

Chapter 3

Research Methodology

This research is qualitative research. The details about the research method as follows:

1. Research design
2. Population and sample size
3. Research instruments
4. Data collection
5. Data analysis

Research Design

1. Review the existing literature and research status both domestically and internationally.

This thesis summarizes existing literature from three aspects: the usefulness of online word-of-mouth, the impact of online word-of-mouth on consumer decision-making, and the impact of online word-of-mouth on product sales.

2. Build a multi-dimensional word-of-mouth model.

Based on existing literature research, this thesis identifies the independent variables that need to be included in the research model, learns the quantification methods of each variable, and ultimately constructs an research model and proposes research hypotheses.

3. Extracting word-of-mouth information.

Firstly, select appropriate data sources. Then, develop crawl in Python to get the data required for this thesis. Finally, filter and clean the data, get the sample data.

4. Empirical analysis to draw conclusions.

The sample data is input into the model for analysis using SPSS software. After statistical analysis, the proposed hypothesis in the previous section is verified to be true or false, and the final empirical conclusion is drawn.

5. Suggestions and Prospects

Based on empirical research conclusions, this thesis provides opinions on the product word-of-mouth management and consumer purchasing decision-making, and presents the future research directions.

Population and Sample Size

At present, many researchers have adopted Nelson's classification of products for research. Nelson divides products into experiential products and search products, and this classification method is also used for product types in this thesis. Consumers' evaluations and reactions to personalized recommendation systems vary depending on the type of product. Research has shown that compared to search based products, consumers have a higher evaluation of personalized recommendations for experiential products, so recommendations have a greater impact on consumers. The reason is that consumers tend to rely more on their own judgment when making decisions about search based products, while they tend to draw on the judgment from others when making decisions about experiential products. Therefore, this thesis mainly focuses on experiential products (facial essence in skin care products). First, skin care products have many comments on various websites meeting the needs of this thesis. Second, it is difficult for consumers to obtain quality information of experiential products before purchasing products. Consumers can only confirm whether the products are suitable for themselves after real use and get a comprehensive and detailed understanding of the product.

One of the research objects of this thesis is JD e-commercial platform. JD was equipped with an online comment system at the beginning of its establishment. With the development of the JD, the online comment system continues to upgrade and improve to meet business needs. In JD e-commercial platform, producers'

comments have specific review standards, comments that not meet the standards will not be displayed. Only approved comments can be displayed to other consumers. For example, the evaluation information text is not related to the product; Comments that copy more than 80% of the content from others' comments (based on the number of words). Compared with other online shopping platforms, the authenticity of comments in JD e-commercial platform is higher. Because if the purchased consumer post product comments in JD product comment system, they will be rewarded with Jingdou from JD e-commercial platform. This motivates consumer to comment on the purchased product, promotes interaction between consumers, and thus increases the number of visits to the platform and products.

Another research object is Xiaohongshu. Xiaohongshu is a social application software released in June 2013, targeting mainly young women. The platform initially started by sharing users' experiences in beauty, personal care, shopping, etc., and later added content sections such as tourism, home furnishings, hotels, restaurants, etc. It is a comprehensive platform of online communities, cross-border e-commerce, sharing platforms, and word-of-mouth databases. According to the latest data from Xiaohongshu, there are currently over 200 million monthly active users and over 43 million sharers. As a "Consumption Decision-making" platform in the hearts of many young users, the commercial value of Xiaohongshu cannot be underestimated, especially driven by the upgrading of national consumption, "Her Economy", and healthy living environment. Many brands need to expand the market and reach precise users through Xiaohongshu. Due to the fact that this research is conducted on two platforms simultaneously, the sample selection was mainly based on the existing indicators of both platforms as shown in the figure 3.1 and figure 3.2.



Figure 3.1 JD Mall Online Comments

Source: <https://item.jd.com/100013709921.html#comment> (10/06/2023)



Figure 3.2 Notes on Xiaohongshu Platform

Source: <http://xhslink.com/vBoalp> (10/06/2023)

In this thesis, sample data (that is, word-of-mouth data) we select covers JD product comments, Xiaohongshu note topics, notes and other information. we ensure that the sample data has a relatively wide coverage so that consumers can have comprehensive understanding facial essence product reviews and trends worldwide. By covering the data of domestic and foreign products, we can obtain a wider range of consumer comments to enhance the credibility and applicability of the research.

In addition, during the sample selection process, we pay special attention to facial essence products with different functions. Since the efficacy of facial essence products have their own characteristics due to the differences in ingredients and functions, such as basic moisturizing essence, whitening essence, anti-aging essence, etc. In order to delve into their impact on consumer word-of-mouth to sales, we select more than a dozen representative facial essence products. As for the sales data of the sample, in order to ensure the accuracy and consistency of the data, we will select data of JD.com's self-operated stores, because these data are usually subject to stricter supervision and quality assurance, thereby reducing data deviation and uncertainty.

In general, by selecting sample data covering domestic and foreign products with different functions, combined with the sales evaluation data of JD.com's self-operated stores, this thesis will obtain more extensive and in-depth research results, which will help to understand facial essence products more comprehensively. At the same time, strict checks on data sources to ensure the quality and credibility of the data will further enhance the persuasiveness of the research conclusions.

Research Instruments

This thesis will adopt a quantitative research method. Because the size of samples data in this thesis is large, it is difficult to collect data directly. Therefore, we develop a data crawler program in Python to capture data related to products comments from JD and Xiaohongshu, and then save the collected data to a local folder. We develop a data processing program in Python to process and gather statistics on the collected data, and then use SPSS to perform statistical analysis.

Data Collection

1. Data collection process

Obtaining product comments data on multiple platforms is the basic of thesis. Therefore, we develop crawler program as follow steps:

1.1 Target determination: Determine the e-commerce website to be crawled and the type of data to be collected, such as product information, prices, comments, etc.

1.2 Web page analysis: Analyze the structure of the target web page to determine the location and acquisition method of the required data. You can use your browser's developer tools to view the HTML structure and CSS selectors of a web page.

1.3 Choose a suitable crawler library: There are many excellent crawler libraries in Python, such as Requests, BeautifulSoup, Scrapy, etc. Choose the appropriate library based on the complexity of the task and your own preferences.

1.4 Write a crawler program: Use the selected crawler library to write a crawler program, visit the target web page and extract the required data. The general process includes sending HTTP requests, parsing web page content, extracting data, etc.

1.5 Dealing with anti-crawling mechanism: In order to prevent being blocked by the website, some anti-crawling mechanisms can be added, such as setting request headers, using proxy IP, etc.

1.6 Store data: Store the crawled data in local files or databases. Common options include CSV, JSON, SQLite, MySQL, etc.

2. E-commerce platform data acquisition

JD provides a browser webpage access address <https://www.jd.com/>, and also has a mobile phone APP. You can directly access every product and product comments information in the mall through the URL of the JD shopping website, and you can directly use the crawler program to crawl the product review information on the website more simply and directly. The steps to obtain product comments data in JD mall are as follows:

2.1 Determination of the skin care product webpage: In the JD shopping website, each product has a unique number, and the corresponding website of the product is: <https://item.jd.com/product number.html>. After determining the products to be crawled, you can find the websites corresponding to the self-operated products of JD.com.

2.2 Acquisition of product comments information: Use crawler program to obtain the corresponding webpage content of the product, analyze the webpage content, and obtain the information corresponding to each comment of the product: comment user ID, number of ratings, specific review content, number of review pictures, number of review videos, and review time.

2.3 Saving the obtained information: On the basis of obtaining the comment information, use the file saving function provided by the pandas library to save the obtained information into a text file or a CSV file. Sample data crawled from JD e-commerce platform is shown as table 3.1.

Table 3.1 Sample data from JD

Products	Comment user	Star rating	Comment content	Comment time	Image num
Proya : Ruby Essence	彬***句	5	早 c 晚 a 套装效果不错，用了两年了，而且物流特别快	2023-06-28	1
Clinique: Dark Spot Double Action Serum	M***1	5	除了正装还送了好几瓶赠品，现在还在用赠品呢！效果也不错，脸上的斑确实又减弱，平时也要注意防晒，效果更好	2023-06-28	3
Lancome: Little Black Bottle	u***m	5	物流很快，用起来很不错，吸收快，不油腻，我是混油皮，冬天干夏天油，现在夏天用着也很好，保湿效果都很好，用完感觉皮肤稳定，不容易过敏，**修复很不错！	2023-06-28	2

3. Social media platform data acquisition

Xiaohongshu only provides users with an APP mobile phone access client, and cannot directly provide browser webpages to access the website, so access to

Xiaohongshu can only be accessed through the mobile client. The Xiaohongshu APP client provides a content search function. You can use the content search function to enter the product name information to directly search for the subject note information list of the product. Click each item in the notebook information list to get specific information. Therefore, to obtain the note information about the product in the Xiaohongshu APP, you need to write crawler information to obtain each item in the subject note information list about the product on the mobile client. The specific steps for obtaining product review information in Xiaohongshu APP are as follows:

3.1 Obtain the interface information of the Xiaohongshu APP client: Connect the mobile phone with the developer mode to computer, and use the AirtestIDE tool to obtain the information of the mobile phone interface.

3.2 Acquisition of product note information: Develop a Python script program to call the search function of the Xiaohongshu APP client to obtain a list of product theme notes, automatically click on each list information, obtain the page corresponding to each list information, analyze each page, and obtain information on the topic of the notebook in the web page, the specific content of the note, the number of basic likes, the number of notebook favorites, and the number of comments on the note.

3.3 Save comment information: On the basis of obtaining comment information, use the file save function provided by the pandas library to save the obtained information into a text file or CSV file. Samples data crawled from Xiaohongshu social media platform is shown as table 3.2.

Table 3.2 Sample data from XHS

Products	comment user	note theme	note content	comment time
Proya : Ruby Essence	瞎*享	相见恨晚 肤感绝了 适合夏日 的精华	听说了太多翻车的，我又是敏感肌肤，所以买了小样，隔挺久用一次。珀莱雅红宝石精华最近清理小样中样，看着这个用过了，拿出来赶紧消耗掉。隔一天一用，肤感也太好了.....	2023-06- 26
Olay: light spot small white bottle	Pea** yee!	Olay 淡斑 小白瓶测 评	用的第二瓶了总结一下 439RMB60ml 不便宜我觉得而且想要快速变白这瓶还是算了两个月类做好防晒有可见效果我用完一瓶才感觉变白了些优点是温和不是很贵肤感很好用的时候一定要用量足不然真的没效果	2023-07- 01
SK- ll:Small light bulb	紫**辑	淡斑精华 黑名单关于祛斑的一些大实话	精华推荐平价精华祛斑精华姐妹们信我祛斑五年再没点经验真说不过去了辛苦整理了一些淡斑精华清单都是大实话有斑的姐妹参考一下 1CJ 淡斑精华液斑点克星淡斑提亮，性价比真的没话说而且质地好吸收，不挑肤质 2SKII 小灯泡精华想白里透红的就选它！让你气色好的像二十出头！	2023-07- 01

4. Data acquisition

This research uses web crawler technology to obtain e-commerce website data and third-party platform Xiaohongshu comment data, and the crawler program is mainly developed in Python language. This thesis selects 14 skin care products and obtains the relevant data of daily comments on the self-operated platform of JD.com and Xiaohongshu theme notes. The specific time of the samples is from June 26, 2023 - July 25, 2023. For the sake of research rigor and data reliability, daily product sales and online word-of-mouth data are collected for facial essence

products of each brand. The final structured data content is: review ID, product name, total number of reviews, and number of image reviews , star rating, comment date, note number. After the original data is captured, the data is cleaned and invalid samples are eliminated. In the end, a total of 20,578 samples of valid review data were obtained, including 10,433 samples from JD.com and 10,145 samples from Xiaohongshu.

Table 3.3 Data samples

Skin Care Products	JD Sample Size	XHS Sample Size	Total Sample Size
HR: Green Aquarius Essence	159	631	790
SHISEIDO: Red Kidney Essence	209	1130	1339
SK-II:Small light bulb	1852	269	2121
SkinCeuticals: Color Repair Essence	171	1433	1604
Clinique: Dark Spot Double Action Serum	128	212	340
Lancome: Little Black Bottle	647	420	1067
Kefumei: Collagen stick Human-like recombinant collagen muscle repair essence	959	534	1493
Guerlain: Imperial Honey Repair Gold Restoration Honey Essence	185	219	404
Clarins: Double Essence	339	1698	2037

Olay: Anti-sugar small white bottle	1315	754	2069
Olay: light spot small white bottle	1505	773	2278
Proya : Double Anti-Essence	1435	1195	2630

Data Analysis

1. Variable design

1.1 Dependent variable

For merchants, product sales directly affect store operations and are an important factor in measuring product performance and quality. Product sales are core data for enterprises and must be kept confidential. However, in the process of research, in order to replace sales, scholars often use the number of reviews or sales ranking, both of which have been recognized in the academic circle. However, the actual situation of JD Mall data obtained in this thesis shows that the sales ranking is reordered every week, which is not suitable for analysis instead of sales. Therefore, this thesis chooses to use the number of product online reviews as an alternative indicator of sales.

1.2 Independent variable

The independent variables in this study mainly include the number of comments, star rating, and images from JD e-commerce platform related products, as well as the number of comments on Xiaohongshu social media platforms. The specific meanings of all variables in this study are shown in table 3.4.

Table 3.4 Variable name and description

Variable Category	Variable Name	Variable Description	
Dependent Variable	Sales volume of products	Actual sales volume of JD platform products	
Independent Variable	Num of JD's Comments	Total number of JD platform product reviews	
	E-commerce platform	Num of JD's Image	Total number of JD platform product reviews with images
		Star rating of JD	JD platform star rating
	Social media platforms	Num of XHS's Comments	The total number of product-related notes released by Xiaohongshu platform

2. Descriptive statistical analysis

In this thesis, the statistical software SPSS is used to summarize and describe the main features of valid sample data, such as its central tendency, variability, and distribution.

3. Correlation analysis

Correlation analysis of data is relatively common in statistics. Correlation analysis can not only judge whether there is correlation of data, but also describe the degree and direction of correlation of data. Before regression analysis of variables, Pearson correlation analysis matrix can be used to preliminarily test whether there is correlation among independent variables and between independent variables and dependent variables. We can analyze which variables are significantly related to sales volume from the correlation analysis results.

4. Hypothesis test

4.1 Regression model construction

Through sorting out the relevant literature, scholars generally use regression analysis methods for empirical analysis to study the impact of product comments on sales, and compare the importance of different coefficients in the regression results to obtain the analysis objectives. The regression equation is established according to the influence of different dimension information on sales volume. The expression is as follows:

$$sales = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b$$

Among them: sales represent product sales, w is the index weight, x is the value of each independent variable, and b is a offset constant.

4.2 Hypothesis test results

Summarize and sort out the hypothesis test results of the product sales volume from the 4 subdividing dimensions of product reviews.