级调整前表5.2京东商城会员等级调整后

| 会员等级 | 年最低消费金额（元） | 会员数 |  | 会员等级 | 年最低消费金额（元） | 会员数 |
| --- | --- | --- | --- | --- | --- | --- |
| 注册会员 | 0 | 142 |  | 注册会员 | 0 | 142 |
| 铁牌会员 | 100 | 12019 |  | 铜牌会员 | 100 | 12019 |
| 铜牌会员 | 2000 | 38557 |  | 银牌会员 | 2000 | 76728 |
| 银牌会员 | 5000 | 12019 |
| 金牌会员 | 10000 | 51772 |  | 金牌会员 | 10000 | 51772 |
| 钻石会员 | 30000 | 23726 |  | 钻石会员 | 30000 | 28044 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 双钻会员 | 100000 | 3464 |  |  |  |  |
| 三钻会员 | 300000 | 490 |
| 四钻会员 | 600000 | 149 |
| 五钻会员 | 1000000 | 215 |

在对数据进行会员与省份关系的具体分析之前，我们提出一个基本假设。

H1：假设网络购物人员的数量和消费能力与网络购物消费者所在省（市）的经济发展水平有正向相关关系。一个省（市）的经济发展水平越高，该省（市）的网络购物消费者数量越多，网络购物消费能力越强；反之，一个省（市）的经济发展水平越低，则该省（市）的网络购物消费者数量越少，网络购物消费能力越弱。

提出这样的假设是因为王秀丽和田祯祎[67]（2011）的研究表示，网络购物的参与度随着网民居住地行政级别的提升而提高，即网络购物的普及率在大城市中比中等城市高，中等城市比小城市、县城、乡镇和农村高。在农村、乡镇、县城和小城市网民中，网络购物的普及率为81%, 比大城市低10个百分点。由此，本研究假设经济发达水平较高的省市网络购物消费者人数较多，消费能力比较强。

由于不同等级会员的年消费能力不同，如表5.3所示，根据本文获取到的具有

省份地址内容的168683条网络购物用户信息进行统计，可以看到不同省（市）地区各个等级会员的数量情况，其中的“五钻会员”是钻石会员的一部分，由于“五钻会员”的条件是年消费金额在100万元以上，本文把“五钻会员”进行单独研究，把这种“五钻会员”归类为超级网络购物用户，他们的分布情况能够反映出一个省

（市）超级消费的能力。

表5.3 不同省份会员数量分布情况

|  | 注册会员 | 铜牌会员 | 银牌会员 | 金牌会员 | 钻石会员 | 五钻会员\* | 会员总数 | 2012 年  GDP 9 （亿元） |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 北京 | 32 | 1313 | 15985 | 16434 | 9898 | 37 | 43662 | 17879.40 |
| 广东 | 16 | 2862 | 13032 | 7959 | 4473 | 142 | 28342 | 57067.92 |
| 上海 | 11 | 310 | 6043 | 6086 | 3922 | 24 | 16372 | 20181.72 |
| 江苏 | 6 | 852 | 6124 | 4257 | 2447 | 3 | 13686 | 54058.22 |
| 四川 | 2 | 738 | 4613 | 2870 | 1339 | 0 | 9562 | 23872.80 |

9 2012年各省GDP数据来自于国家统计局网站：[http: //data. stats. gov. cn/workspace/index? m=fsnd](http://data.stats.gov.cn/workspace/index?m=fsnd)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 浙江 | 14 | 495 | 3538 | 1810 | 958 | 1 | 6815 | 34665.33 |
| 湖北 | 5 | 514 | 3007 | 1617 | 784 | 1 | 5927 | 22250.45 |
| ft东 | 8 | 674 | 2964 | 1088 | 464 | 1 | 5198 | 50013.24 |
| 天津 | 4 | 119 | 1839 | 1467 | 649 | 0 | 4078 | 12893.88 |
| 河南 | 6 | 563 | 2174 | 679 | 226 | 0 | 3648 | 29599.31 |
| 河北 | 4 | 478 | 2004 | 781 | 304 | 3 | 3571 | 26575.01 |
| 辽宁 | 3 | 208 | 1684 | 1086 | 388 | 0 | 3369 | 24846.43 |
| 福建 | 10 | 425 | 1573 | 675 | 332 | 0 | 3015 | 19701.78 |
| 陕西 | 3 | 338 | 1544 | 562 | 214 | 0 | 2661 | 14453.68 |
| 安徽 | 1 | 384 | 1396 | 502 | 236 | 0 | 2519 | 17212.05 |
| 湖南 | 2 | 321 | 1268 | 450 | 174 | 0 | 2215 | 22154.23 |
| 重庆 | 1 | 116 | 1131 | 650 | 247 | 0 | 2145 | 11409.60 |
| 黑龙江 | 4 | 144 | 994 | 384 | 132 | 0 | 1658 | 13691.58 |
| 广西 | 0 | 160 | 888 | 337 | 140 | 0 | 1525 | 13035.10 |
| 江西 | 1 | 259 | 858 | 281 | 109 | 0 | 1508 | 12948.88 |
| ft西 | 2 | 155 | 766 | 259 | 92 | 0 | 1274 | 12112.83 |
| 吉林 | 0 | 123 | 686 | 298 | 101 | 0 | 1208 | 11939.24 |
| 云南 | 4 | 102 | 521 | 238 | 91 | 0 | 956 | 10309.47 |
| 内蒙古 | 2 | 47 | 495 | 255 | 77 | 0 | 876 | 15880.58 |
| 贵州 | 0 | 111 | 419 | 183 | 52 | 0 | 765 | 6852.20 |
| 甘肃 | 1 | 78 | 313 | 155 | 33 | 0 | 580 | 5650.20 |
| 新疆 | 0 | 40 | 310 | 146 | 49 | 1 | 545 | 7505.31 |
| 海南 | 0 | 48 | 238 | 102 | 45 | 0 | 433 | 2855.54 |
| 宁夏 | 0 | 11 | 149 | 43 | 15 | 0 | 218 | 2341.29 |
| 青海 | 0 | 24 | 89 | 45 | 25 | 0 | 183 | 1893.54 |
| 西藏 | 0 | 2 | 66 | 56 | 19 | 0 | 143 | 701.03 |
| 香港 | 0 | 3 | 6 | 4 | 0 | 0 | 13 |  |
| 澳门 | 0 | 2 | 2 | 6 | 0 | 0 | 10 |  |
| 台湾 | 0 | 0 | 1 | 1 | 1 | 0 | 3 |  |
| 总计 | 142 | 12019 | 76720 | 51766 | 28036 | 213 | 168683 |  |

由于我们抓取到的用户信息是京东商城上手机购物用户的信息，从数量上不能

全面反映京东商城用户的所有用户信息，而利用“占比”来分析相关信息更有说服力。本文利用各省（市）不同等级会员数量占比对各个省份的网络购物用户的消费能力进行分析。各省（市）不同等级会员数量占比是指一省（市）不同等级的会员数量占所有会员数量的比率，反映一个省（市）的消费能力。各省（市）不同等级会员数量占比计算公式为Aij=Xpj/ Xcj \*100% (5-1)

其中，Xpj表示p省（市）j等级会员的数量，Xcj表示j等级会员的所有数量；

Aij表示i省j等级会员的占比，Aij越大表示该省（市）j等级的会员数量占总会员数量的比重越大。

经过分析可以得出，京东商城在北京市具有最多数量的会员，北京会员数量占比达到25.88%，其次是广东会员数量占比达到16.8%，上海会员数量占比9.7%排在第三位。表5.4列取了会员数量占总会员数量排名前十位的省份五种会员等级的百分占比。

表 5.4 会员数量占总会员数量排名前十位的省份

| 会员等级 | 北京 | 广东 | 上海 | 江苏 | 四川 | 浙江 | 湖北 | ft东 | 天津 | 河南 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 注册会员 | 22.5% | 11.3% | 7.7% | 4.2% | 1.4% | 9.9% | 3.5% | 5.6% | 2.8% | 4.2% |
| 铜牌会员 | 10.9% | 23.8% | 2.6% | 7.1% | 6.1% | 4.1% | 4.3% | 5.6% | 1.0% | 4.7% |
| 银牌会员 | 20.8% | 17.0% | 7.9% | 8.0% | 6.0% | 4.6% | 3.9% | 3.9% | 2.4% | 2.8% |
| 金牌会员 | 31.7% | 15.4% | 11.8% | 8.2% | 5.5% | 3.5% | 3.1% | 2.1% | 2.8% | 1.3% |
| 钻石会员 | 35.3% | 15.9% | 14.0% | 8.7% | 4.8% | 3.4% | 2.8% | 1.7% | 2.3% | 0.8% |
| 年消费百万  元 | 17.2% | 66.0% | 11.2% | 1.4% | 0.9% | 0.5% | 0.5% | 0.5% | 0.0% | 0.0% |
| **Total** | 25.9% | 16.8% | 9.7% | 8.1% | 5.7% | 4.0% | 3.5% | 3.1% | 2.4% | 2.2% |



图 5.1 各省会员数量与各省2012年GDP值的柱形图

从表5.4中可以看出北京市、广东省和上海市三省市地区的所有会员数量占总会员数的52.38%。年消费能力在30000元以上的钻石会员数量北京占比最多，占比数为35.3%，排名第二的是广东省占比为15.9%，上海市钻石会员占比为14%，排名第三，稍微低于广东省，这种年消费在30000元以上的会员占比情况与北京市排第一，广东省排第二，上海市排第三的会员占比情况吻合。

但是，如果按照十等级的会员分级情况详细分析年消费能力百万元以上的超级网络购物用户，会发现一个特殊情况，广东省的“五钻会员”占比最大达66.05%，排名第一，北京“五钻会员”占比为17.21%排名第二，上海“五钻会员”占比11.16%，排名第三。从表5-3中也可以看出，168683个具有省份信息的手机购物用户中只有北京市、广东省、上海市、江苏省、浙江省、ft东省、河北省和新疆八个省市自治区中有213个年消费能力在百万级以上的用户，占比为0.13%，由于消费能力在百万元以上的前提是消费者有百万元的存储或支付能力，这与当地的经济发展水平有直接关系，年消费百万元以上的“五钻会员”分布情况能够在一定程度上反映出当地的经济发展水平。

然后再看看会员数量相对较少的几个省（市）和地区的数据。根据表5-3的数据，京东商城会员数量最少的是台湾、澳门和香港三个地区，但是由于这几个地区的行政和物流等方面有着政治方面特殊限制，所以无法从经济发展水平程度与会员数量做直接关联研究。所以台湾、澳门和香港三个地区不在本文研究的省市和地区范围之内。西藏、青海、宁夏、海南、新疆、甘肃和内蒙古等地处偏远经济发展水平不发达的会员数量比较少，在这些省市中年消费水平在30000元以上的“钻石会员”数量也比较少，这些数据可以说明，一个省（市）的经济发展水平越低，则该省（市）的网络购物消费者数量越少，网络购物消费能力越弱。

综上所述，结合各省会员数量与各省2012年GDP值的柱形图5.1来看，会员数量、购买能力与经济发展水平有一定关系。通过数据分析可以验证假设1是准确的，即北京、广东和上海这种经济发达水平较高的省市网络购物消费者人数较多，网络购物消费能力比较强。反之，西藏、青海、宁夏海南、新疆、甘肃和内蒙古等经济发展水平越低的省市，网络购物消费者数量越少，网络购物消费能力越弱。

### 5.1.2 不同等级网络购物会员用户对手机品牌的购买情况

第二章中提过本研究共选取了京东商城销售的十三个品牌三十款手机进行研

究，这十三个手机品牌的选取参考京东商城上的手机销售情况进行选取的共213516

条包含会员等级和手机品牌型号信息的在线评价。

三十款手机的213516个不同等级消费者的数量情况如表5.5所示，根据不同手机型号的消费者数量来看，购买苹果手机iPhone4的消费者数量是41312人，排在第一位，购买苹果手机iPhone5的消费者数量是28587人，排在第二位，购买三星

Galaxy\_SIII的消费者数量是18154人，排在第三位，购买苹果iPhone4s为15748人排在第四位，排第五位的是购买小米2S型号手机的人数是15332，第六位的是购买小米2A型号的14761人。所以，根据我们的数据很容易看出京东商城上手机销售量最好的三个品牌是苹果、三星和小米三个品牌的手机，与现实中人们对手机品牌火热度的感觉比较吻合。

表 5.5 三十款手机的不同等级消费者的数量情况

| 品牌 | 型号 | 注册  会员 | 铜牌  会员 | 银牌  会员 | 金牌  会员 | 钻石  会员 | 总计 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 苹果 | iPhone4 | 32 | 264 | 19128 | 14607 | 7281 | 41312 |
| 苹果 | iPhone5 | 53 | 68 | 8856 | 11256 | 8354 | 28587 |
| 三星 | Galaxy\_SIII | 17 | 31 | 8204 | 5832 | 4070 | 18154 |
| 苹果 | iPhone4S | 14 | 33 | 5555 | 6009 | 4137 | 15748 |
| 小米 | MI\_2S | 39 | 1614 | 8232 | 3661 | 1786 | 15332 |
| 小米 | MI\_2A | 7 | 3458 | 6822 | 3224 | 1250 | 14761 |
| 中兴 | ZTE\_Z5 | 9 | 2168 | 4008 | 1390 | 478 | 8053 |
| 联想 | Lenovo\_K900 | 2 | 9 | 5714 | 775 | 401 | 6901 |
| 三星 | Galaxy\_S4 | 8 | 12 | 2033 | 2314 | 1538 | 5905 |
| 三星 | Galaxy\_Note | 3 | 10 | 1791 | 2329 | 1420 | 5553 |
| 魅族 | MX2 | 16 | 809 | 3309 | 966 | 317 | 5417 |
| 诺基亚 | Nokia\_520 | 12 | 1503 | 2412 | 1082 | 336 | 5345 |
| 三星 | Galaxy\_Note\_II | 3 | 9 | 1426 | 1823 | 1343 | 4604 |
| 联想 | Lenovo\_A820T | 12 | 1907 | 1564 | 550 | 182 | 4215 |
| 中兴 | ZTE\_V956 | 2 | 1328 | 1619 | 710 | 257 | 3916 |
| 步步高 | vivo\_X510w | 2 | 13 | 1955 | 960 | 339 | 3269 |
| 华为 | Huawei\_A199 | 1 | 475 | 1394 | 768 | 343 | 2981 |
| HTC | HTC\_329W | 2 | 512 | 1195 | 681 | 300 | 2690 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 索尼 | Z-L36h | 5 | 2 | 1167 | 860 | 452 | 2486 |
| HTC | HTC\_One | 8 | 9 | 1122 | 837 | 491 | 2467 |
| 诺基亚 | Nokia\_920T | 3 | 4 | 1463 | 571 | 188 | 2229 |
| 天语 | K-Touch\_T619 | 5 | 990 | 704 | 300 | 88 | 2087 |
| HTC | HTC\_T528t | 3 | 191 | 939 | 581 | 256 | 1970 |
| 酷派 | coolpad\_5890 | 5 | 651 | 809 | 376 | 128 | 1969 |
| 酷派 | CoolPad\_8720 | 3 | 369 | 732 | 335 | 149 | 1588 |
| 华为 | Huawei\_D2 | 2 | 3 | 630 | 501 | 278 | 1414 |
| 华为 | Huawei\_Mate | 6 | 117 | 612 | 337 | 141 | 1213 |
| 索尼 | SONY\_LT30p | 1 | 1 | 755 | 318 | 115 | 1190 |
| 华为 | Huawei\_P6 | 3 | 2 | 701 | 268 | 152 | 1126 |
| 中兴 | ZTE\_N5 | 1 | 33 | 581 | 294 | 125 | 1034 |

为了更直观的观察不同等级会员消费者对各个品牌不同型号的手机消费情况，本文依然采用百分占比数据来进一步分析会员对不同手机的消费情况。不同等级会员对不同品牌手机消费占比公式为：Bpj=Ypj/ Y \*100% (5-2)

其中，Ypj表示p品牌手机j等级会员的数量，Y表示会员的所有数量；Bij表示购买p品牌手机j等级会员的占比，Bpj越大表示j等级会员的数量占总会员数量的比重越大。

所以与表5.5对应的占比数据如表5.6所示，比较直观的可以看出，购买苹果手机iPhone4的消费者占比19.35%排在第一位，购买苹果手机iPhone5的消费者数量占比是13.39%排在第二位，购买三星Galaxy\_SIII的消费者数量占比是8.5%排在第三位，购买苹果iPhone4s为7.38%排在第四位，排第五位的是购买小米2S型号手机的人数占比是7.18%，第六位的是购买小米2A型号的6.91%。

表5.6 三十款手机的不同等级消费者的数量占比

| 品牌 | 型号 | 注册会员 | 铜牌会员 | 银牌会员 | 金牌会员 | 钻石会员 | 总计 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 苹果 | iPhone4 | 0.08% | 0.64% | 46.30% | 35.36% | 17.62% | 19.35% |
| 苹果 | iPhone5 | 0.19% | 0.24% | 30.98% | 39.37% | 29.22% | 13.39% |
| 三星 | Galaxy\_SIII | 0.09% | 0.17% | 45.19% | 32.13% | 22.42% | 8.50% |
| 苹果 | iPhone4S | 0.09% | 0.21% | 35.27% | 38.16% | 26.27% | 7.38% |
| 小米 | MI\_2S | 0.25% | 10.53% | 53.69% | 23.88% | 11.65% | 7.18% |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小米 | MI\_2A | 0.05% | 23.43% | 46.22% | 21.84% | 8.47% | 6.91% |
| 中兴 | ZTE\_Z5 | 0.11% | 26.92% | 49.77% | 17.26% | 5.94% | 3.77% |
| 联想 | Lenovo\_K900 | 0.03% | 0.13% | 82.80% | 11.23% | 5.81% | 3.23% |
| 三星 | Galaxy\_S4 | 0.14% | 0.20% | 34.43% | 39.19% | 26.05% | 2.77% |
| 三星 | Galaxy\_Note | 0.05% | 0.18% | 32.25% | 41.94% | 25.57% | 2.60% |
| 魅族 | MX2 | 0.30% | 14.93% | 61.09% | 17.83% | 5.85% | 2.54% |
| 诺基亚 | Nokia\_520 | 0.22% | 28.12% | 45.13% | 20.24% | 6.29% | 2.50% |
| 三星 | Galaxy\_Note\_II | 0.07% | 0.20% | 30.97% | 39.60% | 29.17% | 2.16% |
| 联想 | Lenovo\_A820T | 0.28% | 45.24% | 37.11% | 13.05% | 4.32% | 1.97% |
| 中兴 | ZTE\_V956 | 0.05% | 33.91% | 41.34% | 18.13% | 6.56% | 1.83% |
| 步步高 | vivo\_X510w | 0.06% | 0.40% | 59.80% | 29.37% | 10.37% | 1.53% |
| 华为 | Huawei\_A199 | 0.03% | 15.93% | 46.76% | 25.76% | 11.51% | 1.40% |
| HTC | HTC\_329W | 0.07% | 19.03% | 44.42% | 25.32% | 11.15% | 1.26% |
| 索尼 | Z-L36h | 0.20% | 0.08% | 46.94% | 34.59% | 18.18% | 1.16% |
| HTC | HTC\_One | 0.32% | 0.36% | 45.48% | 33.93% | 19.90% | 1.16% |
| 诺基亚 | Nokia\_920T | 0.13% | 0.18% | 65.63% | 25.62% | 8.43% | 1.04% |
| 天语 | K-Touch\_T619 | 0.24% | 47.44% | 33.73% | 14.37% | 4.22% | 0.98% |
| HTC | HTC\_T528t | 0.15% | 9.70% | 47.66% | 29.49% | 12.99% | 0.92% |
| 酷派 | coolpad\_5890 | 0.25% | 33.06% | 41.09% | 19.10% | 6.50% | 0.92% |
| 酷派 | CoolPad\_8720 | 0.19% | 23.24% | 46.10% | 21.10% | 9.38% | 0.74% |
| 华为 | Huawei\_D2 | 0.14% | 0.21% | 44.55% | 35.43% | 19.66% | 0.66% |
| 华为 | Huawei\_Mate | 0.49% | 9.65% | 50.45% | 27.78% | 11.62% | 0.57% |
| 索尼 | SONY\_LT30p | 0.08% | 0.08% | 63.45% | 26.72% | 9.66% | 0.56% |
| 华为 | Huawei\_P6 | 0.27% | 0.18% | 62.26% | 23.80% | 13.50% | 0.53% |
| 中兴 | ZTE\_N5 | 0.10% | 3.19% | 56.19% | 28.43% | 12.09% | 0.48% |
| 总计 | | 0.13% | 7.77% | 44.70% | 30.22% | 17.19% | 100.00% |

为了直接研究手机不同手机品牌的销售情况，根据已经获取到的评论信息进行研究，把表5.5中相同品牌的不同型号的手机消费者数量进行加总得到表5.7数据，

然后进行占比分析得到表5.8数据，对手机品牌和对应的会员等级进行分析。从表

5.8 所对应的图我们可以看出，苹果手机销售占比最大且有明显优势，购买占比为

40.1%，排名第二的为三星手机，销售量占比为16%，小米品牌手机销售量占比为

14.1%排名第三，苹果、三星和小米三大品牌的手机销量占统计量的70.2%，中兴、联想、诺基亚、HTC和华为五个品牌的占比在3.2%-6.1%之间，魅族、索尼、酷派、步步高和天语五个品牌的占比在0.98%-2.5%之间，占比较少。

表 5.7 不同等级用户对不同品牌的购买数量表

| 品牌 | 注册会员 | 铜牌会员 | 银牌会员 | 金牌会员 | 钻石会员 | 总计 |
| --- | --- | --- | --- | --- | --- | --- |
| 苹果 | 99 | 365 | 33539 | 31872 | 19772 | 85647 |
| 三星 | 31 | 62 | 13454 | 12298 | 8371 | 34216 |
| 小米 | 46 | 5072 | 15054 | 6885 | 3036 | 30093 |
| 中兴 | 12 | 3529 | 6208 | 2394 | 860 | 13003 |
| 联想 | 14 | 1916 | 7278 | 1325 | 583 | 11116 |
| 诺基亚 | 15 | 1507 | 3875 | 1653 | 524 | 7574 |
| HTC | 13 | 712 | 3256 | 2099 | 1047 | 7127 |
| 华为 | 12 | 597 | 3337 | 1874 | 914 | 6734 |
| 魅族 | 16 | 809 | 3309 | 966 | 317 | 5417 |
| 索尼 | 6 | 3 | 1922 | 1178 | 567 | 3676 |
| 酷派 | 8 | 1020 | 1541 | 711 | 277 | 3557 |
| 步步高 | 2 | 13 | 1955 | 960 | 339 | 3269 |
| 天语 | 5 | 990 | 704 | 300 | 88 | 2087 |
|  | 279 | 16595 | 95432 | 64515 | 36695 | 213516 |

表 5.8 不同等级用户对不同品牌的购买占比表

| 品牌 | 注册会员 | 铜牌会员 | 银牌会员 | 金牌会员 | 钻石会员 | 品牌占比 |
| --- | --- | --- | --- | --- | --- | --- |
| 苹果 | 0.0464% | 0.1709% | 15.7080% | 14.9272% | 9.2602% | 40.1127% |
| 三星 | 0.0145% | 0.0290% | 6.3012% | 5.7598% | 3.9205% | 16.0250% |
| 小米 | 0.0215% | 2.3755% | 7.0505% | 3.2246% | 1.4219% | 14.0940% |
| 中兴 | 0.0056% | 1.6528% | 2.9075% | 1.1212% | 0.4028% | 6.0899% |
| 联想 | 0.0066% | 0.8974% | 3.4086% | 0.6206% | 0.2730% | 5.2062% |
| 诺基亚 | 0.0070% | 0.7058% | 1.8149% | 0.7742% | 0.2454% | 3.5473% |
| HTC | 0.0061% | 0.3335% | 1.5249% | 0.9831% | 0.4904% | 3.3379% |
| 华为 | 0.0056% | 0.2796% | 1.5629% | 0.8777% | 0.4281% | 3.1539% |
| 魅族 | 0.0075% | 0.3789% | 1.5498% | 0.4524% | 0.1485% | 2.5370% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 索尼 | 0.0028% | 0.0014% | 0.9002% | 0.5517% | 0.2656% | 1.7217% |
| 酷派 | 0.0037% | 0.4777% | 0.7217% | 0.3330% | 0.1297% | 1.6659% |
| 步步高 | 0.0009% | 0.0061% | 0.9156% | 0.4496% | 0.1588% | 1.5310% |
| 天语 | 0.0023% | 0.4637% | 0.3297% | 0.1405% | 0.0412% | 0.9774% |
| 用户占比 | 0.1307% | 7.7723% | 44.6955% | 30.2155% | 17.1861% | 100.0000% |



图 5.2 不同等级用户对不同品牌的购买情况

从以上数据中还可以看出，年消费能力在2000-9999元之间的“银牌会员”是除了“天语”品牌之外的十二个品牌的第一大消费群体，“天语”品牌的第一大消费群体是“铜牌会员”；年消费能力在10000-29999元之间的“金牌会员”是苹果、三星、小米、诺基亚、HTC、华为、魅族、索尼和步步高品牌的第二大消费群体，年消费能力在100-1999元之间的“铜牌会员”却是中兴、联想、酷派的第二大消费群体。

苹果、三星、华为、索尼、HTC和步步高六个品牌的消费者共同特征是前三大主要消费群体是“银牌会员”最多，“金牌会员”排第二，“钻石会员”排第三位，

“铜牌会员”为第四大消费群体，“注册会员”最少；其它七个品牌的前三大消费群体是“银牌会员”、“金牌会员”和“铜牌会员”，“钻石会员”为第四大消费群体，

“注册会员”占比最少。

## 5.2 网络购物用户的咨询、购买行为以及对商品进行评论等行为规律研究

目前国内对于网络购物用户行为研究的还不是很多，大多数研究者都是采用问卷调查方式获取几百到上千份问卷调查然后对这些调查数据进行分析，被调查者中还有很多人从来没有进行过网络购物活动。饶曦（2010）提出网络购物用户通过网络进行购物行为以后网络购物消费者发表买后在线评论的主要渠道在购物网站上，其中90．2％的消费者在原购物网站商品下方发表评论，有近10％的消费者在原购物网站社区中发表评论，另有部分消费者在其他网站上或自己的博客中发表商品评论。但是，饶曦的这些具体数据的来源或者计算方式并没有阐述。2011年，北京大学王秀丽和田祯祎（2011）通过问卷调查方式获取了2625份有网络购物经历的有效问卷，通过对问卷进行分析以后发现，网民是否参与网络购物与性别、年龄、学历、职业、收入和居住地等特征有关，而与婚姻状况无关。或许是由于获取相关信息等原因，对于网络购物用户具体发起网络购物意向、下订单以及对所购买的商品进行评价的具体行为目前还没有人进行研究。本文利用专业网页信息抓取技术获取到的数据对网络购物用户的咨询、购买以及在线评价等行为进行了详细研究。

本论文根据网络用户的上午日期和时间对用户咨询和评价信息的时间利用

Excel的条件查询函数进行了相应统计，把咨询日期、在线评价日期、购买日期转化成“星期一”、“星期二”、“星期三”、“星期四”、“星期五”、“星期六”和“星期日”七个日期，本论文在统计过程中把“星期一”至“星期五”定为工作日，“星期六”和“星期日”定为周末节假日。把上网时间从0: 00点到24: 00点分成四个时间段，其中0: 00-6:00为“凌晨”时间、6: 00-12:00为“中午”时间、12: 00-18:00为“下午”时间和18: 00-24:00为“晚上”这四个时间段。在把日期和时间进行以上划分后，我们对网络购物消费者的网络购物行为习惯进行研究。

在进行详细的统计分析之前，我们对网络购物消费者的上网行为进行了一些基本假设。

H2：假设网络购物用户大多都是利用下班以后晚上18点至24点这一时间段上网的时间进行网络购物活动。

按照一般人的行为和生活规律，凌晨0点至早上6点这一时间段，正常生活规律的人应该处于“睡眠休息”时间，所以“凌晨”时间段进行网络购物的人最少，上午6点-下午18点应该处于上班路途中或者工作行为中，这些过程中可能由于工

作纪律要求、上班途中的网络连接状态等影响，“凌晨”、“上午”和“下午”这三个时间段进行网络购物行为的人应该比较少，而晚上18点至24点这一时间段进行网络购物的人比较多。

H3：假设网络购物用户利用周末时间“星期六”和“星期日”进行网络购物活动比较多。

按照王秀丽和田祯祎（2011）的研究，网民参与网络购物的比例与学历、收入均呈现较为显著的正向相关关系，学历越高的网民中，网络购物的参与度越高，而且网民收入水平越高，越青睐于通过网络进行购物。所以，我们可以认为通过网络进行购物的人应该以在职人员为主。由于工作纪律的要求等这些人应该主要在周末节假日时间进行网络购物活动。这里的网络购物活动包括在线浏览商品、对关注商品进行咨询活动、下订单活动、收货活动、支付活动、在线评价活动等相关行为。

### 5.2.1 网络购物用户咨询、购买、评价时间特征

1．网络购物用户咨询行为分析

研究中把23482条咨询信息按照用户的上网时间进行分析，利用不同时间段的咨询信息量对总咨询量的占比数值来分析不同时间段进行咨询活动的人的咨询活动情况，其中，把“年-月-日”形式的“咨询日期”利用Excel日期转换功能转换成“星期一”至“星期日”的格式，具体时间段的咨询数量分布情况如表5.9所示。

表 5.9 网络购物用户不同时间段咨询数量分布表

|  | 凌晨 | 上午 | 下午 | 晚上 | 总计 |
| --- | --- | --- | --- | --- | --- |
| 星期一 | 191 | 761 | 1314 | 1159 | 3425 |
| 星期二 | 227 | 778 | 1389 | 1261 | 3655 |
| 星期三 | 241 | 899 | 1245 | 1221 | 3606 |
| 星期四 | 232 | 760 | 1199 | 1219 | 3410 |
| 星期五 | 211 | 746 | 1259 | 1113 | 3329 |
| 星期六 | 240 | 717 | 1076 | 1098 | 3131 |
| 星期日 | 225 | 675 | 1052 | 974 | 2926 |
| 总计 | 1567 | 5336 | 8534 | 8045 | 23482 |

不同时间段的咨询信息量与总咨询量的占比公式为：Cm=Xm/ X \*100% (5-3)。

其中，Xm表示m时间段进行咨询活动的咨询信息数量，X表示所有咨询信息数

量；Cm表示m时间段的咨询信息量对总咨询量的占比，Cm越大表示在m时间段进行咨询活动的信息占总咨询数信息的比重越大。其中，m的取值为“凌晨”、“上午”、

“下午”和“晚上”四个时间段。

利用公式5.3和表5.9的数据进行分析得到5.10表中的数据，可以看出，根据

用户在上网活动在每天中的时间段分布的情况来看，在“下午”时间12: 00-18:00

点进行咨询的人占比最多，占总人数的36.34%，其次是利用“晚上”时间18: 00-24:00进行咨询的人占比排在第二，占总人数的34.26%，所以“下午”和“晚上”进行咨询的人比较多，共占总人数的70.6%。在“上午”进行咨询的人数占比只有22.72%，

“凌晨”咨询占比只有6.67%，这符合夜间休息时间没有多少人上网这一生活规律。但是，分析结果对于我们的假设2却不十分相符，我们的分析结果是在“下午”

时间12: 00-18:00点进行咨询的人占比最多，“晚上”时间18: 00-24:00进行咨询的人排在第二，这说明在下午上班时间有很多人在对自己感兴趣的商品进行咨询，或许网络购物用户是为了得到自己想知道的咨询问题的答案尽快得到商家的回复，所以在“下午”发起咨询活动，这样就有机会在当天“晚上”时间就可以看到答案。而在“晚上”发起咨询活动也可能是由于咨询者在晚上有购买意向，很着急得到回复，就在“晚上”时间进行咨询。

根据每周七天来分析，从表5.10中可以看出“星期三”进行咨询活动的最多，咨询数量占比为15.57% ，“星期二”进行咨询活动的数量占比为15.36%，排在第二位，比“星期三”进行咨询活动的数量非常接近，只是数量占比少了0.21%；在

“星期一”、“星期四”和“星期五”进行咨询活动的用户在14.18%-14.59%，而在“星期六”和“星期日”周末假期时间进行咨询活动的用户相对较少，分别为13.33%和

12.46%。通过实际的咨询活动的数据分析得出结论与设想的非常不一样，如果按照中国惯例，把“星期一”作为一个星期第一天来算，每个星期的第二天和第三天是进行购物咨询活动最多的两天，在每个星期的最后两天“星期六”和“星期日”这两天节假日中进行网上购物咨询活动的人最少。

表 5.10 网络购物用户咨询时间占比

| 咨询信息 | 凌晨 | 上午 | 下午 | 晚上 | 按星期 |
| --- | --- | --- | --- | --- | --- |
| 星期一 | 0.81% | 3.24% | 5.60% | 4.94% | 14.59% |
| 星期二 | 0.97% | 3.31% | 5.92% | 5.37% | 15.57% |
| 星期三 | 1.03% | 3.83% | 5.30% | 5.20% | 15.36% |
| 星期四 | 0.99% | 3.24% | 5.11% | 5.19% | 14.52% |
| 星期五 | 0.90% | 3.18% | 5.36% | 4.74% | 14.18% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 星期六 | 1.02% | 3.05% | 4.58% | 4.68% | 13.33% |
| 星期日 | 0.96% | 2.87% | 4.48% | 4.15% | 12.46% |
| 按时间段 | 6.67% | 22.72% | 36.34% | 34.26% | 100.00% |

2．网络购物用户在线评价活动分析

根据表5.11的213516条在线商品评论信息来按照用户的上网在线时间的基础数据情况进行不同时间段的数量占比进行分析，由于在线评价日期和时间是在同一个单元格内的，所以需要利用Excel的“数据分列”功能把在线评价的日期和时间拆分成两个单元格，然后把“年-月-日”形式的“评价日期”利用Excel日期转换功能转换成“星期一”至“星期日”的格式，得到表5.12的占比数据表。

表 5.11 网络购物用户不同时间段在线评价数量分布表

|  | 凌晨 | 上午 | 下午 | 晚上 | 总计 |
| --- | --- | --- | --- | --- | --- |
| 星期一 | 1151 | 9108 | 11429 | 9894 | 31582 |
| 星期二 | 1233 | 8266 | 11239 | 9421 | 30159 |
| 星期三 | 1321 | 8201 | 13306 | 10579 | 33407 |
| 星期四 | 2344 | 8443 | 10658 | 9900 | 31345 |
| 星期五 | 1288 | 8914 | 10966 | 9683 | 30851 |
| 星期六 | 1521 | 7067 | 9526 | 9175 | 27289 |
| 星期日 | 2083 | 7495 | 9811 | 9494 | 28883 |
| 总计 | 10941 | 57494 | 76935 | 68146 | 213516 |

表5.12的数据显示，从每天“凌晨”、“上午”、“下午”和“晚上”四个时间段方面来看，每天“下午”进行在线评价的人数最多，占到总评价数的36.03%，利用“晚上”时间进行评论的人比较多，占总评价数的31.92%，“下午”和“晚上”两个时间段的总评价人数占比总共达到67.95%，“上午”在线评论数量占比较少，只有26.93%，“凌晨”时间段的商品在线评价数量占比最少，只有5.12%，这符合休息时间没有多少人上网这一生活规律。而“上午”时间段在线评价数量占比不如“下午”和“晚上”数量占比多，可能是由于“上午”时间网络购物用户更多的精力在工作上，而到了“下午”时间，网络购物用户处理完一天的工作开始上网评价自己购买过的商品。

根据表5.12的数据还可以看出，每周“星期三”进行商品评论的人数占比最多，为15.65%，每周“星期一”、“星期二”、“星期四”和“星期五”进行评论的用户数量占比比较均匀，都在在14.12%-14.79%，而周末假期时间“星期六”和“星期日”

进行评论的用户数量占比相对较少，分别为12.78%和13.53%。

表 5.12 网络购物用户评价时间占比数据表

| 评价时间 | 凌晨 | 上午 | 下午 | 晚上 | 按星期 |
| --- | --- | --- | --- | --- | --- |
| 星期一 | 0.54% | 4.27% | 5.35% | 4.63% | 14.79% |
| 星期二 | 0.58% | 3.87% | 5.26% | 4.41% | 14.12% |
| 星期三 | 0.62% | 3.84% | 6.23% | 4.95% | 15.65% |
| 星期四 | 1.10% | 3.95% | 4.99% | 4.64% | 14.68% |
| 星期五 | 0.60% | 4.17% | 5.14% | 4.54% | 14.45% |
| 星期六 | 0.71% | 3.31% | 4.46% | 4.30% | 12.78% |
| 星期日 | 0.98% | 3.51% | 4.59% | 4.45% | 13.53% |
| 按时间段 | 5.12% | 26.93% | 36.03% | 31.92% | 100.00% |

3．网络购物用户下订单购买活动分析

利用网页信息抓取工具抓取到的213516条在线商品评论信息中有154808条商品评价信息具有“购买日期”，把“年-月-日”形式的“购买日期”利用Excel日期转换功能转换成“星期一”至“星期日”的格式，根据“星期X”的信息形式对在线商品评价信息按照手机购买品牌进行分析得到表5-13。

表 5.13 网络购物用户下订单购买商品不同时间数量占比

| 购买信息 | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 | 星期六 | 星期日 | 按品牌 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 苹果 | 5.91% | 4.72% | 4.44% | 4.38% | 4.41% | 4.27% | 4.14% | 32.27% |
| 三星 | 2.74% | 2.97% | 3.24% | 2.65% | 2.57% | 2.22% | 2.04% | 18.42% |
| 小米 | 2.24% | 2.87% | 2.44% | 2.08% | 2.15% | 2.11% | 1.79% | 15.68% |
| 中兴 | 1.00% | 1.34% | 1.35% | 1.31% | 1.22% | 1.05% | 0.98% | 8.25% |
| 联想 | 0.80% | 0.95% | 0.90% | 0.72% | 0.71% | 0.56% | 0.57% | 5.22% |
| 诺基亚 | 0.79% | 0.72% | 0.57% | 0.61% | 0.60% | 0.57% | 0.59% | 4.44% |
| HTC | 0.53% | 0.57% | 0.58% | 0.58% | 0.52% | 0.52% | 0.45% | 3.74% |
| 魅族 | 0.42% | 0.54% | 0.50% | 0.42% | 0.58% | 0.51% | 0.46% | 3.44% |
| 华为 | 0.42% | 0.49% | 0.44% | 0.44% | 0.40% | 0.41% | 0.37% | 2.96% |
| 索尼 | 0.29% | 0.34% | 0.35% | 0.27% | 0.25% | 0.24% | 0.28% | 2.03% |
| 酷派 | 0.26% | 0.27% | 0.28% | 0.28% | 0.26% | 0.28% | 0.26% | 1.88% |
| 天语 | 0.19% | 0.20% | 0.17% | 0.17% | 0.17% | 0.18% | 0.17% | 1.24% |
| 步步高 | 0.07% | 0.06% | 0.06% | 0.06% | 0.06% | 0.06% | 0.05% | 0.42% |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 按星期 | 15.65% | 16.04% | 15.32% | 13.98% | 13.89% | 12.97% | 12.15% | 100.00% |

从表5.13的数据中可以看出，每周“星期二”下订单购买手机商品的人数占比最多，占总人数的16.04%，每周“星期一”、和“星期三”下订单购买商品的用户数量占比相对较多，分别为15.65%和15.32%，每周“星期四”和周“星期五”下订单购买手机的人数占比分别为样本总量的13.98%和13.89%，而每周“星期六”和“星期日”下订单买手机的人数量占比相对较少，分别为12.97%和12.15%。

根据每周不同时间用户购买商品的时间分布来看，在每周的“星期一”至“星期三”购买商品的活动数量占总购买活动数量的比例最多，说明网络购物用户在周一至周三购买行为比较多，作为每个周的后两个工作日每周“星期四”和“星期五”购买行为次数一般，而在周末节假日休息期间的“星期六”和“星期日”购买行为活动数量占比最少。由此，可以验证假设3网络购物用户利用周末节假日时间“星期六”和“星期日”进行网络购物活动比较多这一假设不成立。

### 5.2.2 网络购物用户咨询与购买行为关系分析

本研究把获取到的23482条手机咨询信息和213516条在线评论信息进行关联分析，研究中把咨询信息与评论信息根据网络购物用户姓名唯一性进行关联，找出了

13个手机品牌每个品牌网络用户咨询相关手机信息后对应的购买数量，并进行分析。

如果有两个以上的Excel表格，或者一个表格内两个以上的sheet页面，拥有共同的数据——基础数据表，当其他的几个表格或者页面需要共享这个基础数据表内的部分数据，或者我们想实现当修改一个表格其他表格内共有的数据可以跟随更新的功能，都可以通过Vlookup函数实现。

在本研究中基础数据表是包含213516条在线评价的数据表，内涵“用户名称、用户等级、省份、评价日期和时间、购买日期、商品信息、评价星级”等信息，而另一个表是咨询信息表，包含“用户名称，咨询日期和时间，咨询内容、回复内容”等信息，由于评价信息和咨询信息表格的“用户名称”顺序和数量是不同的，如果想要将两个表格匹配到一个表格内，或者想将咨询信息表内的信息添加到在线评价信息表格中，而当我们修改基础数据的同时，新表格数据也随之更新。这样我们免去了一个一个查找，复制，粘贴的麻烦，也同时免去了修改多个表格的麻烦。

通过把咨询信息和在线评价信息进行关联以后，利用“咨询者中的购买用户数”除以“咨询用户数”就可以得到咨询者中的购买占比数据。通过计算得到表5.14的数据，可以看出，中兴品牌手机咨询者在进行咨询活动以后会有28.66%人购买中兴

手机，而步步高品牌的手机咨询者在进行咨询活动以后只有8.28%的人购买步步高手机，综合十三个手机品牌来看，网络购物用户在咨询活动以后大约会有15.52%的咨询者最终会购买手机产品，有84.48%的咨询者不会咨询完以后不会进行购买活动。

表 5.14 咨询与购买行为分析

| 品牌 | 咨询用户数 | 购买用户数 | 购买占比 |
| --- | --- | --- | --- |
| 天语 | 208 | 33 | 15.87% |
| 酷派 | 385 | 47 | 12.21% |
| 魅族 | 542 | 100 | 18.45% |
| 诺基亚 | 682 | 117 | 17.16% |
| 索尼 | 743 | 105 | 14.13% |
| HTC | 985 | 145 | 14.72% |
| 联想 | 1183 | 168 | 14.20% |
| 华为 | 1203 | 150 | 12.47% |
| 步步高 | 1413 | 117 | 8.28% |
| 中兴 | 2331 | 668 | 28.66% |
| 小米 | 3671 | 500 | 13.62% |
| 三星 | 3856 | 571 | 14.81% |
| 苹果 | 6280 | 923 | 14.70% |
| 总计 | 23482 | 3644 | 15.52% |

本文获取到213516条在线评价信息，这些评价的前提是购买了手机才能评价，

所以这 213516 条在线评价表示 213516 次手机购买行为。23482 条咨询信息经过研

究以后只有3644个用户在213516条评价信息中找到了完全一致的购物者，说明在

京东商城的213516 次购买行为中只有3644 次购买行为有咨询过程，

3644/213516\*100%=1.7%，所以，1.7%的手机购买者有买前咨询行为，98.3%的手机购买者并没有提前进行咨询活动。

### 5.2.3 网络购物用户购买行为与在线发表评价行为的关系分析

由于在评价信息中包含评价日期和时间以及购买日期和时间，京东商城上网络购物用户只有在购买完商品后才能进行评价，所以在线评价日期一定比购物日期要晚，研究中利用评价日期与购买日期的天数差额来计算网络购物用户购买行为与在

线评价行为的关系，结合表5.7的数据并通过计算可以得到表5.15数据。

表 5.15 购买与发表在线评价行为分析

| 会员等  级  手机品牌 | 注册会员 | 铜牌会员 | 银牌会员 | 金牌会员 | 钻石会员 | 平均数 | 评价总数  （销售数量） |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 步步高 | 7.9 | 33.8 | 34.7 | 37.8 | 32.0 | 35.3 | 3269 |
| 苹果 | 28.4 | 30.2 | 20.9 | 27.3 | 25.2 | 24.2 | 85647 |
| HTC | 5.7 | 16.2 | 23.2 | 24.6 | 23.7 | 22.9 | 7127 |
| 索尼 | 12.5 | 1.2 | 18.2 | 23.1 | 24.8 | 20.6 | 3676 |
| 三星 | 13.6 | 19.1 | 17.2 | 23.8 | 21.1 | 20.4 | 34216 |
| 酷派 | 4.9 | 13.1 | 20.8 | 21.9 | 19.7 | 18.6 | 3557 |
| 诺基亚 | 4.9 | 13.5 | 17.7 | 21.6 | 18.5 | 17.8 | 7574 |
| 小米 | 7.1 | 11.8 | 17.8 | 19.5 | 18.5 | 17.3 | 30093 |
| 联想 | 3.9 | 9.2 | 12.9 | 17.1 | 24.6 | 12.9 | 11116 |
| 天语 | 2.9 | 11.1 | 14.7 | 14.1 | 9.9 | 12.6 | 2087 |
| 华为 | 4.8 | 8.1 | 10.9 | 13.5 | 13.4 | 11.6 | 6734 |
| 中兴 | 9.5 | 9.4 | 11.0 | 11.1 | 11.2 | 10.6 | 13003 |
| 魅族 | 3.4 | 6.8 | 9.2 | 9.8 | 8.8 | 8.9 | 5417 |
| 平均数 2 | 13.1 | 11.0 | 17.4 | 23.3 | 22.1 | 19.2 | 213516 |

通过表5.15的数据可以得出以下结论，如果不考虑手机品牌因素，在“注册会员”、“铜牌会员”、“银牌会员”、“金牌会员”和“钻石会员”五个等级的会员中，

“铜牌会员”的在线评价与购买时间间隔最短，他们会在购买手机11天以后对手机

做出评价，然后就“注册会员”在购买手机13.1 天以后进行评价，“银牌会员”在

购买手机17.4天以后做出评价，“钻石会员”在购买手机22.1天以后进行评价，“金

牌会员“评价与购买间隔时间最长，他们会在购买手机23.3天以后才对所购买的手机进行评价。

不同品牌的手机购买用户评价时间与购买时间的间隔也有很大不同。“步步高”品牌手机用户对手机的评价与购买时间间隔最长，间隔时间为35.3天，“魅族”品

牌手机用户对手机的评价与购买时间间隔最短，间隔时间仅仅只有8.9天。结合销售数据来看，不同手机品牌网络购物用户的购买与评价间隔日期与手机的在线评价的数量（销售数量）没有显著关系。综合所有品牌的手机购物者，他们通常在购买相应产品19.2天以后对相关产品进行评价。

### 5.2.4 在线评价信息购买与评价时间差的多元线性回归分析

#### 5.2.4.1 多元线性回归分析理论介绍以及数据描述

多元线性回归分析也称为多重线性回归分析10，是最为常用的一种回归分析

（Regression）方法。多元线性回归分析设计多个自变量，用来处理一个因变量与多个自变量之间的线性关系，建立变量之间的线性模型并根据模型做评价和预测。

本研究中在抓取的213516条评价信息中包含了“用户等级”、“评价星级”、“评价回复数”、“购买日期”、“评价日期”以及“评价有用数”等信息。对于回归分析中一些数据的来源作如下解释。

其中“用户等级”是用户在京东商城中根据前一年的消费金额自动生成的会员等级，包括“注册会员”、“铜牌会员”、“银牌会员”、“金牌会员”和“钻石会员”五个等级。由于字符型变量无法做回归分析，为了便于进行回归分析研究中给五个等级的会员赋予0-4的值，用数字0代替“注册会员”，用数字1代替“铜牌会员”，

用数字2代替“银牌会员”，用数字3代替“金牌会员”，用数字4代替“钻石会员”。“评价星级”是网络购物用户在购买商品以后对商品进行评价时给的星数，感

觉质量差到质量好星数最少一颗最多五颗，如果感觉商品很差就给一颗星，如果感觉商品非常好最多可以给五颗星。研究中，把星数转换成数字形式，取值为1、2、

3、4和5。网络购物用户给的星级数与京东商城上商品的“好评数”、“中评数”和“差评数”是有关系的，京东商城默认的统计规则为：某种商品的“好评数”是用户给出四颗星和五颗星的评价的数量的累计值；“中评数”是用户给出三颗星和两颗星的评价的数量的累计值；而“差评数”是用户给出一颗星的评价的数量的累计值。这种包含四颗星和五颗星的“好评数”比较符合实际网络购物用户的感受，而“中评数”与“差评数”却与实际网络购物用户想表示的真实感受有所差别。由于一共有1至五颗星的五种感受，“中评”通常是取一至五数字的中间数即三最合适，所以三颗星的用户评价应该归为“中评”，而一颗星或者两颗星的评价应该归为“差评”，在京东商城的这种统计方式下，本来应该归为“差评”的两颗星的评价被归为“中评”了，所以京东商城的“中评数”比实际“中评数”大，京东商城的“差评数”比实际“差评数”要小。京东商城这种没有明显备注的调整某种程度上减少了其所售商品的差评数量或许是为了在感官上提升其所售商品的整体质量。但在本研究中，

10 定义来源于杨维忠、张甜编著，清华大学出版社出版的《SPSS统计分析与行业应用案例详解》，第85

页。

为了更真实的反应网络购物用户对商品的感受，研究中的“好评数”、“中评数”和“差评数”不直接抓取京东商城上的数值，而是按照实际抓取的评价的星级情况重新统计获取数值，“好评数”是用户给出四颗星和五颗星的评价的数量的累计值；“中评数”是用户给出三颗星的评价的数量的累计值；“差评数”是用户给出两颗星和一颗星的评价的数量的累计值。

“评价回复数”是某用户的评价一共得到了多少条回复。而“评价有用数”是指京东商城上每条成功发表的评价都带有一个“有用”按钮，如果其他浏览者认为这条评价对自己有用就可以点击这个“有用”按钮，“评价有用数”就是这条评价的“有用”按钮被点击的累计次数。

“购买日期”是指网络购物用户下订单的日期，“评价日期”是网络购物用户购买商品以后通过一段时间的使用对商品做出评价的日期。根据京东商城的规则，“评价日期”一定比“购买日期”大，研究中把“评价日期”与“购买日期”的差额天数定义为一个新的变量“购买评价日期差”。

#### 5.2.4.2 “购买评价日期差”与“用户等级”、“评论星级”、“评价回复数”以及“评价有用数”的多元回归分析

研究中把“购买评价日期差”作为被解释变量，把“用户等级”、“评论星级”、

“评价回复数”以及“评价有用数”作为解释变量，寻找“用户等级”、“评论星级”、“评价回复数”以及“评价有用数”与“购买评价日期差”变量之间的解释与被解释关系。

为了简化解释变量之间的多重共线性检验的操作，如表5.16所示在SPSS16.0版本中使用多元线性回归逐步回归分析法来进行分析。这个方法虽然好用，但是最后可能出现几个模型都比较合适，这时就需要比较这几个模型的优劣。

表5.16 使用逐步回归分析法

Variables Entered/Removeda

| Model | Variables Entered | Variables Removed | Method |
| --- | --- | --- | --- |
| 1 | 用户等级 | . | Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100). |
| 2 | 评价回复数 | . | Stepwise (Criteria: Probability-of-F-to-enter <= .050,  Probability-of-F-to-remove >= .100). |

**Variables Entered/Removeda**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Variables Entered | Variables Removed | Method |
| 1  2 | 用户等级  评价回复数 | .  . | Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).  Stepwise (Criteria: Probability-of-F-to-enter <= .050,  Probability-of-F-to-remove >= .100). |

a. Dependent Variable: 日期差

根据表5.17的回归模型总结表来看，调整后的R Square值非常小，这是因为购买评价日期差的实际决定因素是由网络购物用户购买商品的日期和评价日期同时决定，这种购买行为与评价行为与网络购物用户本身的行为特征有直接关系，比如购买商品的行为与购买者的收入、性别、年龄、受教育程度等个人特征有密切关系，所以不需要特别关注R Square值，只是通过R Square值可以知道目前本研究中的这些影响因素对于网络购物用户的购买评价日期差的解释程度不是很大。用于检验变量之间多重共线性的Durbin-Watson检验值是1.703，比较接近2，所以解释变量之间的多重共线性不严重。

表5.17 回归模型总结

Model Summaryc

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
| --- | --- | --- | --- | --- | --- |
| 1  2 | .118a  .120b | .014  .014 | .014  .014 | 28.113  28.107 | 1.703 |

a. Predictors: (Constant), 用户等级

b. Predictors: (Constant), 用户等级，评价回复数

c. Dependent Variable: 日期差

表5.18 系数结果表

Coefficientsa

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| --- | --- | --- | --- | --- | --- | --- |
| B | Std. Error | Beta |
| 1 | （Constant） | 10.077 | .218 |  | 46.227 | .000 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 用户等级 | 3.848 | .082 | .118 | 46.748 | .000 |
| 2 | （Constant） | 10.168 | .218 |  | 46.596 | .000 |
|  | 用户等级 | 3.834 | .082 | .118 | 46.582 | .000 |
|  | 评价回复数 | -.782 | .094 | -.021 | -8.330 | .000 |

a. Dependent Variable: 日期差

结合表5.17以及表5.18的结果来看，使用多元线性回归逐步回归分析法得到两个模型方程，P值为0.000，非常显著，模型具有统计学意义。第一个模型的结果是“购买评价日期差”只与“用户等级”有关系，解释变量只有“用户等级”一个变量，即模型一为：购买评价日期差=10.077+3.848\*用户等级。第二个模型的结果是“购买评价日期差”由“用户等级”和“评价回复数”两个变量解释，即模型二为：购买评价日期差=10.168+3.834\*用户等级-0.782\*评价回复数。这就需要对这两个模型的优劣进行比较。从购买行为、评价行为以及评价得到回复行为的先后顺序来考虑，由于网络购物用户需要先购买然后才能进行评价、评价结束以后其他用户才能进行回复行为，而购买行为决定购买日期，评价行为决定评价日期，这两个日期决定购买与评价日期差，而评价回复数对评价日期差没有多大的因果解释，所以选择第一个模型更合理。



图5.3 回归标准化残差的标准P-P图

为了判断随机扰动项是否服从正态分布，观察图5.3所示的标准化残差的P-P图可以发现，各观测的散点基本都在对角线上，据此可以判断残差服从正态分布。综上所述，“购买评价日期差”与“用户等级”、“评论星级”、“评价回复数”

以及“评价有用数”四个解释变量中的“用户等级”有关系，调整R方不是足够大，拟合优度不好，说明虽然知道了用户等级，但大体预测这个用户购买商品的日期与对商品做出评价日期的间隔不能仅仅依靠“用户等级”这个特征。

## 5.3 本章小结

本章根据研究抓取到的213516条在线评价和23482条咨询信息，使用Excel的透视表以及Excel自带的各种函数功能对网络购物用户的省份、会员等级、消费品牌、咨询和评论时间及关注焦点等网络行为进行分析和研究。经过分析得出以下

结论：

1. 网络购物人员的数量和消费能力与网络购物消费者所在省（市）的经济发展水平有正向相关关系。一个省（市）的经济发展水平越高，该省（市）的网络购物消费者数量越多，网络购物消费能力越强；反之，一个省（市）的经济发展水平越低，则该省（市）的网络购物消费者数量越少，网络购物消费能力越弱。

北京、广东和上海这种经济发达水平较高的省市网络购物消费者人数较多，网络购物消费能力比较强。反之，西藏、青海、宁夏海南、新疆、甘肃和内蒙古等经济发展水平越低的省市，网络购物消费者数量越少，网络购物消费能力越弱。

2. 每周“星期二”下订单购买手机商品的人数占比最多，每周“星期三”进行咨询和商品评论活动人数占比最多。而每周“星期六”和“星期日”进行咨询、下订单买手机以及购买以后进行商品评价的人数量占比相对较少。

每周“星期三”进行咨询活动的最多，“星期二”进行咨询活动的数量占比排在第二位，与“星期三”进行咨询活动的数量非常接近，只是数量占比少了0.21%；在“星期一”、“星期四”和“星期五”进行咨询活动的用户一般，而在“星期六”和“星期日”周末假期时间进行咨询活动的用户相对较少。

每周“星期三”进行商品评论的人数占比最多，每周“星期一”、“星期二”、“星期四”和“星期五”进行评论的用户数量占比比较均匀，而周末假期时间“星期六”和“星期日”进行评论的用户数量占比相对较少。

每周“星期二”下订单购买手机商品的人数占比最多，每周“星期一”、和“星期三”下订单购买商品的用户数量占比相对较多，，每周“星期四”和周“星期五”下订单购买手机的人数占比一般，而每周“星期六”和“星期日”下订单买手机的人数量占比相对较少。

3. 在“下午”时间进行咨询和在线评价的人占比最多，利用“晚上”时间进行咨询和在线评价的人占比次之，“上午”进行咨询和在线评价的人数占比较少，“凌晨”时间段进行咨询和在线评价的人占比最少。

在“下午”时间12: 00-18:00点进行咨询的人占比最多，占总人数的36.34%，其次是利用“晚上”时间18: 00-24:00进行咨询的人占比排在第二，占总人数的34.26%，所以“下午”和“晚上”进行咨询的人比较多，共占总人数的70.6%。在“上午”进行咨询的人数占比只有22.72%，“凌晨”咨询占比只有6.67%。

每天“下午”进行在线评价的人数最多，占到总评价数的36.03%，利用“晚上”时间进行评论的人比较多，占总评价数的31.92%，“下午”和“晚上”两个时间段的总评价人数占比总共达到67.95%，“上午”在线评论数量占比较少，只有26.93%，

“凌晨”时间段的商品在线评价数量占比最少，只有5.12%，这符合休息时间没有多少人上网这一生活规律。

4. 网络购物用户在咨询活动以后大约会有15.52%的咨询者最终会购买手机产品，有84.48%的咨询者不会咨询完以后不会进行购买活动。很多网络购物用户购买行为发生之前并不会对目标商品发起在线咨询行为，只有1.7%的手机购买者有买前对关注商品进行咨询的行为，98.3%的手机购买者并没有提前进行咨询活动就对商品进行购买。

很多网络用户对商品进行关注以后对相关产品进行咨询，得到的答复以后如果达不到自己的期望就不会购买相应产品，这可以解释为什么有那么多的咨询者咨询以后并不会购买相应产品。很多实际的网络购物用户有购物行为，但并没有进行咨询，是因为其他网络购物用户的咨询信息和购买评价信息都是公开可以看到的，很多网络购物者有相同的问题，看到别人问了相关问题并获得答案以后就没有必要再重复去问，而且网络卖家通常会把典型问题归结为“常见问题”放在明显的位置供上网者浏览。所以，绝大多数网络购物用户不会发起咨询活动。

5. 购买手机的网络购物用户通常在购买相应手机产品19.2天以后对相关产品进行评价。

下订单购买以后通过需要1天以上时间才能收到货物，然后用户需要办理手机卡对手机进行使用，通过充电、放电、打电话、发短信等多种功能多天的体验以后，才能对相应产品的相关功能进行评价，所以购买与评价之间十几天的间隔期间是一个合理范围。

6. “购买评价日期差”与“用户等级”、“评论星级”、“评价回复数”以及“评价有用数”等影响因素的多元回归分析

研究中把“购买评价日期差”作为被解释变量，把“用户等级”、“评论星级”、

“评价回复数”以及“评价有用数”作为解释变量，寻找“用户等级”、“评论星级”、“评价回复数”以及“评价有用数”与“购买评价日期差”变量之间的解释与被解释关系。通过多元逐步回归分析发现，“购买评价日期差”只受与网络购物用户特征有关的“用户等级”影响。

# **第 6 章** **B2C**网络购物用户商品评价效应的相关实证研究

在本章中，对在线评价信息结合咨询信息进行因子分子找出网络购物用户在对商品做出评价时对商品使用过程中的关注焦点以及购买之前进行咨询活动时对网络商城服务商的服务以及商品特征的关注焦点。然后通过回归分析得出在线评价信息所包含的的因素中对网络销售数量的影响因素。

## 6.1 网络购物用户在线评价和咨询信息的关注焦点研究

### 6.1.1 因子分析理论介绍

在科学研究或日常生活中，常常需要判断某一事物在同类事物中的好坏、优劣程度及其发展规律等问题。而影响事物的特征及其发展规律的因素（指标）是多方面的，因此，在对该事物进行研究时，为了能更全面、准确地反映出它的特征及其发展规律，就不应仅从单个指标或单方面去评价它，而应考虑到与其有关的多方面的因素，即研究中需要引入更多的与该事物有关系的变量，来对其进行综合分析和评价。多变量大样本资料无疑能给研究人员或决策者提供很多有价值的信息，但在分析处理多变量问题时，由于众变量之间往往存在一定的相关性，使得观测数据所反映的信息存在重叠现象。因此为了尽量避免信息重叠和减轻工作量，人们就往往希望能找出少数几个互不相关的综合变量来尽可能地反映原来数据所含有的绝大部分信息。而主成分分析和因子分析正是为解决此类问题而产生的多元统计分析方法。

根据清华大学出版社出版的陈胜可（2013）编著的SPSS教材定义，因子分析(Factor Analysis)是一种数据简化的技术。它通过研究众多变量之间的内部依赖关系，探求观测数据中的基本结构，并用少数几个独立的不可观测的假想变量来表示其基本的数据结构。这几个不可观测的假想变量能够反映原来众多变量的主要信息。原始的变量是可观测的显式变量，而假想变量是不可观测的潜在变量，称为因子。

行为及社会科学研究领域中，在建构效度的检验上，研究者及学者最常使用的方法就是因子分析法，因为因子分析主要的目的是用以认定心理学上的特征，借着公共因子的发现而确定观念的结构成分，根据量表或测验所提取的公共因子，可以获悉测验或量表有效测量的特征或态度。

吴明隆、涂金堂在《SPSS与统计应用分析》（2012）中提到，社会科学中，因子分析通常应用在以下三个层面：

1．显示变量间因子分析的组型（pattern）。

2．侦测变量间的群组（clusters），每个群组所包括的变量彼此间相关性很高，同质性较大，即将关系密切的个别变量合并为一个子群。

3．减少大量变量数目，使之成为一组涵括变量较少的统计自变量（成为因子），每个因子与原始变量间有某种线性关系存在，而用较少公共因子来代表多数、个别、独立的变量。

进行因子分析的基本步骤一共有四步，详细如下：

1.对数据进行标准化处理

2.估计因子的载荷矩阵

因子分析的基本模型如下：

*Z*1 *a*11*F*1 *a*12*F*2 ... *a*1 *p Fp* *c*1*U*1

*Z* *a F**a F* ... *a F* *c U*

2 21 1



22 2

2 *p p* 2 2

...

*Z*  *a*

*F*  *a*

*F*...  *a*

*F* *c U*

*M* *m*1 1

*M*2 2

*Mp p* m  *m*

其中，*Z*1、*Z*2…*Z*m是原始变量，*F*1、*F*2…*F*p

p <m。

*a*

是公共变量因子，

*ij*

是因子载荷，

因子分析基本模型表示为矩阵形式为Z=A·F+C U，

其中Z为m\*1的矩阵，A为m\*p因子载荷矩阵，F为p\*1因子变量矩阵，C 为

m\*m对角矩阵，U为m\*1矩阵。模型中相关要素的概念为

#### （1）因子载荷

在因子变量不相关的条件下，aij就是第*i*个原始变量与第*j*个因子变量的相关系数。aij绝对值越大，则Zi与Fi的关系越强。

#### （2）变量的共同度(Communality)

也称公共方差。Zi的变量共同度为因子载荷矩阵A中第*i*行元素的平方和。可见：Zi的共同度反应了全部因子变量对Zi总方差的解释能力。

#### （3）因子变量Fj的方差贡献

因子变量Fj的方差贡献为因子载荷矩阵A中第j列各元素的平方可

见：因子变量Fj的方差贡献体现了同一因子Fj对原始所有变量总方差的解释能力，

Sj/p表示了第j个因子解释原所有变量总方差的比例。

3.因子旋转（rotation）

建立因子分析数学模型的目的不仅要找出公共因子并对变量进行分组，更重要的是要知道没个因子的意义，以便对实际的现实问题作出科学分析。当因子载荷矩阵A的结构不便对主因子进行解释的时候，可以用一个正交矩阵右乘A（即对A实施一个正交变换）。由线性代数的知识可以知道，对A实施一次正交变换，对应坐标系就有一次旋转，便于对因子的意义进行解释。

4.估计因子得分

以公共因子表示原因变量的线性组合而得到的因子得分函数。可以通过因子得分函数计算观测记录在各个公共因子上的得分，从而解决公共因子不可观测的问题。

### 6.1.2 在线评价信息的因子分析

进行在线评价要素因子分析前，需要先进行KMO样本测度分析，对样本数据进行因子分析的适合程度进行判断。由表6.1的结果可见在线评价要素量表的KMO值为0．817，并且巴特莱特统计值不显著(sig．=．ooo)，根据马庆国[68]（2002）提出的标准，在线评价信息的数据很适合做因子分析找出在线评价者的关注焦点。

表6.1 在线评价信息KMO和Bartlett检验结果

| KMO and Bartlett's Test | | |
| --- | --- | --- |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .817 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 582.566 |
|  | df | 45 |
|  | Sig. | .000 |

表6.2 使用主成分法作因子提取方法

Communalities

|  | Initial | Extraction |
| --- | --- | --- |
| 外观漂亮 | 1.000 | .985 |
| 性价比高 | 1.000 | .939 |
| 系统流畅 | 1.000 | .990 |
| 功能齐全 | 1.000 | .984 |
| 照相分辨率高 | 1.000 | .965 |
| 通话质量与音质 | 1.000 | .967 |

|  |  |  |
| --- | --- | --- |
| 反应快 | 1.000 | .982 |
| 屏幕大 | 1.000 | .877 |
| 支持国产机 | 1.000 | .930 |
| 信号稳定 | 1.000 | .951 |

Extraction Method: Principal Component Analysis.

在SPSS 16．0软件中采用特征值大于1的标准，如表6.2所示，使用主成分法作为因子提取方法，并进行方差最大化(Varimax)正交旋转，得到未经旋转表6.3和旋转以后的因子载荷表6.4。

**表**6.3 未旋转的因子载荷表

**Component Matrixa**

表6.4 旋转以后的因子载荷表

**Rotated Component Matrixa**

|  | Component | | |  | Component | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 1 | 2 | 3 |
| 外观漂亮 | .974 | -.187 | -.043 | 外观漂亮 | .932 | .340 | -.033 |
| 性价比高 | .936 | .251 | .022 | 性价比高 | .712 | .612 | .241 |
| 系统流畅 | .983 | -.103 | .115 | 系统流畅 | .928 | .329 | .146 |
| 功能齐全 | .987 | -.077 | .053 | 功能齐全 | .910 | .380 | .106 |
| 照相分辨率高 | .972 | -.117 | -.074 | 照相分辨率高 | .894 | .406 | -.024 |
| 通话质量与音质 | .776 | -.576 | .182 | 通话质量与音质 | .970 | -.149 | -.057 |
| 反应快 | .985 | -.067 | .083 | 反应快 | .909 | .372 | .137 |
| 屏幕大 | .575 | .375 | -.637 | 屏幕大 | .231 | .858 | -.296 |
| 支持国产机 | .666 | .693 | -.086 | 支持国产机 | .258 | .863 | .345 |
| 信号稳定 | .155 | .484 | .832 | 信号稳定 | .053 | .020 | .973 |

Extraction Method: Principal Component Analysis. Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

A. 3 components extracted. a. Rotation converged in 4 iterations.

通过如表6.5分析数据和包含特征值的碎石图图6.1所示对在线评价信息共提取前3个因子作为主因子，而且前三个因子可以解释全部变量的95.69%，解释效果非常好。

表6.5 因子提取方法提取3个因子

**Total Variance Explained**

| Compone nt | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tota l | % of  Varian ce | Cumulative  % | Tota l | % of  Varian ce | Cumulative  % | Tota l | % of  Varian ce | Cumulative  % |
| 1 | 7.08  1 | 70.805 | 70.805 | 7.08  1 | 70.805 | 70.805 | 5.75  3 | 57.528 | 57.528 |
| 2 | 1.31  9 | 13.194 | 83.999 | 1.31  9 | 13.194 | 83.999 | 2.54  8 | 25.476 | 83.003 |
| 3 | 1.16  9 | 11.693 | 95.692 | 1.16  9 | 11.693 | 95.692 | 1.26  9 | 12.688 | 95.692 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| 1 |  |  |  |  |  |  |  |  |  |

4 .252 2.521 98.212

5 .096 .961 99.174

6 .031 .312 99.486

7 .020 .205 99.690

8 .020 .198 99.888

9 .009 .091 99.979

0 .002 .021 100.000

xtraction Method: Principal Component Analysis.

E



图6.1 在线评价因素特征值的碎石图

表6.6得到因子变量得分系数矩阵，通过因子变量得分系数矩阵就可以把在线评价信息包含的“外观漂亮”、“性价比高”、“系统流畅”、“功能齐全”、“照相分辨率高”、

“通话质量与音质”、“反应快”、“屏幕大”、“支持国产”和“信号稳定”十个要素总结为三个关注焦点。

表6.6 在线评价信息成份得分系数矩阵

Component Score Coefficient Matrix

|  | Component | | |
| --- | --- | --- | --- |
| 1 | 2 | 3 |
| 外观漂亮 | .179 | -.022 | -.089 |

|  |  |  |  |
| --- | --- | --- | --- |
| 性价比高 | .033 | .192 | .126 |
| 系统流畅 | .173 | -.040 | .059 |
| 功能齐全 | .156 | .000 | .023 |
| 照相分辨率高 | .150 | .030 | -.085 |
| 通话质量与音质 | .318 | -.346 | -.077 |
| 反应快 | .157 | -.007 | .049 |
| 屏幕大 | -.146 | .515 | -.314 |
| 支持国产机 | -.166 | .466 | .213 |
| 信号稳定 | -.030 | -.069 | .797 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

根据在线评价信息旋转后的成分载荷矩阵表6.4，本研究把“外观漂亮”、“性价

比高”、“系统流畅”、“功能齐全”、“照相分辨率高”、“通话质量与音质”和“反应快”7个要素在第一个因子上的载荷较大，因此作为第一公因子“手机基本功能”因子，把“屏幕大”和“支持国产”两个要素在第二个因子上的载荷较大，作为第二公因子“屏幕与国产性”因子；把“信号稳定”作为第三因子“信号稳定性”因子。根据表表6.7因子得分的协方差矩阵可知，新得到的“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子是完全正交的。

表 6.7 在线评价信息因子得分的协方差矩阵

Component Score Covariance Matrix

| Component | 1 | 2 | 3 |
| --- | --- | --- | --- |
| 1 | 1.000 | .000 | .000 |
| 2 | .000 | 1.000 | .000 |
| 3 | .000 | .000 | 1.000 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Component Scores.

综上所述，通过对在线评价要素进行因子分析，得到了三个新的不可观测因子：即“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子，这三个

因子对原来的十个要素的解释贡献率达到95.69%，而且三个因子完全正交。在后续章节的分析研究中，关于网络购物用户对商品的在线评价信息就用“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子要素。

根据表6.6的因子得分系数矩阵中得到的因子得分系数和原始变量的标准化值就可以计算不可观测的“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子的得分。

“手机基本功能”因子=0.179\*“外观漂亮”+0.033\*“性价比高”+0.173\*“系统流畅”+0.156\*“功能齐全”+0.150\*“照相分辨率高”+0.318\*“通话质量与音质”

+0.157\*“反应快”-0.146\*“屏幕大”-0.166\*“支持国产”-0.030\*“信号稳定”

“屏幕与国产”因子=-0.022\*“外观漂亮”+0.192\*“性价比高”-0.040\*“系统流畅”+0.000\*“功能齐全”+0.030\*“照相分辨率高”-0.346\*“通话质量与音质”

-0.007\*“反应快”+0.515\*“屏幕大”+0.466\*“支持国产”-0.069\*“信号稳定”

“信号稳定性”因子=-0.089\*“外观漂亮”+0.126\*“性价比高”+0.059\*“系统流畅”+0.023\*“功能齐全”-0.085\*“照相分辨率高”-0.077\*“通话质量与音质”

+0.049\*“反应快”-0.314\*“屏幕大”+0.213\*“支持国产”+0.797\*“信号稳定”

### 6.1.3 咨询信息的因子分析

一．咨询信息的初步因子分析

进行在线咨询信息的十一个要素因子分析前，也需要先进行KMO样本测度分析，对样本数据进行因子分析的适合程度进行判断。由表6.8的结果可见在线咨询要素量表的KMO值为0．767，并且巴特莱特统计值不显著(sig．=．ooo)，根据马庆国（2002）提出的标准，在线咨询信息的数据适合做因子分析找出在线咨询者的关注焦点。

表6.8 在线咨询信息KMO和Bartlett检验结果

| KMO and Bartlett's Test | | |
| --- | --- | --- |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .767 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 490.612 |
|  | df | 55 |
|  | Sig. | .000 |

表6.9 对咨询信息使用主成分法作因子提取方法

Communalities

|  | Initial | Extraction |
| --- | --- | --- |
| 价格方面 | 1.000 | .899 |
| 支付方面 | 1.000 | .859 |
| 物流方面 | 1.000 | .689 |
| 售后服务方面 | 1.000 | .955 |
| 卡支持方面 | 1.000 | .649 |
| 颜色等外部特征 | 1.000 | .873 |
| 功能方面特征 | 1.000 | .819 |
| 配件方面 | 1.000 | .915 |
| 软件及操作系统方面 | 1.000 | .943 |
| 使用区域方面 | 1.000 | .723 |
| 质量方面 | 1.000 | .924 |

Extraction Method: Principal Component analysis.

在SPSS 16．0分析软件中采用特征值大于1的标准，如表6.9所示，对咨询信息使用主成分法作为因子提取方法，并进行方差最大化(Varimax)正交旋转，得到未经旋转表6.10和旋转以后的因子载荷表6.11。

表 6.10 未旋转的因子载荷表

Component Matrixa

表 6.11 旋转以后的因子载荷表

Rotated Component Matrixa

|  | Component | |  | Component | |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 1 | 2 |
| 价格方面 | .887 | -.334 | 价格方面 | .917 | .240 |
| 支付方面 | .890 | -.257 | 支付方面 | .875 | .304 |
| 物流方面 | .707 | .434 | 物流方面 | .327 | .763 |
| 售后服务方面 | .947 | -.241 | 售后服务方面 | .913 | .350 |
| 卡支持方面 | .806 | .013 | 卡支持方面 | .650 | .476 |
| 颜色等外部特征 | .877 | .323 | 颜色等外部特征 | .529 | .770 |
| 功能方面特征 | .688 | .588 | 功能方面特征 | .222 | .877 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 配件方面 | .689 | .663 | 配件方面 | .180 | .940 |
| 软件及操作系统方面 | .959 | -.156 | 软件及操作系统方面 | .873 | .426 |
| 使用区域方面 | .787 | -.321 | 使用区域方面 | .828 | .192 |
| 质量方面 | .899 | -.341 | 质量方面 | .930 | .241 |

Extraction Method: Principal Component Analysis. Extraction Method: Principal Component Analysis.

A. 2 components extracted.

Rotation Method: Varimax with Kaiser Normalization.

A. Rotation converged in 3 iterations.

通过如表6.12分析数据和包含特征值的碎石图图6.2所示，对在线评价信息共提取前2个因子作为主因子，而且前两个因子可以解释全部变量的84.08%。

表6.12 因子提取方法提取2个因子

Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 7.690 | 69.909 | 69.909 | 7.690 | 69.909 | 69.909 | 5.644 | 51.311 | 51.311 |
| 2 | 1.559 | 14.169 | 84.078 | 1.559 | 14.169 | 84.078 | 3.604 | 32.767 | 84.078 |
| 3 | .515 | 4.685 | 88.763 |  |  |  |  |  |  |
| 4 | .467 | 4.244 | 93.007 |  |  |  |  |  |  |
| 5 | .391 | 3.555 | 96.562 |  |  |  |  |  |  |
| 6 | .197 | 1.787 | 98.349 |  |  |  |  |  |  |
| 7 | .087 | .788 | 99.137 |  |  |  |  |  |  |
| 8 | .047 | .431 | 99.568 |  |  |  |  |  |  |
| 9 | .030 | .273 | 99.841 |  |  |  |  |  |  |
| 10 | .011 | .096 | 99.937 |  |  |  |  |  |  |
| 11 | .007 | .063 | 100.000 |  |  |  |  |  |  |

Extraction Method: Principal Component Analysis.



图6.2 在线咨询因素特征值的碎石图

表6.13得到在线咨询信息因子变量得分系数矩阵，通过因子变量得分系数矩阵就可以把在线咨询信息包含的“价格方面”、“支付方面”、“物流方面”、“售后服务方面”、

“卡支持方面”、“颜色等外部特征”、“功能方面特征”、“配件方面”、“软件及操作系统方面”、“使用区域方面”以及“质量方面”十一个要素总结为两个关注焦点。

表6.13 在线咨询信息成份得分系数矩阵

Component Score Coefficient Matrix

|  | Component | |
| --- | --- | --- |
| 1 | 2 |
| 价格方面 | .218 | -.108 |
| 支付方面 | .190 | -.068 |
| 物流方面 | -.086 | .281 |
| 售后服务方面 | .190 | -.055 |
| 卡支持方面 | .081 | .067 |
| 颜色等外部特征 | -.027 | .235 |
| 功能方面特征 | -.145 | .360 |
| 配件方面 | -.173 | .399 |
| 软件及操作系统方面 | .160 | -.010 |
| 使用区域方面 | .203 | -.109 |

|  |  |  |
| --- | --- | --- |
| 质量方面 | .222 | -.111 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

根据在线咨询信息旋转后的成分载荷矩阵表6.11，本研究把“价格方面”、“支付

方面”、“售后服务方面”、“卡支持方面”、“软件及操作系统方面”、“使用区域方面”以及“质量方面”在第一个因子上的载荷较大，将第一个因子命名为“咨询信息因子一”；“物流方面”、“颜色等外部特征”、“功能方面特征”和“配件方面”在第二个因子上载荷较大，将其命名为“咨询要素因子二”。

表 6.14 在线咨询信息因子得分的协方差矩阵

Component Score Covariance Matrix

| Component | 1 | 2 |
| --- | --- | --- |
| 1 | 1.000 | .000 |
| 2 | .000 | 1.000 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

通过表6.14可以看出，利用因子分析方法求出的两个咨询信息的要素因子是正交

的。

二．对咨询信息进行进一步的因子分析

仔细分析“咨询信息因子一”包含的“价格方面”、“支付方面”、“售后服务方面”、“卡支持方面”、“软件及操作系统方面”、“使用区域方面”以及“质量方面”七个要素很难用一个特征词来描述新形成的不可观测因子一；而“物流方面”、“颜色等外部特征”、“功能方面特征”和“配件方面”四个要素也不容易用一个特征来描述。所以，根据研究的需要，本研究中把在线咨询信息的十一个要素根据需要进行多次因子分析以后，发现把“价格方面”、“支付方面”、“物流服务”和“售后服务”四个方面的要素不纳入因子分析的因素，而把“卡支持方面”、“软件及操作系统方面”、“使用区域方面”、“质量方面”、“颜色等外部特征”、“功能方面特征”和“配件方面”七个要素进行因子分析以后得到的两个因子更便于解释和研究。

因子分析前，仍然需要先进行KMO样本测度分析，对样本数据进行因子分析的适合程度进行判断。由表6.15的结果可见在线咨询要素量表的KMO值为0．683，

并且巴特莱特统计值不显著(sig．=0．000)，根据学者Kaiser（1974）的观点，如果

KMO的值小于0.5，则不太适合进行因子分析，进行因子分析时要求KMO值至少在0.6以上，所以数据适合做因子分析以达到降低维度目的。

表6.15 咨询信息的七要素KMO和Bartlett检验结果

| KMO and Bartlett's Test | | |
| --- | --- | --- |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .683 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 239.343 |
|  | df | 21 |
|  | Sig. | .000 |

通过如表6.16和表6.17分析数据和包含特征值的碎石图图6.3所示，对在线评

价信息共提取前2个因子作为主因子，而且前两个因子可以解释全部变量的85.21%，

解释效果相对较好。

表6.16 旋转以后的因子载荷表

Rotated Component Matrixa

|  | Component | |
| --- | --- | --- |
| 1 | 2 |
| 卡支持方面 | .729 | .399 |
| 颜色等外部特征 | .571 | .764 |
| 功能方面特征 | .261 | .898 |
| 配件方面 | .197 | .941 |
| 软件及操作系统方面 | .863 | .411 |
| 使用区域方面 | .864 | .147 |
| 质量方面 | .912 | .233 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

A. Rotation converged in 3 iterations.

表6.17 因子提取方法提取2个因子

**Total Variance Explained**

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 4.847 | 69.245 | 69.245 | 4.847 | 69.245 | 69.245 | 3.286 | 46.948 | 46.948 |
| 2 | 1.118 | 15.965 | 85.211 | 1.118 | 15.965 | 85.211 | 2.678 | 38.262 | 85.211 |
| 3 | .437 | 6.243 | 91.454 |  |  |  |  |  |  |
| 4 | .344 | 4.918 | 96.372 |  |  |  |  |  |  |
| 5 | .181 | 2.591 | 98.963 |  |  |  |  |  |  |
| 6 | .062 | .890 | 99.853 |  |  |  |  |  |  |
| 7 | .010 | .147 | 100.000 |  |  |  |  |  |  |

Extraction Method: Principal Component Analysis.



图6.3 在线咨询因素特征值的碎石图

表6.17得到在线咨询信息因子变量得分系数矩阵，通过因子变量得分系数矩阵就可以把在线咨询信息包含的“卡支持方面”、“软件及操作系统方面”、“使用区域方面”、

“质量方面”、“颜色等外部特征”、“功能方面特征”和“配件方面”七个要素进行因子分析以后得到的两个因子，即可以总结为两个关注焦点。

根据表6.16旋转以后的因子载荷表，本研究把“卡支持方面”、“软件及操作系统方面”、“使用区域方面”和“质量方面”四个在第一个因子上载荷较大的要素总结为

“产品性能”因子；把“颜色等外部特征”、“功能方面特征”和“配件方面”三个在第二个因子上载荷较大的要素命名为“产品规格”因子。这两个因子对原来的七个要素的解释贡献率达到85.21%，而且两个因子完全正交。

表6.18 在线咨询信息因子得分的协方差矩阵

Component Score Coefficient Matrix

|  | Component | |
| --- | --- | --- |
| 1 | 2 |

|  |  |  |
| --- | --- | --- |
| 卡支持方面 | .225 | -.006 |
| 颜色等外部特征 | .023 | .270 |
| 功能方面特征 | -.176 | .456 |
| 配件方面 | -.222 | .504 |
| 软件及操作系统方面 | .287 | -.044 |
| 使用区域方面 | .377 | -.204 |
| 质量方面 | .372 | -.169 |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Component Scores.

根据表6.18的因子得分系数矩阵中得到的因子得分系数和原始变量的标准化值

就可以计算每个观测的“产品性能”因子和“产品规格”因子的得分。

“产品性能”因子=0.225\*“卡支持方面”+ 0.287\*“软件及操作系统方面”+0.377\*“使用区域方面”+0.372\*“质量方面”+0.023\*“颜色等外部特征”-0.176\*“功能方面特征”-0.222\*“配件方面”

“产品规格”因子=-0.006\*“卡支持方面”-0.044\*“软件及操作系统方面”-0.204\*“使用区域方面”-0.169\*“质量方面”+0.270\*“颜色等外部特征”+0.456\*“功能方面特征”+0.504\*“配件方面”

## 6.2 在线评价信息对销售量的影响

通过前期的咨询信息和在线评价信息数据的抓取以及对数据的量化，然后再对相关数据进行因子分析得到新的因子数据等基础上，本研究先通过最小二乘法

（OLS）在不考虑购买日期的基础上利用已有的横截面数据对在线评价信息对手机销量的影响做了基本研究。然后，研究中根据购买日期数据增加月份维度将所有评价信息变成面板数据，通过对在线评价信息进行因子分析以后发现网络购物用户对国产手机与非国产手机有明显的关注，考虑到“国产”和“非国产”手机特性的不同，本文创新性的将“国产”和“非国产”特征考虑在内，利用固定效应模型对面板数据进行了第二步研究，第三步研究是在第二步固定效应模型基础上考虑内生性问题，通过滞后一期变量来考察其对销量的影响。在销量方面，由于上一期的销量可能会影响本期销量，即考虑到本期的网络购物用户的行为可能存在追随前期用户

的效应，所以最后一步研究是使用动态面板效应模型广义矩（GMM）来做进一步研究与证实。

### 6.2.1 基本变量的定义与描述以及相关的基本假设

由于京东商城上只有购买了商品才能对商品进行评价，所以根据前期的数据处理，由于抓取到的在线评价信息中包含手机购买时间信息，研究中根据月份把购买时间数据不同年份不同月份的数据进行分类然后统计，生成不同型号手机不同月份的评价数量。由于无法获取京东商城每个月手机的销售数量，研究中就用每个月的评价数量模拟当月的销售数量，而每款手机的总的销售数量用每款手机对应的在线评价的总数来模拟。研究中的面板数据来源根据每一个月作为一期，把前一个月的销售数量作为前一期的销售数量、本月的销售数量作为本期的销售数量，评价效价、在线评价的有用数也以月份做分期。

#### 6.2.1.1 手机价格与销售数量关系

“手机价格”来源就是抓取数据时京东商城网上显示的手机价格。通过哈尔. 范里安（2006）的《微观经济学原理》可以知道，由于典型的需求曲线具有负斜率，所以通常情况下，当一种商品的价格上升时，对这种商品的需求就会减少，因此只用商品的价格和需求数量的变动方向是相反的。也就是说，商品价格与销售数量是反方向变动的，即商品价格与销售数量负相关。

假设一：京东商城的手机销售数量与手机价格显著负相关。

#### 6.2.1.2 上市时间与销售数量关系

研究中还需要用到相应手机的“上市时间”，本文中某款手机在京东商城的上市时间就是抓取到该款手机的在线评价信息中最早的购买时间与抓取该款手机评价信息的日期的天数差。根据常识判断，在一个产品的生命周期内，产品上市时间（天数）越长，总的销售数量越多。

假设二：京东商城的手机销售数量与上市时间显著正相关。

#### 6.2.1.3 有用的评价数与销售数量关系

京东商城在每条评价旁边都设置一个“有用”按钮，浏览在线评价的任何一位浏览者不管是不是京东商城的会员如果认为这条评价对自己有用就可以点击这个按钮，所以每条评价旁边的“有用”按钮上的数字就是有多少人认为这条评价是有用的。不同款号手机对应的被“认为有用的评价数”是指该款手机所有在线评价对应的“有用”按钮被点击数量的总和。

Mudambi和Schuf（f 2010）明确提出了在线评价信息有用性（helpfulness of online

review）的理论概念，把在线评价信息有用性定义为在消费者决策过程中的感知价值。Pan和Zhang（2011）提出有用性越高的在线评价对消费者的说服效果越好。所以本文提出以下假设：

假设三：京东商城的手机销售数量与有用的评价数显著正相关。

#### 6.2.1.4 咨询数量与销售数量关系 “咨询数量”是指信息抓取时每款手机对应的咨询信息的数量。对产品进行咨

询信息的数量越多，说明对该产品关注的消费者越多，那么相应的购买该产品的消费者就越多。

假设四：京东商城的手机销售数量与该款手机对应的咨询数量显著正相关。

#### 6.2.1.5 评价效价与销售数量关系

由于研究中的每月的销售数量与每月的在线评价数量有关系，对于在线评价信息对销售量的影响方面本文使用了“评价效价”。评价效价是指消费者对产品或者服务的好坏进行评价。一般用评价分数或者正负面的评价的比例来衡量在线评价效价。评价效价对产品销售量的影响常常被称为说服效应（persuasive effect）：对产品或服务的评价分数越高，说明产品或服务的使用者对产品或者服务的满意度越高，越容易引发其他消费者对该产品或服务的态度转变，从而影响或说服他们更大可能的购买该产品或服务。

国内外有一些学者对于评价效价对商品销售量的影响进行了研究，Wenjing

Duan等学者（2008）发现在线评论的正面口碑效应对电影票房有正向影响，Garrett

Sonnier等学者（2011）进一步研究发现正面评论、中性评论和负面评论对电影票房的影响有着显著影响。Judith Chevalier等人（2006）通过提高在线评论的正面口碑效应对Amazon. com和Barnesandnoble. com网站的图书销售量有提升作用，而且一星评论的影响比5星评论的影响大。Clemons等学者（2006）对消费者对啤酒厂商的在线评价进行了研究，发现在线评价的分数与瓶酒的销售量正相关，并且认为前25％的评分情况对销量的预测最为精确。Qiang Ye等（2009）通过使用与Judith Chevalier相同的研究方法经过研究发现在线评论对中国旅馆的网络订购数量有显著关系。本研究中把评价效价定义为所有评价的分数之和与评价人数的比值，即评价平均分值。研究中对评价效价对销售数量的影响进行了研究，并做出以下假设。

假设五：京东商城的手机销售数量与该款手机对应在线咨询信息评价效价有显著正相关关系。

#### 6.2.1.6 前一期销售数量与本期销售数量关系（需补充理论或文献）

由于在线评价信息对于所有网络用户都是长时间可以看到的，评价的人越多说

明购买相应产品的人越多，本研究中用每月的评价数量模拟每月的销售数量，通常受从众效应的影响，如果前一期购买数量越多那么本期的购买数量也越多，即可以理解为前一期的销售数量越多那么本期的销售数量越多，于是提出假设六。

假设六：京东商城本期手机销售数量与该款手机前一期的销售数量有显著正相关关系。

6.2.1.7前一期评价有用数与本期销售数量关系

对于评价是否有用的判断前提是判断者应该先阅读评价，通常评价阅读者越多并且被阅读的在线评价确实对阅读者有用，阅读者才会点击“有用”按钮，点击有用按钮的阅读者越多，这些阅读者越有可能根据自己所阅读的在线评价信息做出购买决策，越有可能购买。

假设七：本期手机销售数量与前一期的评价有用数存在显著正向关系。

6.2.1.8前一期的评价效价与本期销售数量关系

评价效价是网络购物用户对所购买的商品进行评分的平均分值，是对所购商品满意度的综合判断表现。评价效价越高说明网络购物用户对该该商品的评分越高，网络购物用户满意度也越高，应该越有利于产品的销售。

假设八：本期手机销售数量与前一期的评价效价存在显著正向关系。

### 6.2.2 在线评价信息对销售量的影响

6.2.2.1最小二乘法对横截面数据分析在线评价对销售量的影响

在本节中，在不考虑购买日期的基础上抓取的数据是横截面数据，通过最小二乘法（OLS）对在线评价信息的评价效价、手机价格、上市时间、有用评价的数量、咨询数量与手机销量的关系做了基本研究。并且在此基础上考虑了这些变量对国产手机与非国产手机对销售量的影响。

研究中用到的OLS模型如公式6-1所示：

*LnVoli* **0**1*Ln*Pr*icei***2 *LnVali* **3*LnAgei* **4 *LnUsefuli* **5*LnConsultsi* *i*

（6-1）

下标i=1,2，…30代表不同款手机，*Voli*是指第i款手机的在线评价数量，

Pr *icei*代表第i款手机的价格，*Vali*代表网络购物用户评价的平均分值即所有评价的分数之和与评价人数的比值，*Agei*表示第i款手机的上市时间的月数，*Usefuli*表示第i款手机对应的在线评价的有用数，*Consultsi*是指第i款手机产品对应的咨询信息的数量，*i*表示该模型的误差项。

表 6.19 横截面数据描述性统计

| 变量名称 | Variable | Obs | Mean | Std. Dev. | Min | Max |
| --- | --- | --- | --- | --- | --- | --- |
| 销售量 | LnVol | 30 | 8.329645 | 0.999686 | 6.94119 | 10.62891 |
| 价格 | LnPrice | 30 | 7.499445 | 0.704245 | 5.293305 | 8.556222 |
| 评价效价 | LnVal | 30 | 3.067015 | 0.33698 | 2.523988 | 3.840018 |
| 上市时间 | LnAge | 30 | 5.310971 | 0.601012 | 4.219508 | 6.46925 |
| 评价有用数 | LnUseful | 30 | 5.354566 | 0.916616 | 4.077538 | 7.321188 |
| 咨询数 | LnConsults | 30 | 6.229163 | 0.951163 | 4.744932 | 8.022897 |

在模型中我们对这些因变量和自变量进行了对数变换。对数变换是因为在线评价数量取对数后与销售量相关，而且把潜在的非线性关系变为线性关系，使回归模型的结果更加稳健。表6. 19是对横截面数据的基础描述性统计，描述了经过对数转

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| VARIABLES | lnvol | lnvol | lnvol | lnvol |
| Lnprice | -0.620\*\*\* | -0.597\*\*\* | -0.588\*\* | -0.665\*\*\* |
| -0.163 | -0.162 | -0.231 | -0.23 |
| Lnval | 0.705\*\* | 0.645\*\* | 0.566\* | 0.52 |
| -0.281 | -0.263 | -0.312 | -0.351 |
| Lnage | 0.0497 | 0.0863 | -0.313 | -0.0166 |
| -0.12 | -0.115 | -0.304 | -0.214 |
| Lnuseful | 1.101\*\*\* | 0.897\*\*\* | 1.085\*\* | 0.786\*\*\* |
| -0.0768 | -0.248 | -0.435 | -0.262 |
| Lnconsults |  | 0.205 | 0.155 | 0.266 |
|  | -0.233 | -0.305 | -0.246 |
| Constant | 4.657\*\*\* | 4.288\*\*\* | 6.119\*\*\* | 5.803\*\* |
| -0.858 | -0.987 | -1.142 | -2.244 |
| Observations | 30 | 30 | 11 | 19 |
| R-squared | 0.877 | 0.882 | 0.939 | 0.827 |
| 注意：括号中的值为异方差稳健标准误 | | | | |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | | | |

换以后样本中各个变量的均值、标准差、最小值和最大值，表6.20是回归结果。表6.20评价信息对销售量的影响（横截面数据，OLS回归）

根据表6.20结果来看，模型1中仅仅针对评价信息相关因素不考虑咨询信息的

影响将在线评价信息总数量与手机价格、评价效价、手机上市时间、和评价有用数进行了回归。模型1的回归结果显示，手机价格、在线评价效价以及在线评价的有用数三个控制变量对手机的销售量这个因变量有显著影响。而手机上市时间对手机销售量没有显著影响。其中，Lnprice的系数是负向显著的（**1 =-0.620, p<0.01）,

说明手机价格越高销售量越低，这比较符合假设一的预期，价格越高，则购买的消

费者越少。Lnval的系数是正向显著的（**2 =0.705, p<0.05），说明评价效价越高越有利于提高销售量，实证结果显著符合假设五的预期，评价效价越高越有利于促进消费者的购买行为，反之如果评价效价越低越不利于促进消费者的购买。Lnuseful

的系数是正向显著的（**4 =1.101, p<0.01），符合假设三的预期，说明评价有用数越大越有利于销售。

由于模型1中没有把咨询信息数量对销售量的影响放入模型，所以在模型2 中

加入咨询信息数量并进行回归。模型2得到的结论与模型1类似，手机价格对手机的销售量有负向显著影响（**1 =-0.597, p<0.01），符合假设一的预期。在线评价效价对手机的销售量这个因变量有正向显著影响（**2 =0.645, p<0.05），验证了假设五的预期。在线评价的有用数对手机的销售量有正向显著影响（**4 =0.897, p<0.01），这一结果符合假设三的预期。而手机上市时间对手机销售量依然没有显著影响，即

假设二没有得到有效支持。咨询信息数量对销售数量的影响也不显著，假设四的预期没有得到显著支持。

模型3与模型4是在考虑国产手机与非国产手机不同的情况下，因变量销售量

与价格、评价效价、咨询数量等控制变量之间的关系。其中模型3是针对苹果、三

星、诺基亚和索尼四个品牌的11款非国产手机进行的回归分析，模型4是对华为、

中兴、小米、联想、HTC、酷派、魅族、步步高和天语九个品牌的19款国产手机进行的回归分析。将两个模型的回归结果结合来看，不管国产手机还是非国产手机的销售量与价格都是负相关，国产手机价格对销量的负影响系数**1 =-0.665（p<0.01），非国产手机价格对销量的负影响系数**1 =-0.588（p<0.01）。非国产手机的评价效价

比国产手机的评价效价对销量的影响明显，这种结果与预期结果可能有差异，预期

结果应该是消费者对国产手机的信任度不如非国产手机，消费者应该更看重国产手机的评价效价，而OLS实证结果并非如此。OLS实证结果与预期结果存在差异的原因可能在于OLS回归模型中使用横截面数据，没有考虑时间因素，而且样本容量不是足够大，这里只能明确判断评价效价对销售量有正相关关系，不能说是因果关系，或许也与非国产手机的评价确实又多又好。国产手机与非国产手机中，评价的有用数对销售量有正向显著影响，国产手机的影响系数是（**4 =0.786, p<0.01），

非国产手机的影响系数是（**4 =1.085, p<0.05）。

综上所述，在不考虑时间因素的前提下通过对样本的横截面数据进行包含多种控制因素的多次最小二乘法回归分析可以得出以下结论：

1.手机价格对手机销售数量呈显著的负影响，假设一的结果得到了实证结果的显著支持，即手机价格越高，其销售量越低。

2.评价效价对手机销售数量呈显著的正影响，假设五的结果得到了实证结果的显著支持，即对手机的在线评价的平均分值越高，其销售量越高。

3.上市时间对手机销售量的影响不显著，假设二没有得到显著支持。这是因为面板数据没有考虑时间因素，也有可能网络销售卖家出现供货不及时导致商品短缺等因素影响。

4.评价有用数对手机销售数量呈显著的正影响，假设三的结果得到了实证结果的显著支持，即对手机的在线评价的评价有用数越高，其销售量越高。

5.咨询数量与手机销售数量有正相关关系，但并不显著，所以假设四没有得到显著支持。

6.2.2.2固定效应模型对面板数据分析在线评价对销售量的影响

研究在线评价信息对销量的影响时，很可能产生遗漏相关变量的问题（Omitted Variable Problem）。在回归模型中，遗漏了重要的相关变量是导致内生性

（Heterogeneity）的重要原因。因为一旦这个被遗漏的变量同时与因变量和目标自变量相关，那么目标自变量与随机扰动项的无关性假定就不再成立了，这时如果用最小二乘法（OLS）进行估计，目标自变量的估计结果就是有偏的。在本小节中，研究中考虑了时间因素，根据不同款型手机的用户购买日期所在的月份进行统计形成面板数据，然后用固定效应模型（Fixed Effects Model）进行回归分析，研究中使用固定效应模型（Fixed Effects Model）而非随机效应模型(Random Effects Model)是因为豪斯曼检验（Hausman Test）的结果支持固定效应（Fixed Effects Model）优于随机效应模型(Random Effects Model)。在回归过程中，引入时间控制变量Year

Dummy，并在回归过程中对时间因素进行了控制，把OLS回归的结果与固定效应模型（Fixed Effects Model）回归的结果放在一起进行比较。考虑到在线评价信息与网络购物用户的购买行为之间存在很强的内生性问题，本文中不仅用同期的因变量与自变量进行了回归（回归结果如表6.21 所示），还用因变量与滞后一期的自变量

（在线评价的效价和有用数量）进行了回归（自变量滞后一期的回归结果如表6.23

所示），这样做不仅处理了内生性问题，也考察了检验实证结果的稳健性。通过表

6.21和表6.23的回归结果来看，回归结果还是有些差异的。

研究中用到的固定效应模型如公式6-2所示：

*LnVolit***0**1*LnValit***2 *LnUsefulit**i**t**it*

（6-2），

其中*Volit*代表第i款手机第t月的在线评价数量用来模拟因变量第i款手机第 t

月的手机销售量，*Valit* 代表第i款手机第t月的评价效价，*Usefulit*代表第i款手机第

t月在线评价的有用数，*i*是指随手机个体变化而不随时间变化的要素，t表示随时间变化而不随个体变化的因素即时间要素，*it*表示该模型的误差项。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ALL | ALL | CHN Brand11 | | Foreign Brand12 | |
|  | OLS | FE | OLS | FE | OLS | FE |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| VARIABLES | lnvol | lnvol | lnvol | lnvol | lnvol | lnvol |
| lnval | 5.101 | 2.021 | 6.422\* | 7.628\*\* | 1.147 | -7.587 |
|  | (3.821) | (4.911) | (3.554) | (3.506) | (5.514) | (5.113) |
| lnuse | 0.621\*\*\* | 0.519\*\*\* | 0.431\*\*\* | 0.488\*\*\* | 0.831\*\*\* | 0.521\*\*\* |
|  | (0.112) | (0.0746) | (0.0991) | (0.0846) | (0.164) | (0.102) |
| Year Dummy | YES | YES | YES | YES | YES | YES |
| Constant | -5.572 | -1.374 | -6.802 | -10.25\* | -0.463 | 13.88 |
|  | (5.647) | (7.557) | (5.045) | (5.302) | (8.199) | (8.010) |
| Observations | 266 | 266 | 123 | 123 | 143 | 143 |
| R-squared | 0.618 | 0.619 | 0.612 | 0.579 | 0.700 | 0.728 |
| Number of id |  | 30 |  | 19 |  | 11 |
| 注意：括号中的值为异方差稳健标准误 | | | | | | |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | | | | | |

表 6.21 评价信息对销售量的影响（面板数据，固定效应模型）

自变量滞后一期的固定效应模型回归用到的模型与自变量没有滞后的回归模型有稍微差异，自变量滞后一期的固定效应的模型如公式6-3所示：

*LnVolit***0**1*LnValit*1**2 *LnUsefulit*1*i**t**it*

（6-3），其中*Volit* 代表第 i

款手机第t月的在线评价数量用来模拟因变量第i款手机第t月的手机销售量，*Valit*1

11 CHN Brand是指本研究中的华为、中兴、小米、联想、HTC、酷派、魅族、步步高和天语九个国产品牌的所有手机样本。

12 Foreign Brand是指针对苹果、三星、诺基亚和索尼四个非国产品牌的所有手机样本。

代表第i款手机第t-1月的评价效价，*Usefulit*1代表第i款手机第t-1月在线评价的有用数，*i*是指随手机个体变化而不随时间变化的要素，t表示随时间变化而不随个体变化的因素即时间要素，*it*表示该模型的误差项。

表 6.22 滞后一期自变量的描述统计

| 变量 | Variable | Obs | Mean | Std. Dev. | Min | Max |
| --- | --- | --- | --- | --- | --- | --- |
| 销售量 | LnVol | 271 | 5.863936 | 1.402266 | 0.693147 | 8.850087 |
| 评价效价 | LnVal | 271 | 1.524533 | 0.034907 | 1.386294 | 1.609438 |
| 评价有用数 | LnUseful | 266 | 3.906712 | 1.512412 | 0 | 6.839477 |
| 滞后一期评价效价 | Lag1Val | 231 | 1.525266 | 0.034171 | 1.386294 | 1.609438 |
| 滞后一期评价有用数 | Lag1Useful | 230 | 4.045674 | 1.422574 | 0 | 6.839477 |

表6.22 是对在线评价的效价和有用数滞后一期以后的变量进行描述性统计，描

述了样本中相关变量的均值、标准差、最小值和最大值。表6.23是自变量滞后一期的回归结果。

表 6.23 自变量滞后一期评价信息对销售量的影响（面板数据，固定效应模

型）

|  | ALL | ALL | CHN Brand | | Foreign Brand | |
| --- | --- | --- | --- | --- | --- | --- |
|  | OLS | FE | OLS | FE | OLS | FE |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| VARIABLES | lnvol | lnvol | lnvol | lnvol | lnvol | lnvol |
| lag1val | 3.318 | -0.450 | 2.052 | 5.025 | 3.803 | -7.727\*\*\* |
|  | （2.212） | （2.664） | （2.952） | （3.260） | （4.167） | （2.254） |
| lag1use | 0.561\*\*\* | 0.335\*\*\* | 0.437\*\*\* | 0.361\*\*\* | 0.726\*\*\* | 0.302\*\* |
|  | （0.0858） | （0.0725） | （0.0953） | （0.103） | （0.120） | （0.102） |
| Constant | -2.147 | 4.154 | -0.443 | -4.646 | -3.556 | 15.85\*\*\* |
|  | （3.338） | （4.187） | （4.459） | （4.992） | （6.446） | （3.752） |
| Year Dummy | YES | YES | YES | YES | YES | YES |
| Observations | 230 | 230 | 104 | 104 | 126 | 126 |
| R-squared | 0.604 | 0.569 | 0.589 | 0.601 | 0.666 | 0.614 |
| Number of id |  | 30 |  | 19 |  | 11 |
| 注意：括号中的值为异方差稳健标准误 | | | | | | |

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

结合表6.21和表6.23的回归结果来看，在线评价信息的有用性数量对手机销售量的影响不管在OLS方法还是固定效应模型回归方法中、不管是考虑不考虑时间因素以及国产非国产的情况下回归结果都是一致的，在线评价信息的有用性数量对手机销售量的影响存在显著正向影响。评价效价对手机销量的影响在不考虑国产手机和非国产手机差别但考虑时间因素的情况下，OLS回归分析中评价效价对销量有正相关关系；在考虑国产手机和非国产手机的不同情况下，评价效价对国产手机销量有正向的促进关系。但是，评价效价对非国产手机销量的影响对于OLS回归方法和固定效应模型回归方法存在一定差异，为了进一步完善模型，在下一小节中使用广义矩估计（GMM）来进一步回归分析。

6.2.2.3 GMM模型对动态面板数据分析在线评价对销售量的影响

在一段时期内有时比较明显的是上一期手机的销量会影响这一期，网络购物用户在消费行为方面可能会存在消费追随效应。即上一期卖的多，这一期也会也很多人购买，因此需要使用动态面板效应来处理这个问题。动态面板数据模型（Dynamic Panel Data Model）是指通过在静态面板数据模型中引入滞后期的被解释变量来反映动态滞后效应的模型。

本文研究中用到的GMM模型如公式6-4所示：

*LnVolit***0**1*LnVolit*1**2 *LnValit*1**3*LnUsefulit*1*i**t**it*

（6-4），其中

*Volit*代表第i款手机第t月的在线评价数量用来模拟因变量第i款手机第t月的手机销售量，*Volit*1用来模拟因变量第i款手机第t-1月的手机销售量，*Valit*1代表第i款手机第t-1月的评价效价，*Usefulit*1代表第i款手机第t-1月在线评价的有用数，*i* 是指随手机个体变化而不随时间变化的要素，t表示随时间变化而不随个体变化的因素即时间要素，*it*表示该模型的误差项。

对于GMM模型，研究中做了异方差处理结果是稳健的，由于做了Robust检验就无法做Sargan检验，但是本研究中做了Abond检验，通过检验值显示使用GMM模型是合理的。通过对动态面板数据进行回归，得到表6.24结果。

表 6.24 评价信息对销售量的影响（动态面板数据, GMM模型）

|  | ALL | CHN Brand | Foreign Brand |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| VARIABLES | lnvol | lnvol | lnvol |
| L.lnvol | 0.303\*\*\* | 0.0372 | 0.491\*\*\* |

|  |  |  |  |
| --- | --- | --- | --- |
|  | （0.0864） | （0.0912） | （0.104） |
| lag1val | 0.0913 | 6.749\*\* | -0.553 |
| （2.779） | （3.110） | （3.947） |
| lag1useful | 0.142 | 0.292\*\*\* | 0.0202 |
| （0.0945） | （0.105） | （0.119） |
| Constant | 3.544 | -5.552 | 3.964 |
| （4.341） | （4.676） | （6.112） |
| Observations | 194 | 84 | 110 |
| YEAR DUMMY | YES | YES | YES |
| Number of id | 30 | 19 | 11 |
| Arellano-BondAR(2) | 0.07 | 0.6 | 0.0538 |
| 注意：括号中的值为异方差稳健标准误 | | | |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | | |

通过表6.24结果来看，经过GMM处理其内生性问题，在不考虑国产与非国产手机差别的情况下，上一期手机的销售量对当期有显著正向影响（**1 =0.303，

p<0.01），评价效价和评价的有用数对销售量存在正向促进关系。在考虑国产手机和非国产手机的情况下，国产手机的上一期销售量对当期销售量存在正向促进关系，但不显著，说明国产手机消费没有明显的追随现象；国产手机的滞后一期的评价效

价对当期销售量有显著正向影响（**2 =6.749, p<0.05）；国产手机的滞后一期的评价

有用数量对当期销售量有显著正向影响（**3 =0.292, p<0.01）；也就是说国产手机的销量受在线评价的影响比较大，消费者买国产手机时更注重对在线评价信息的参考。对于苹果、三星、索尼、诺基亚等非国产手机，上一期手机的销售量对当期有显著

正向影响（**1 =0.491, p<0.01），但是评价有用数对手机销售数量的影响并不显著；

可以理解为网络购物用户对非国产手机的消费存在明显追随现象，如果这个网络购物用户生活工作中周围购买非国产手机的人很多，那么这个用户很可能也去购买非国产手机，而他在购买非国产手机时对在线评价的关注度不如购买国产手机的用户。

### 6.2.3 本节小结

本节对销售量的影响因素的假设以及回归分析结果进行总结。

假设一：京东商城的手机销售数量与手机价格存在显著负向影响。当一种商品

的价格上升时，对这种商品的需求就会减少，因此只用商品的价格和需求数量的变动方向是相反的。

通过回归结果显示，在不考虑手机的国产性与非国产性特征下，手机价格对销售量的影响系数-0.597，显著性水平为0.01。国产手机价格对销售量的影响系数是

-0.665，显著水平0.01，非国产手机价格对销售量的影响系数是-0.588，显著水平0.05。由于0.588比0.665明显要小，说明购买华为、中兴、小米等国产手机品牌的网络购物用户受手机价格的影响比购买苹果、三星、诺基亚、索尼等非国产手机品牌的用户大。进一步可以理解为，国产手机价格对手机销售量的影响比非国产手机价格对销售量的影响大。

假设二：京东商城的手机销售数量与上市时间显著正相关。

本研究的实证结果显示，不管考虑还是不考虑手机是否国产的因素情况下，手机的上市时间对手机销售数量的影响都没有显著影响关系。无法根据手机的上市时间判断手机销售数量，这比较符合实际。因为现实中手机更新换代的速度非常快，研究过程中可能存在不同款式手机上市时间不一致，而且存在某些款式手机供应商某段时间内对手机的供货可能存在不足或是出于营销手段考虑故意使用饥饿营销手段等情况，导致手机上市以后销售量与上市时间没有明显关系。

假设三：京东商城的手机销售数量与有用的评价数存在显著正向相关关系。根据OLS和固定效应模型回归结果，可以明显得出的结论无论国产手机还是非

国产手机在线评价的有用数对手机销售数量存在显著正向影响关系。这是因为在线评级的有用数是网络购物用户对所看到的在线商品评价做出的比较客观实在的点评，网络购物用户认为评价对自己的购买行为决策有用才会去点击有用按钮，所以可信度比较大。在考虑手机是不是国产的因素的前提下，把前期销售量对当期销售量的影响也纳入考虑范围时，回归结果出现了比较值得关注地方，就是虽然在线评价的有用数对手机销售量的影响仍然是正向影响，但非国产手机的影响不如国产手机的影响大。这说明购买国产手机的网络购物用户对在线评价的有用数更关注，他们对于国产手机的在线评价的有用数更关注，购买行为更容易受到有用贴的影响。而非国产手机的消费者对在线评价的有用数不如国产手机的关注大，网络购物用户对评价的有用数相对国产手机不那么关注，可以解释为他们更相信非国产手机的品牌或者质量。

假设四：京东商城的手机销售数量与该款手机对应的咨询数量显著正相关。通过OLS实证分析结果显示，手机销售数量与网络购物用户的在线咨询数量存

在正向关系，但是关系并不显著。对手机进行咨询的人越多说明对该手机关注的人

越多，他们购买咨询过的手机的可能性应该越大，但是影响关系不显著，是因为任何一个网络购物用户的咨询信息都是公开的，所有人都可以看到，所以如果有潜在的购物用户看到了想问的问题已经被问过了就不会再去问同样的问题了，这也正是咨询数量与销售数量正向影响关系不显著的原因。

假设五：京东商城的手机销售数量与该款手机对应在线咨询信息评价效价有显著正相关关系。

根据实证结果综合来看，如果不考虑时间因素，评价效价对手机销售量的影响存在显著正向影响。国产手机的评价效价比非国产手机的评价效价对销售量的影响更大。由于本研究中评价效价是在线评价的平均分值。所以，评价分值的高低对国产手机销售量的影响比非国产手机的销售量影响大。

假设六：京东商城本期手机销售数量与该款手机前一期的销售数量有显著正相关关系。

如果考虑时间因素，不管国产手机还是非国产手机，前一期的在线评价效价对本期手机销售量的影响都不显著，但是国产手机前一期评价效价对本期销售量存在正向相关关系。

通过动态固定效应模型回归分析，得到一个值得关注的结果，研究中的手机的总体样本本期销售数量与前一期的销售数量存在显著正向关系，影响系数是0.303

（p=0.01）。考虑了手机的国产与非国产因素时，非国产手机前一期销售量对本期销售量有显著正向影响，前一期销售量对本期销售量的影响系数是0.491（p=0.01），而国产手机前一期销售量对本期销售量的影响不如非国产手机的影响显著。这个结果可以理解为，非国产手机消费者购买手机时从众效应比国产手机的从众效应明显。

假设七：本期手机销售数量与前一期的评价有用数存在显著正向关系。

回归结果显示，手机总体样本的前一期的评价有用数对本期销售量的影响不显著，但前一期的评价有用数对国产手机的销售量存在显著影响，前一期评价有用数对国产手机销售量的影响系数是0.292（p=0.01），而非国产手机的前一期的评价效价与前一期的评价有用数对本期的销售数量影响不显著。

假设八：本期手机销售数量与前一期的评价效价存在显著正向关系。

根据GMM回归结果，总体样本前一期评价效价对本期手机销售量的影响不显著，但前一期的评价效价对国产手机本期的销售量存在显著影响，其影响系数为

6.749（p=0.05），但前一期的评价效价对非国产手机本期销售量没有显著影响。综合假设七和假设八来看，前一期由于前一期评价效价和评价有用数都属于评

价信息，属于网络口碑范畴，所以可以针对国产手机与非国产手机销售量受在线评

价信息的口碑影响做个结论，就是国产手机消费者的消费行为受前一期的评价信息影响比非国产手机消费者显著，即口碑效应对国产手机销售量影响更大。

## 6.3 本章小结

本章在第5章基本数据分析的基础上对在线评价信息的效应进行了进一步实证分析。

第一节中对在线评价信息和在线咨询信息进行因子分析，找出网络购物用户对商品进行在线评价和咨询时到底关注哪些焦点，为第二节和第三节的分析奠定了基础。第二节对网络购物用户的购买与评价行为时间差进行了分析，寻找这种时间差的影响因素。第三节中通过在线评价信息包含的多种因素对销售量的影响关系分析。

1．对在线评价信息和咨询信息的因子分析

通过对在线评价信息包含的“外观漂亮”、“性价比高”、“系统流畅”、“功能齐全”、“照相分辨率高”、“通话质量与音质”、“反应快”、“屏幕大”、“支持国产”和

“信号稳定”十个要素进行因子分析，得到了三个新的不可观测因子：即“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子，这三个因子对原来的十个要素的解释贡献率达到95.69%，而且三个因子完全正交。

根据表6.6的因子得分系数矩阵中得到的因子得分系数和原始变量的标准化值就可以计算不可观测的“手机基本功能”因子、“屏幕与国产”因子和“信号稳定性”三个因子的得分。

“手机基本功能”因子=0.179\*“外观漂亮”+0.033\*“性价比高”+0.173\*“系统流畅”+0.156\*“功能齐全”+0.150\*“照相分辨率高”+0.318\*“通话质量与音质”

+0.157\*“反应快”-0.146\*“屏幕大”-0.166\*“支持国产”-0.030\*“信号稳定”

“屏幕与国产”因子=-0.022\*“外观漂亮”+0.192\*“性价比高”-0.040\*“系统流畅”+0.000\*“功能齐全”+0.030\*“照相分辨率高”-0.346\*“通话质量与音质”

-0.007\*“反应快”+0.515\*“屏幕大”+0.466\*“支持国产”-0.069\*“信号稳定”

“信号稳定性”因子=-0.089\*“外观漂亮”+0.126\*“性价比高”+0.059\*“系统流畅”+0.023\*“功能齐全”-0.085\*“照相分辨率高”-0.077\*“通话质量与音质”

+0.049\*“反应快”-0.314\*“屏幕大”+0.213\*“支持国产”+0.797\*“信号稳定”通过对在线评价信息的因子分子，研究中发现网络购物用户对手机是不是国产

的特征有特别的关注，于是在第三节对手机销售量的影响因素中，把国产手机和非国产手机区分开来做回归分析。

研究中还对咨询信息包含的“价格方面”、“支付方面”、“物流方面”、“售后服务方面”、“卡支持方面”、“颜色等外部特征”、“功能方面特征”、“配件方面”、“软件及操作系统方面”、“使用区域方面”以及“质量方面”十一个要素进行了两次因子分析。发现把“价格方面”、“支付方面”、“物流服务”和“售后服务”四个方面的要素不纳入因子分析的因素，而把“卡支持方面”、“软件及操作系统方面”、“使用区域方面”、“质量方面”、“颜色等外部特征”、“功能方面特征”和“配件方面”七个要素进行因子分析以后得到的两个因子更便于解释和研究。所以，在线咨询信息的关注点最终总结为“价格方面”、“支付方面”、“物流方面”、“售后服务方面”、“产品性能”和“产品规格”六个关注焦点。

根据表6.18第二次因子分析的因子得分系数矩阵中得到的因子得分系数和原始变量的标准化值就可以计算每个观测的“产品性能”因子和“产品规格”因子的得分。

“产品性能”因子=0.225\*“卡支持方面”+ 0.287\*“软件及操作系统方面”+0.377\*“使用区域方面”+0.372\*“质量方面”+0.023\*“颜色等外部特征”-0.176\*“功能方面特征”-0.222\*“配件方面”

“产品规格”因子=-0.006\*“卡支持方面”-0.044\*“软件及操作系统方面”-0.204\*“使用区域方面”-0.169\*“质量方面”+0.270\*“颜色等外部特征”+0.456\*“功能方面特征”+0.504\*“配件方面”

2.通过因子分析的研究发现，手机是不是国产的这一特征很受网络购物用户关注，因此这一特征可能影响网络购物用户购买手机的行为，由此这一特征对手机销售量的影响比较明显。在第二节，本文对在线评价信息的多项因素对手机销售量的影响通过总体样本、国产手机样本和非国产手机样本都进行了回归分析研究。

在研究中不考虑前一期影响因素对本期销售量影响的前提下提出了手机价格、上市时间、有用的评价数量、咨询数、手机评价效价对手机销售量影响的五个假设，并通过OLS回归方法对横截面数据进行分析对五个假设进行了验证，验证结果如表

## 6.25 所示。

表 6.25 OLS回归分析结果表

| 回归分析方法 | OLS 回归方法分析 | | |
| --- | --- | --- | --- |
| 影响销售量因素的五个假设 | 总体样本 | 国产手机 | 非国产手机 |
| 假设一：京东商城的手机销售数量与  手机价格存在显著负向影响。 | 显著  支持 | 显著  支持 | 显著  支持 |

|  |  |  |  |
| --- | --- | --- | --- |
| 假设二：京东商城的手机销售数量与  上市时间显著正相关。 | 不显著  支持 | 不显著  不支持 | 不显著  不支持 |
| 假设三：京东商城的手机销售数量与有用的评价数存在显著正相关关系。 | 显著支持 | 显著支持 | 显著支持 |
| 假设四：京东商城的手机销售数量与  该款手机对应的咨询数量显著正相关。 | 不显著  支持 | 不显著  支持 | 不显著  支持 |
| 假设五：京东商城的手机销售数量与该款手机对应在线咨询信息评价效价  有显著正相关关系。 | 显著支持 | 不显著支持 | 显著支持 |

通过来自京东商城的信息分析得到的结果是，对于京东商城来说手机价格、在线评价有用数、以及咨询信息数量对手机销售量有显著影响，但是手机上市时间的长短对手机销售量的影响不显著，咨询信息数量能够促进手机销售，但由于在线咨询信息的公开性以及长久保存可视性等特点，咨询信息数量与手机销售量之间的关系不显著。

考虑了前一期影响因素对本期销售量影响的情况下进行进一步分析，结果如表

## 6.26 所示。

表 6.26 固定效应模型和GMM模型回归分析结果表

| 回归分析方  法 | 固定效应模型回归方法分析 | | | GMM 回归方法分析 | | |
| --- | --- | --- | --- | --- | --- | --- |
| 结论 | 总体样  本 | 国产手  机 | 非国产手  机 | 总体样  本 | 国产手  机 | 非国产手  机 |
| 假设六：本期手机销售数 量与前一期 的销售数量  显著正相关。 | - | - | - | 显著支持 | 不显著支持 | 显著支持 |
| 假设七：本期手机销售数量与前一期的评价有用  数存在显著 | 显著支持 | 显著支持 | 显著支持 | 不显著支持 | 显著支持 | 不显著支持 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 正相关关系。 |  |  |  |  |  |  |
| 假设八：本期手机销售数 量与前一期 的评价效价  显著正相关。 | 不显著支持 | 不显著支持 | 显著不支持 | 显著不支持 | 显著支持 | 不显著不支持 |

通过考虑时间因素特征结合表6.26的总结，国产手机消费者的消费行为受前一期的评价信息影响比非国产手机消费者显著，即口碑效应对国产手机销售量影响更大。非国产手机消费者购买手机时从众效应比国产手机的从众效应明显。

# **第7 章** **B2C**网络购物用户商品评价的效应研究的总结及前

**景预测**

在线评价信息对网络商城竞争力以及网络购物用户购买行为的影响越来越大的现实背景下，在线评价信息反映网络购物用户怎样的网络行为规律，怎样影响网络商家的销售业绩，消费者尤其是网络商家怎样有效利用在线评价这种大数据信息这种新的信息媒体成为迫切需要回答的问题。本论文从获取在线评价一手信息大数据开始出发，对在线评价信息所反映的网络购物用户的网络行为规律进行分析，然后对在线评价信息对销售量的影响进行了实证分析。本章作为论文的总结，主要对论文的主要研究内容、结论、实践意义、研究中的不足和局限性以及未来的研究方向进行总结。

## 7.1 本文的研究结论

本文利用网络购物消费者二十一万多次真实网络购买手机的行为过程形成的实际一手在线评价信息数据结合咨询数据进行分析得出的网络购物用户的网络行为规律具有现实可信度。对于网络销售商家和网络购物用户都有实际应用价值。主要结论如下：

1. 网络购物用户的数量和消费能力与网络购物消费者所在省（市）的经济发展水平有正向相关关系。一个省（市）的经济发展水平越高，该省（市）的网络购物消费者数量越多，网络购物消费能力越强；反之，一个省（市）的经济发展水平越低，则该省（市）的网络购物消费者数量越少，网络购物消费能力越弱。

2. 每周“星期二”和“星期三”是网络购物用户进行网络购物相关活动的高峰日期，每周“星期六”和“星期日”是网络购物用户进行网络购物相关活动的低峰日期。这个结果与预期中每逢节假日是消费者的消费活动高峰期有所差异。其中，每周“星期二”下订单购买手机商品的人数占比最多，每周“星期三”进行咨询和商品评论活动人数占比最多。而每周“星期六”和“星期日”进行咨询、下订单买手机以及购买以后进行商品评价的人数量占比相对较少。

3. 每天的“下午”和“晚上”时间段是网络购物用户进行网络购物相关活动的高峰时间段，“凌晨”时间段是网络购物用户进行网络购物活动的低峰时间段。其中，在“下午”时间进行咨询和在线评价的人占比最多，利用“晚上”时间进行咨询和

在线评价的人占比次之，“上午”进行咨询和在线评价的人数占比较少，“凌晨”时间段进行咨询和在线评价的人占比最少。

4. 对于网络咨询行为与购买行为的关系结论是，进行了网络咨询的用户最终不一定会购买咨询过的商品，而有网络购买行为的用户很可能并没有进行网络咨询活动，这与网络咨询信息的公开性、长久性以及随时可获得性等因素有关系。其中，网络购物用户在咨询活动以后大约会有15.52%的咨询者最终会购买手机产品，有

84.48%的咨询者不会咨询完以后不会进行购买活动。很多网络购物用户购买行为发生之前并不会对目标商品发起在线咨询行为，只有1.7%的手机购买者有买前对关注商品进行咨询的行为，98.3%的手机购买者并没有提前进行咨询活动就对商品进行购买。

5. 虽然京东商城对购物后进行在线评价的用户有奖励机制，但购买手机的网络购物用户并不会在购买相应产品后立刻对所购买的商品进行评价，京东商城手机购物用户平均在购买相应手机产品19.2天以后对相关产品进行评价，这种时间间隔能够显示网络购物用户对评价行为具有理性、只有真正体验过商品以后才能准确发表有用的在线评价。

下订单购买以后通过需要1天以上时间才能收到货物，然后用户需要办理手机卡对手机进行使用，通过充电、放电、打电话、发短信等多种功能多天的体验以后，才能对相应产品的相关功能进行评价，所以购买与评价之间十几天的间隔期间是一个合理范围。

6. “购买评价日期差”与网络购物用户本身的特征和行为习惯有关系，与评价信息包含的“评论星级”、“评价回复数”以及“评价有用数”等影响因素之间没有明显关系。

研究中把“购买评价日期差”作为被解释变量，把“用户等级”、“评论星级”、

“评价回复数”以及“评价有用数”作为解释变量，寻找“用户等级”、“评论星级”、“评价回复数”以及“评价有用数”与“购买评价日期差”变量之间的解释与被解释关系。通过多元逐步回归分析发现，“购买评价日期差”只受与网络购物用户特征有关的“用户等级”影响。

7.对在线评价的内容进行因子分析研究发现，网络购物用户在对手机商品进行使用以后做出的评价主要集中在手机的基本功能、手机品牌的国产性与非国产性以及手机信号稳定性三个方面。在线咨询信息因子分析结果显示，潜在的网络手机消费者主要关注手机价格、京东商城购物支付、物流、京东商城售后服务、手机产品性能和产品规格六个方面。

8.京东商城手机销售数量与商家制定的手机销售价格存在显著负相关关系，与在线评价信息的有用数和评价效价存在显著正相关关系。手机网络销售数量与手机的上市时间长短没有明显关系，与在线咨询数量没有显著关系。

9.国产品牌手机本期的网络销售数量与前一期在线评价信息的有用数和评价效价之间的正向关系比非国产品牌显著，即国产品牌手机消费者受口碑效应的影响比非国产品牌手机明显。

10.非国产手机本期网络销售数量与前一期的销售数量之间的正向关系比国产品牌手机显著，即非国产品牌手机的消费者受从众效应的影响比国产品牌手机明显。

## 7.2 根据研究结果给网络商家提出合理的建议

本节根据数据综合分析给网络商城卖家提出的建议如下：

1．网络商城应该根据消费者的网络行为习惯合理安排工作人员数量

（1）. 每周一、二、三下订单购买商品的用户比较多，商家应该安排更多的工作人员处理订单，每周四和周五卖家应该安排周一至周三大约7/8数量的人员进行订单处理，而周六和周日大约只需要周一至周三3/4数量的工作人员处理订单。

（2）. 由于大约15.52%的咨询者会在咨询后购买商品，所以卖家必须重视咨询信息的处理，对咨询者的咨询进行耐心答复，促进咨询者购买商品。每周二和周三对商品进行咨询的用户比较多，应该安排合理数量工作人员尽快进行咨询信息处理，每周一、四和五每天大约安排7/8数量的人员进行咨询答复，而周六和周日只需要周一至周三3/4数量的工作人员咨询答复。

（3）. 每天中午12点至晚上24点是网络购物用户进行网络咨询和评论的高峰

时间段，网络卖家应该配置70%左右的人力资源在中午12点至晚上24点，配置25%

左右的资源在上午6点至12点，配置5%左右的资源在凌晨0点至6点。

2．网络商城应该根据消费者的区域分布合理配置资源

（1）. 从网络购物消费者的身份分布数据来看，北京、广东和上海三省市的所有会员数量占总数的52.38%，三省市年消费过百万的顶级消费者比例占总体的

94.42%，所以卖家应该把50%以上的资源分布在北京、广东和上海三省市，重点维护北上广三省市的高端消费者，建立合理的物流体系，提高服务质量。

（2）. 为了增加网络销售额度，卖家应该重点在每周一至周三增加优惠活动，增加网络购物用户的购买热情。

3．网络销售卖家对所售商品的介绍应该符合消费者的关注焦点

网络卖家在对商品进行介绍时不能仅凭自己的感觉对商品进行介绍，应该根据消费者以及潜在消费者在网上做出咨询和评价行为产生的信息和数据进行充分分析，然后及时调整宣传重点，让宣传和介绍符合消费者的信息需求。

对于所销售的商品介绍时，突出特色，让商品介绍能够最大程度的吸引消费者。对于消费者的常见问题设立常见问题显示区，及时消除消费者对于产品消费的疑问，提高在线回答问题工作人员的效率。

4．为提高销售数量对消费者不同的评价行为制定不同的激励政策

由于销售数量与评价效价和评价有用数存在正向促进作用，所以网络商城对进行在线评价的消费者进行了积分奖励，积分可以换成现金在下次消费行为中使用。而这种奖励机制也可以对获得有用数多的评价进行二次奖励，这样可以促进在线评价者发表出对其他消费者更有用的在线评价信息。

由于手机产品更新换代比较快，很多消费者的消费受到口碑效应和从众效应影响比较明显，目前网络商城对于口碑效应的使用已经非常成熟，突出所售商品的好评数、好评率或平均分值等，对于受从众效应影响明显的产品应该好好利用从众效应显示出上一期（比如一个周或一个月）的销售数量，进一步利用网络夸大从众效应，有利于提高商品销售数量。

## 7.3 本文的创新之处

1.探索性的首次把网络购物用户的在线咨询信息纳入销售量影响因素的研究范围。把在线咨询信息与在线评价信息相结合，来获取更多的信息量，比如得到了咨询购买占比等新的数据。

2.使用因子分析研究网络购物用户购物之前对网络商城以及商品的关注焦点以及购物以后进行在线评价时对所购买商品的关注焦点。有利于网络商城在销售商品时对商品做出符合消费者预期的商品介绍和宣传。

3.对样本数据进行横截面数据模型和面板数据模型相结合的分析，得到对商品销售数量影响因素更为详细的、可以对比的分析结果。

## 7.4 本文的不足之处以及进一步研究的方向

研究中由于作者精力和时间有限，数据获取难度大、网络商城经常升级导致信息格式变化等各种原因，导致研究中存在一些不足之处。

1.样本数量局限性

由于时间限制、数据抓取困难、抓取周期长等原因，在样本数据采集上，虽然抓取了二十多万条数据，但是根据大数据思想，样本数量和信息数量越多分析结果越准确，由于受信息抓取时间限制研究中只是抓取了京东商城上的30款手机的在线评价信息数据，因此可能是的研究结论有一定的局限性，分析结果有待于进一步验证。

2.实证商品研究对象类型局限

本文研究中选用京东商城3C产品中的手机产品作为研究对象，3C产品还包括笔记本电脑、iPad等其它产品，本研究没有进一步抓取其它研究对象的相关数据，不同产品的数据对于研究结果可能会存在一定差异。

3.在评价信息对商品销售数量影响的研究中，由于真实的销售数量难以获取，手机每期的销售数量用在线评价信息的数量来模拟，由于存在消费者购买了产品但是没有进行评价的情况，所以研究中的销售数量与真实销售数量存在差异，导致研究结果可能存在不够准确的风险。

4.研究时间以及京东商城信息显示方面限制无法研究产品生命周期因素

由于整个研究时间有限，而且京东商城对于下市商品不再显示等方面影响，导致研究中无法抓取研究商品的整个生命周期中的销售量、评价信息和咨询信息，所以研究中无法对商品处于不同生命周期阶段的销售影响因素的研究。

由于各种限制原因导致研究存在一定不足之处，所以研究工作在未来还有很多可以发展之处。

1.扩大样本容量，可以使研究研究更准确。如果能够合法获取网络商城数据库中的所有样本信息，可以对消费者网络购物行为进行更为准确的研究。

2.可以抓取不同产品的样本进行对比研究，比如把笔记本、Pad等产品纳入研究范围，对不同产品的消费者行为进行研究，并对结果进行对比可以进一步归纳总结消费者的网络购物行为规律。

3.对于产品生命周期不同时间段的销售量影响因素进行研究。如果有机会在

B2C电子商务领域工作，利用工作中的相关产品的销售数据可以对不同产品所处的不同生命周期时间段的销售量的影响因素进行研究，有利于提高商品整个生命周期中的总销售数量。

参考文献

[1] 张海燕． 基于分词的中文文本自动分类研究与实现[Ｄ]． 长沙: 湖南大学, 2002.

[2] 娄德成． 基于ＮＬＰ技术的中文网络评论观点抽取方法的研究[Ｄ]． 上海: 上海交通大学, 2007.

[3] 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年第13卷第8期, 78- 96页．

[4] 施国良, 石桥峰. 基于文本挖掘的不同购物网站商品评论一致性研究[J]. 现代图书情报技术, 2011年第12期, 64-68页.

[5] 李健. 在线商品评论对产品销量影响研究[J]. 现代情报, 2012年第1期, 164-167页.

[6] 王琦, 唐世渭, 杨冬青, 等. 基于DOM的网页主题信息自动提取[J]. 计算机研究与发展, 2004年第41卷第10期, 1786-1792页.

[7] CNNIC. 2012年中国网民消费行为调查报告.

[8] CNNIC. 第31次中国互联网络发展状况统计报告.

[9] 埃弗瑞姆. 特伯恩（Efrain Turban)、戴维. 金（David King)著, 电子商务导论（Introduction of Electronic Business), 人民大学出版社, 2010年第一版.

[10] 罗敖生. 基于ASP实现网上数据的自动抓取[J]. 江苏广播电视大学学报, 2002年第13卷第6期, 59-61页.

[11] 朱晓静, 方磊坤, 李太君. 基于P2P网络的信息检索[J]. 现代电子技术, 2007年第4期, 120-122页.

[12] 顾潇华, 郭军城. 网页超链抓取及自动分类技术实现[J]. 河北大学学报（自然科学版）, 2007年第1期, 99-102页.

[13] YU D, SHEIKHOLESLAMI G, ZHANGA. Findout: finding outliers in large datasets[ J ]. Knowledge and Information Systems, 2002年第4卷第4期, 387-412页.

[14] 张博, 蔡皖东. 面向主题的网络蜘蛛技术研究及系统实现[J]. 微电子学与计算机2009年第5期, 52-55页.

[15] 袁浩, 黄烟波. 网页标题分析对主题爬虫的改进[J]. 计算机技术与发展, 2009年第6期, 22-28页.

[16] 白鹤, 汤迪斌, 王劲林. 分布式多主题网络爬虫系统的研究与实现[J]. 计算

机工程, 2009年第35卷第19期, 13-19页.

[17] 郭会斌. 营销口碑的产生路径与创造研究[J]. 经济与管理. 2005年第19卷第l1期, 44-46页.

[18] 阙克儒. 网络匿名性、企业形象与关系品质对网络口碑影响之研究--以线上游戏为例[D]. 国立中兴大学, 2004.

[19] 黄英, 朱顺德. 二十一世纪的口碑营销及其在中国的发展潜力[J]. 管理前沿. 2003年第6期, 33-36页.

[20] 郝媛媛. 在线评论对消费者感知与购买行为影响的实证研究[D]. 黑龙江: 哈尔滨工业大学, 2010.

[21] 章晶晶. 网络环境下口碑再传播意愿的影响因素研究[D]. 浙江: 浙江大学, 2007.

[22] 卢向华、冯越. 网络口碑的价值——基于在线餐馆点评的实证研究[J]. 管理世界, 2009年第7期, 126-132, 171页.

[23] 周晶晶. 在线客户评论对消费者购买决策的影响力研究[D]. 浙江: 浙江大学, 2010.

[24] 张素华. 基于网络外部性与从众效应的锁定策略研究[D]. 福建: 厦门大学, 2008.

[25] 殷国鹏. 消费者认为怎样的在线评论更有用[J]. 管理世界, 2012年第12期, 115-124页.

[26] 姜巍, 张莉, 戴翼, 蒋竞和王刚. 面向用户需求获取的在线评论有用性分析[J]. 计算机学报, 2013年第1期, 121-131页.

[27] 杨爽, 徐畅. 在线产品评论有用性的影响机制[J].求索, 2013年第3期, 249-251页.

[28] 郝媛媛、邹鹏、李一军、叶强. 基于电影面板数据的在线评论情感倾向对销售收入影响的实证研究[J]. 管理评论, 2009年第10期, 95-103页.

[29] 李宏, 喻葵, 夏景波. 负面在线评论对消费者网络购买决策的影响一个实验研究[J]. 情报杂志, 2011年第5期, 202-207页.

[30] 孟美任, 丁晟春. 在线中文商品评论可信度研究[J]. 现代图书情报技术, 2013年第9期, 60-66页.

[31] 李念武, 岳蓉. 网络口碑可信度及其对购买行为之影响的实证研究[J]. 图书情报工作, 2009第53卷第22期, 133-137页.

[32] 施国良石桥峰. 基于文本挖掘的不同购物网站商品评论一致性研究[J]. 现

代图书情报技术,2011年第12期, 64-68页.

[33] 潘勇, 赵军民. 基于顾客满意度的B2C电子商务网站评价[J]. 现代情报, 2008第5期, 220-223页.

[34] 李思曼, 王宇航, 李亚平. 基于顾客满意的消费者网上购物影响因素分析[J]. 商业研究, 2009第01期, 203-206页.

[35] 饶曦. 网络购物中的消费者购买行为分析[J]. 现代企业, 2010第6期, 50- 51, 57页.

[36] 王秀丽田祯祎. 网民网络购物行为调查[J]. 图书情报工作, 2011第6期, 20-23, 95页.

[37] 马庆国. 管理统计: 数据获取、统计原理、SPSS工具应用与研究, 科学出版社, 2002.08.01.

[38] 陈胜可. SPSS统计分析从入门到精通, 清华大学出版社, 2013.05, 第二版.

[39] 吴明隆, 涂金堂. SPSS与统计应用分析, 东北财经大学出版社, 2012年9月第一版.

[40] 杨维忠, 张甜. SPSS统计分析与行业应用案例详解, 清华大学出版社, 2013 年3月第二版.

[41] 朱建平, 方匡南, 章贵军. SPSS统计分析与应用, 首都经济贸易大学出版社, 2013年2月第一版.

[42] Zhang Z． Weighing stars: Aggregating online product reviews for intelligent e—commerce applications[J]． IEEE Intelligent Systems, Vol. 23(5): PP42—49．, 2008.

[43] Tsai F S, Kwee A T. Database Optimization for Novelty Mining of Business Blogs [J]. Expert Systems With Applications, Vol. 38(9): PP 11040-11047., 2011.

[44] Liu Y, Huang X J, An A J, et a1． Modeling and predicting the helpfulness of online reviews[C]. Procedings of the 8th IEEE International Conference on Data Mining, Washington． DC, USA: IEEE Computer Society: PP443—452．, 2008.

[45] 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 Machincry: PP303—310．, 2007.

[46] Otterbacher J． ”Helpfulness" in online communities: A measure of message

Quality[C]. Proceedings of the 27 International Conference on Human Factors in Computing Systems, Boston, MA, USA: Association for Computing Machinery: PP955—964．, 2009.

[47] 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, Vol. 36(3): PP6527—6535., 2009.

[48] 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, Vol. 26(3): PP1—34．, 2008.

[49] Abbasi A, Chen H C, Thoms S, et a1． Affect analysis of web forums and blogs using correlation ensembles[J]． IEEE Transactions on Knowledge and Data Engineering, Vol. 20(9): PP1168—1180., 2008.

[50] Kim S M, Pantel P, Chklovski T, et a1． Automatically assessing review helpfulness[C] Proceedings of the Conference Empirical Methods in Natual Language Processing, Morristown, NJ, USA: Association for Computational Linguistics: PP423—430．, 2006.

[51] Chen C C, Tseng Y D． Quality evaluation of product reviews using an information quality framework[J]． Decision Support Systems, Vol. 50(4): PP755—768．, 2011.

[52] Li N, Wu D D. Using Text Mining and Sentiment Analysis for On Line Forums Hotspot Detection and Forecast [J]. Decision Support Systems, Vol. 48(2): PP354-368., 2010.

[53] Chang C W, Lin C T, Wang L Q. Mining the Text Information to Optimizing the Customer Relationship Management[J]. Expert Systems with Applications, Vol. 36(2): PP1433-1443., 2009.

[54] Xu K, Liao S S, Li J, et al. Mining Comparative Opinions from Customer Reviews for Competitive Intelligence[J]. Decision Support Systems, Vol. 50(4): PP743-754., 2010.

[55] 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, PP841—847．, 2004.

[56] Hu N, Pavlou P A, Zhang J． Overcoming the J-shaped distribution of product reviews [J]． Communications of the ACM, Vol. 52(10): PP144—147．, 2009.

[57] 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, Vol. 21(3): PP111—147．, 2004.

[58] Jiang Z H, Benbasat I． Investigating the irfluence of the functional mechanisms of online product presentations[J]． Information Systems Research, Vol. 18(4): PP454—470．, 2007.

[59] Pavlou P A, Fygenson M． Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior[J]． MIS Quarterly, Vol. 30(1): PP115—143．, 2006.

[60] Pavlou P A, Liang H G, Xue Y J． Understanding and mitigating uncertainty in online exchange relationships: A principal—agent perspective[J]． MIS Quarterly, Vol. 31(1): PP105—136．, 2007.

[61] Kotler P, Keller K． Marketing Management[M]．12 ed．, Upper saddle River, NJ: Prentice—Hall, 2005．

[62] Kohli R, Devar S, Mahmood M A． Understanding determinants of online consumer satisfaction: A decision process perspective[J]． Journal of Management Information Systems, Vol. 21(1): PP115—135．, 2004.

[63] Mudambi S M． Schuf D． What makes a helpful online reviewAstudyofcustomerreviewsonAmazon．Com, MISQuarterly, Vol. 34(1): PP185—200．, 2010.

[64] Hevner A R, March S T, Park J, et a1． Design science in information systems research[J]． MIS Quarterly, Vol. 28(1): PP75—105．, 2004.

[65] Walls J G, Widmeyer G R, Sawy A E． Building an information system design theory for vigilant EIS[J]． Information Systems Research, Vol. 3(1): PP36—59．, 1992.

[66] March S T, Simith G F． Design and natural—science research on information technology[J]． Decision Support System, Vol. 15(4): PP251—266．, 1995.

[67] March S T, Storey V C． Design science in the information systems deccipline: An introduction to the special issue on design science

Research[J]．MIS Quarterly, 32(4): PP725—730．, 2008.

[68] Stigle G J． The econnomics of information[J]． Journal of Political Economy, Vol. 69(3): PP213—225．, 1961.

[69] Chen Y B, Xie J H． Online consumer review: Word-of-mouth as a news element of marketing communication mix[J]． Management Science, Vol. 54(3): PP477—491．, 2008.

[70] Lee T, Bradlow E． Automatic Construction of Conjoint Attributes and Levels from Online Customer Reviews[R]． Florida Decision and Information Sciences Workshop, 2007.

[71] 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 Linguistics: PP334-342., 2007.

[72] Emanuel Rosen, Roger E. Scholl. The Anatomy of Buzz: How to Create Word-of-Mouth Marketing[M]. New York: Random House: PP44-47., 2000.

[73] Blackwell, Roged D., Paul W. Miniard and James F. Engle. Consumer Behavior[M]. NewYork: Harcount, Inc, 2001.9.

[74] Christiansen T., Tax S. S. Measuring word of mouth: the questions of who and when[J]. JournalofMarketingCommunications. 6(3): 185-199., 2000.

[75] Dellaroeas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms[J]. Management Science. Vol. 49(10): PP1407-1424., 2003.

[76] Hennig Thurau, K. P. Gwinner, G. Walsh. D. D. Gremler. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet[J]. JournalofInteractiveMarketing. Vol. 18(1): PP38-52. 2004.

[77] Datta, P. R., Chowdhury, D. N. and Chakraborty, B. R. Viral marketing: new form of word-of-mouth through Internet [J]. The Business Review, Vol. 3(2): PP69-75., 2005.

[78] Katz, E., Lazarsfeld P. F. Personal Influence: The Part Played by People in the Flow of Mass Communication[M]. New York: The Free Press, 1964.

[79] Leibenstein, H. Bandwagon, snob, and Veblen. Effeets in the Theory of

Consumers' Demand [J]. The Quarterly Journal of Eeonomies, Vol.64(2): PP183-207., 1950.

[80] Lascu D-N., Zinkhan G. Consumer Conformity: Review and Applications for Marketing Theory and Practice[J]. Jounal of Marketing Theory &Practie. Vol. 7(3): PPl-12., 1999.

[81] Banerjee, A. V. A Simple Model of Herd Behavior[J]. Quarterly Jounal of Economics, Vol. 107(3): PP797一817., 1992.

[82] Soo-Min Kim, Patrick Pantel, Tim Chklovski and Marco Pennacchiotti. Automatically Assessing Review Helpfulness [C]. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP: PP423-430., 2006.

[83] Yuanyuan Hao, Qiang Ye, YiJun Li and Zhuo Cheng. How does the Valence of Online Consumer Reviews Matter in Consumer Decision MakingDifferencesbetweenSearchGoodsandExperienceGoods[C]. Proceedingsofthe43rdHawaiiInternationalConferenceonSystemSciences, 2010.

[84] Do-Hyung Park and Sara Kim. The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews [J]. Electronic Commerce Research and Applications, (Vol. 7): PP399–410., 2008.

[85] Jumin Leea, Park D H, Han I. The effect of negative online consumer reviews on product attitude: An information processing view[J]. Electronic Commerce Research and Applications, (Vol. 7): PP341-352., 2008 .

[86] Wenjing Duana, Bin Gub, Andrew B. Whinston. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry[J]. Journal of Retailing, (Vol. 84): PP233-242., 2008 .

[87] Garrett P. Sonnier, Leigh McAlister. A Dynamic Model of the Effect of Online Communications on Firm Sales. Marketing Science, Vol.30, No.4: PP702- 716., 2011.

[88] Judith A. Chevalier and Dina Mayzlin. The Effect of Word of Mouth on Sales: Online Book Reviews[J]. Journal of Marketing Research, Vol.43, No. 3: PP345-354., 2006.

[89] Qiang Ye, Rob Law, Bin Gu. The impact of online user reviews on hotel

Room sales[J]. International Journal of Hospitality Management, Vol.28: PP180-182., 2009.

[90] Xia Liu, Mengqiao He, Fang Gao and Peihong Xie. An empirical study of online shopping customer satisfaction in China - a holistic perspective[J]. International Journal of Retail& Distribution Management, Vol.36 No. 11: PP919-940., 2008.

[91] CLEMONS E K, GAO Guodong, Hitt L M． When Online Reviews Meet Hyper Differentiation: A Study of the Craft Beer Industry[J]． Journal of Management Information Systerms, Vol. 23(2): PP149—171．, 2006.

[92] Hal R. Varian. Intermediate Microeconomics: A Modern Approach. 2003, W. W. Norton& Company, Inc. Sixth Edition.

[93] Mudambi, S. M. and Schuff. D. What Makes a Helpful Online ReviewAStudyofCustomerReviewsonAmazon. com[J]. MISQuarterly. Vol. 34: PP185-200., 2010.

[94] Pan, Y. and Zhang, J. Q. Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews[J]. Journal of Retailing, Vol. 87: PP598~612., 2011.

[95] Stephen Litvin, Ronald Goldsmith, Bing Pan. Electronic word-of-mouth in hospitality and tourism management[J]. Tourism Management, (29): PP458-468., 2008.

[96] Chrysanthos Dellarocas. The digitization of word of mouth promise and challenges of online feedback mechanisms [J]. Management Science, Vol.49, No. 10: PP1407-1424., 2003.

[97] Wenjing Duan Bin Gu Andrew B. Whionston. The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry [J]. Journal of Retailing, Vol. 84(2): PP233-242., 2008.

[98] Christy M. K. Cheung, Matthew K. O Lee and Neil Rabjohn. The impact of electronic word-of-mouth-The adoption of online opinions in online customer communities[J]. Internet Research, Vol.18, No.3: PP229-247., 2008.

[99] Yi-Fen Chen. Herd behavior in purchasing books online[J]. Computers in Human Behavior, (24): PP1977-1992., 2008.

[100] Steven Jones, Darrell Lee, Edward Weis. Herding and feedback trading

By different types of institutions and the effects on stock prices[C], Financial Management Association Meetings, 1999.

[101] Jen-Hung Huang. Herding in online product choice [J]. Psychology & Marketing, Vol. 23(5): PP413-428., 2006.

附录 A

在线咨询词频统计语料库

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **手机特征** | | | | | | | | | | |
| **价格方面** | **支付方面** | **物流方面** | **售后服务** | **颜色等**  **外部特征** | **手机卡支持** | **功能** | **配件** | **软件操作系统** | **质量** | **使用区域** |
| 价格 | 付款 | 发货 | 售后  服务 | 颜色 | 电信 | 网络 | 充电  器 | 操作系  统 | 行  货 | 美国 |
| 价钱 | 退款 | 送货 | 退换 | 黑色 | 中国电信 | 上网 | 耳机 | 存储空间 | 水货 | 英国 |
| 降价 | 刷卡 | 提货 | 退货 | 白色 | 移动 | 网卡 | 充电插头 | 内存 | 标  准机 | 国外 |
| 报价 | 分期 | 有  货 | 退款 | 贴 膜 | 中国移动 | 网页 | 电源  插头 | 软件 | 假  货 | 韩国 |
| 便宜 | 发票 | 供货 | 停产 | 机身 | 联通 | 邮件 | 插头 | 激活 | 保  修 | 非洲 |
| 促销 | 开发  票 | 什么  时候 | 保修 | 机壳 | 中国联通 | 游戏 | 贴 膜 | 越狱 | 新  机 | 香港 |
| 优惠 | 支付 | 现货 | 保修  期 | 包装 | 合约 | 视频 | 大 卡 | 版本 | 翻  新 |  |
| 限量 | 信用  卡 | 预订 | 保质  期 | 瑕疵 | 定制 | 无线 | 小 卡 | 下载 | 正  规 |  |
| 官网  贵 | 储蓄  卡 | 到货 | 保障 | 外观 | 电信卡 | 解锁 | 配件 | 升级 | 正  品 |  |
| 花费 |  | 快递 | 验货 | LOGO | 电信号 | 带锁 | 剪 卡 | IOS | 次  品 |  |
| 调价 |  | 邮寄 | 维修 | 划痕 | 话费 | 码锁 | 充电 | WIFI | 原  装 |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 下跌 |  | 库存 | 修理 | 刮痕 | 套餐 | 锁屏 | 屏幕 | CPU | 质  量 |  |
| 下降 |  | 速度 | 检测报告 | 划伤 | SIM | 通话 | 电池 | 数据 | 拆封 |  |
|  |  | 签收 | 保养 | 保护膜 | WCDMA | 打电  话 | 赠品 | 正版 | 封  条 |  |
|  |  | 交付 |  | 包装盒 | CDMA | 接电话 | 卡槽 | 容量 | 声音 |  |
|  |  | 订单  状态 |  | 外包装 | 号段 | 拨打  电话 | 外壳 | 安装软  件 | 音  量 |  |
|  |  |  |  | 塑料 | 神州 | 发信息 | 保护套 |  | 批次 |  |
|  |  |  |  | 金属 | 动感 地  带 | 扬声  器 | 保护  壳 |  | 发  热 |  |
|  |  |  |  | 玻璃 | 全球通 | 振动 | 摄像头 |  | 烫 |  |
|  |  |  |  | 产品包  装 | 换卡 | 震动 | USB |  | 寿  命 |  |
|  |  |  |  | 显示屏 | 大卡 | 彩信 | 套装 |  |  |  |
|  |  |  |  |  | 小卡 | 拍照 | 感应  器 |  |  |  |
|  |  |  |  |  | 单卡 | 关机 | 机盒 |  |  |  |
|  |  |  |  |  | 双卡 | 密码 | 说明书 |  |  |  |
|  |  |  |  |  | 双 卡 | 收音  机 | 连接  线 |  |  |  |
|  |  |  |  |  |  | 手电 | 音箱 |  |  |  |
|  |  |  |  |  |  | 播放 |  |  |  |  |
|  |  |  |  |  |  | 音乐 |  |  |  |  |
|  |  |  |  |  |  | 浏览  器 |  |  |  |  |
|  |  |  |  |  |  | 看电 |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | 视 |  |  |  |  |
|  |  |  |  |  |  | 通讯  录 |  |  |  |  |
|  |  |  |  |  |  | 连接  电脑 |  |  |  |  |
|  |  |  |  |  |  | 充电  时间 |  |  |  |  |
|  |  |  |  |  |  | 待机  时间 |  |  |  |  |
|  |  |  |  |  |  | 播放  歌曲 |  |  |  |  |
|  |  |  |  |  |  | 播放  音乐 |  |  |  |  |
|  |  |  |  |  |  | 待机 |  |  |  |  |
|  |  |  |  |  |  | 免提 |  |  |  |  |

致 **谢**

时间飞逝，博士四年的生活马上就要划上句号，经过将近两年的努力，终于完成了这篇博士论文。回想整个规划和写作的过程，多少个日子在对外经济贸易大学图书馆里努力查找图书和电子资料、阅读文献、确定论文选题、抓取数据、处理分析数据、敲击键盘、多次修改，这些过程中离不开导师陈进教授的指导，离不开同学和好友的帮助和支持，离不开家人的鼓励和帮助，我才能顺利完成论文，在这里向各位帮助过我的老师、同学、朋友和亲人表示深深谢意！

首先，要感谢我的导师陈进教授，此论文的顺利完成得益于我的指导教师陈进教授从论文选题、框架确定、研究方法选择到论文撰写、修改完成整个过程对我的悉心指导和支持鼓励。在整个博士学习四年时间里，陈进教授作为院长不但要处理学院的各种事情，而且非常尽职尽责的指导我的学习和生活，特别是在论文选题方向确定期间，陈老师给我提出了宝贵的建议和详细的指导，使我少走了很多弯路，能够准确把握选题和研究方向。在社会实践方面，陈进教授带着我们去了上海、舟ft等多个城市实地考察调研目前国内电子商务企业的生存环境和经营状况，让我在实际环境中真正体会到中国电子商务的发展状况，这对于我的选题非常有现实意义，帮助我拓宽了研究的视角。在论文写作中，陈进老师要求非常严格，对论文研究发展进度监督严格，经常督促我鼓励我严格按照时间安排完成论文不同阶段的各种研究工作。陈进老师不仅在学术方面有敏锐的观察力和独特的视角，他的严格的管理风格、正直的为人处世的态度、平易近人的性格、对学生善良仁慈的关怀、对困难学生的帮助都给我留下深深的印象，给我们树立了良好的榜样并潜移默化的影响我的思想和行为。

论文的顺利进行还要感谢信息学院李兵老师，他在我的论文完成过程中对于信息抓取工具的建议和使用提供了指导，感谢他让我参与每周的学生科研座谈会，让我学习到了很多论文写作方面的知识。

同时还要感谢华为技术有限公司的张祥森工程师，他在我论文所用数据样本选择以及数据抓取量化过程中给与了很大的帮助和指导。还要感谢我的同学荆逢春和薛蕊同学，在整个博士学习期间，我们一起学习课程、讨论交流，她们在我论文完成过程中所用技术的使用进行了长期的交流、支持和帮助。是他们的大力支持和帮助，才能使我顺利完成博士论文的研究与写作。

同时，也要感谢对外经济贸易大学给我提供了良好的科研和学习的环境，让我在学习和研究中得到足够多的研究资料得以学习知识，增长见识。

最后，我要深深感谢我的家人们，感谢我的父母对我学习的支持和鼓励，感谢老公对我的照顾和支持，只有你们的无私关爱和支持，才能让我全心的投入学习和科研中，并顺利完成博士论文。

再次感谢我的导师、亲人、同学和朋友的帮助、支持和鼓励，在今后的人生道路上我会承载着对你们的感激之情，克服各种困难继续奋斗，实现更高的人生目标，以此汇报给每一位关心、支持和帮助我的人！感谢你们！

# 个人简历在读期间发表的学术论文与研究成果

**个人简历：**

孙妮，女，1982年1月10日生。

2005年7月毕业于ft东财经大学（原ft东经济学院），获管理学士学位。

2007年7月毕业于对外经济贸易大学，获经济学硕士学位。

2008年3月至2010年8月：在海辉高科软件有限公司做数据分析工程师。

2010年9月进入对外经济贸易大学攻读产业经济学专业博士研究生。

**已发表的学术论文与研究成果：**

[1]孙妮、陈进. 信息服务产业贸易竞争力对比研究. 求索, 2012,11.

[2]陈进、孙妮、汪菲. 京沪服务贸易竞争力对比研究.商业时代, 2013, 9.

[3]李淑静、孙妮. 多哈回合农业议题谈判的进步之处.商场现代化, 2013,2.

[4]孙妮、陈进、汪菲. 基于B2C网站商品咨询和评论的网络消费者行为规律研究.图书情报工作, 2014, 增（1）.

**主要科研项目：**

1. 基于B2C电子商务网络购物用户商品评论研究，2013年对外经济贸易大学研究生科研创新项目（201315）, 2013，项目负责人。

2. 金融服务外包产业聚集的区位优势与发展模式研究，教育部人文社科基金项目(09YJA630023)，2009．9，参与者