

A Study on the Perception of Yongfu Siraitia Grosvenorii Consumption Based on Web Text Analysis

Qiongqian Qin^{1,a,*}, Shanxi Chen¹, Rimei Zhong²

¹School of Economics and Management, Guangxi Normal University, Guilin, China

²School of Mathematics and Statistics, Guangxi Normal University, Guilin, China

^aqqx12314@stu.gxnu.edu.cn

*Corresponding author

Abstract: With the continuous development and prosperity of the online shopping industry, the use of the Internet to conduct research related to shopping has become a new form of research. Post-purchase consumer reviews are the most realistic and personalised way of perceiving Siraitia grosvenorii, and they are also the most direct indicator of consumers' perceptions and preferences of Siraitia grosvenorii products. This study uses online text analysis to compare and contrast the consumption preferences of consumers of Siraitia grosvenorii products with the comments of consumers who bought Siraitia grosvenorii products on Taobao, and conducts high-frequency word analysis, semantic network analysis and emotional perception analysis. Suggestions were made to improve the consumer perception of Yongfu Siraitia grosvenorii.

Keywords: Yongfu Siraitia grosvenorii; Consumer perception; Web text analysis

1. Introduction

Siraitia Grosvenorii is the fruit of a perennial Temperate plant of the Cucurbitaceae, with oval leaves, yellowish flowers and rounded fruits, used in medicine to clear heat and relieve cough. Siraitia grosvenorii originated in Yongfu County and is a unique fruit in China with a high economic value. Guangxi is the largest production base of Siraitia grosvenorii in China, and Yongfu in particular has the highest quality. After years of development, Yongfu Siraitia grosvenorii industry has formed a rather large scale characteristic industry. The industry formed by Siraitia grosvenorii integrates agricultural production, processing, sales and rural tourism as well as cultural heritage, which has a strong competitiveness in Guangxi and even the whole country, and has become the main force of the county characteristic industry in Yongfu County. However, in terms of Siraitia grosvenorii market consumption, there are problems of large differences in the degree of consumer perception and insufficient deepening of consumption scenarios. Therefore, this paper deconstructs consumers' consumption preferences and perceptions of Siraitia grosvenorii products based on big data analysis of high-frequency words, semantic network analysis and sentiment analysis to conduct an in-depth study on the industrial development of Yongfu Siraitia grosvenorii and countermeasures to effectively activate the market consumption potential and extend the value chain of Siraitia grosvenorii products.

2. Literature review

In terms of research on the cultivation of Siraitia grosvenorii varieties and technologies such as sweet glucoside extraction, Qin Hongbo et al. explored key techniques for the high-yield cultivation of Siraitia grosvenorii histoponic seedlings by controlling the effects of different planting and pollination times, transplanting methods and other variables on the growth of Siraitia grosvenorii histoponic seedlings [1]. Dai Sheng et al. were interested in exploring the problems faced in the application of Siraitia grosvenorii sweet glucoside in the field of health products and pharmaceuticals to provide a reference for the application of Siraitia grosvenorii sweet glucoside as a natural sweetener [2]. In terms of the industry production, processing and sales of Siraitia grosvenorii, Liao Xiansheng et al. analyzed the "Internet +" marketing model of Siraitia grosvenorii and proposed suggestions for the rational allocation of marketing resources, optimization and upgrading, and product extension of Siraitia grosvenorii [3]. Lu Yan et al. focused on the production and processing of Siraitia grosvenorii and

proposed to strengthen the link between the interests of farmers, enterprises and the government [4]. Lei Xian Yue summarized the development status of the arrowroot industry [5]. Liang Wei found the problems of backward production mode and insufficient innovation in product development and marketing model of *Siraitia grosvenorii* [6]. At present, domestic and international research on *Siraitia grosvenorii* has focused on the cultivation, pharmacology or food industry of *Siraitia grosvenorii*, but few scholars have conducted research on the consumer perception of *Siraitia grosvenorii* using online texts. Therefore, on the basis of previous studies, with the help of ROSTCM6 software, web text analysis and Octopus crawler program, this study takes the comments of consumers who buy *Siraitia grosvenorii* products on Taobao as samples to conduct high-frequency word analysis, semantic network analysis and emotional perception analysis, so as to make a comparative study on the consumption preferences of *Siraitia grosvenorii* consumer groups.

3. Research design

3.1. Overview of the *Siraitia grosvenorii* industry in Yongfu County, Guangxi

Yongfu County is famous as the "hometown of *Siraitia grosvenorii*" and is also the largest production and export base of *Siraitia grosvenorii* in China. Yongfu County has 6 towns and 3 townships under its jurisdiction, of which Longjiang Township, Baishou Township, Suqiao Township and Yongfu Township are the main original production areas of Yongfu *Siraitia grosvenorii*. The local meteorological conditions such as light, temperature and moisture are very suitable for the growth of *Siraitia grosvenorii* which likes shade and does not tolerate high temperatures and sunlight. The selenium-rich, high-quality and tasty *Siraitia grosvenorii* produced here is the finest of fruits, and seedlings of varieties of green-skinned fruit, long-spreading fruit and red-haired fruit are more likely to survive, produce more fruit and have larger fruits, and are widely planted. With a planting area of nearly 120,000 mu and a production and output value that ranks first in the country, the total output value of the whole industry chain reaches more than 10 billion yuan, and the annual export volume accounts for 85% of the global market share [7], and the raw fruit of *Siraitia grosvenorii* and its products are exported to more than 20 countries and regions, including Japan, the United States and South Korea. By 2022, Yongfu County had more than 400 cooperatives for acquisition, processing and sales, and more than 20 enterprises for scientific research and finishing, forming a more complete industrial system.

3.2. Data collection and processing

3.2.1. Acquisition of data

With the help of the micro-word cloud platform and ROSTCM6 software, this study used web-based text analysis to study the consumer perceptions and consumption preferences of *Siraitia grosvenorii* consumers by using post-sale reviews of shops selling *Siraitia grosvenorii* products on the Taobao platform as a sample. A total of 1711 reviews were crawled through filtering using the Octopus crawler program. The basic fields crawled included: username, review time, review content, and other information, as shown in Table 1 below.

Table 1: Crawled web review information

Username	Comment Time	Comment content	Other information
I**2	19 August 2022	Very good value for money, this is the second order back	Viewed 36 times Pointed useful
I**6	17 Mar 2023	The fruit is medium sized, full of fresh fruit and individual	Total number of comments 72 clicked useful 0
I**'s	15 Mar 2023	A lot of portions, the fruit is very big stuff received, very full	Total number of comments 63 clicked useful 1
with** end	13 Feb 2023	I received the baby and can't wait to open it	Total number of comments 56 clicked useful 1
t**2	07 October 2022	A big bag full, sufficient quantity, no crushed, taste	Total number of comments 201 clicked useful3

3.2.2. Comment text pre-processing

Because the content required for crawling belongs to the original data, which contains a lot of worthless information, it is necessary to clean the data before analysis. If the original data is studied

directly, the subsequent modeling process will be disturbed and the model effect will be affected. Generally speaking, there will be water army, repeated comments, meaningless sentences and other cases to make up the number. Through the process of data cleaning, these meaningless statements for analysis can be removed. For the consumer review data of *Siraitia grosvenorii* on websites, OTA platforms and merchants may employ mercenaries to artificially give positive reviews or give negative reviews to competitors in order to improve their own competitiveness by increasing popularity and attracting customers. This phenomenon will lead to meaningless reviews such as "default good reviews" and "default bad reviews". We try to identify and delete misleading information provided by the navy. At the same time, not all customers are willing to take the time to review, so there may be data where users only rate but don't describe it in words. In order to save time, some customers are likely to copy the comments of others. Therefore, we need to delete the data with no comments and only ratings, and de-duplicate the comments.

Comments on the Internet are not limited to the number of words, some customers will only comment a few words, such as "good, good", "really good", these short comments can only reflect the positive emotional color, but for the study of the consumption preference of *Siraitia grosvenorii* cannot play any role, belong to invalid information, should be deleted. Based on the above purposes, the data is cleaned in the following ways.

Delete the data of default praise, default medium comment, default bad comment or user without comment content. The specific number of deleted comments is as Table 2.

Table 2: Deleted system default comment data

Deleted content	Number of comments
Default praise	263
Default medium	3
Default bad	36
Users did not post comments	72

If the comments made by consumers are too few words, it will be regarded as invalid information. We assume that the string length is 10 characters, and the invalid comments with less than 10 characters will be removed.

In summary, a total of 1362 reviews were obtained by eliminating the system default ratings, duplicate data and invalid reviews, with an efficiency rate of 79.60%. After word segmentation and screening of comments by ROSTCM6 software, high-frequency words are accurately extracted, and word frequency analysis and semantic network analysis are conducted on the results.

4. Interpretation of result

4.1. High-frequency word analysis

The sample of web comments is imported into ROSTCM6 software, and the word frequency analysis is carried out on the text after word segmentation. Table 3 shows the top 60 high-frequency words with the highest frequency of consumer reviews.

Through the above analysis, we found that the word "packaging" is the most mentioned by consumers, which indicates that consumers pay the most attention to the external packaging of the *Siraitia grosvenorii*, followed by "Siraitia grosvenorii", "taste", "quality", "freshness", indicating that consumers pay attention to the quality of the *Siraitia grosvenorii* taste and freshness. "size" and "weight" indicate that consumers attach importance to the size of the *Siraitia grosvenorii*; "Clean", "affordable", "health", "shelf life" further shows consumers' concern about the value for money of *Siraitia grosvenorii*; Words such as "cough" and "throat" show consumers' demand for the throat wetting and cough relieving effect of *Siraitia grosvenorii*; The words "acquisition", "logistics", "delivery" and "very soon" indicate that consumers care more about the delivery time and the speed of logistics transportation when they buy the *Siraitia grosvenorii* through the network. "Satisfaction", "repurchase" and "comfort" reflect consumers' pleasant experience and express consumers' willingness to buy again.

4.3. Analysis of emotional perception

Sentiment analysis, as one of the most commonly used research methods, is able to count, collate and analyse the subjective words with emotion in the text of online reviews, and finally come up with three kinds of analysis results classified as positive, neutral and negative sentiment, sentiment analysis can reflect consumers' perception of *Siraitia grosvenorii* products more intuitively, in order to better understand consumers' perception of *Siraitia grosvenorii*, using the sentiment analysis function in ROST-CM6 to analyse the text of online reviews related to *Siraitia grosvenorii* consumers, the results of sentiment analysis are shown in the following table 4.

Table 4: Results of the emotional analysis of consumers

Analytic result	Internet Consumer Reviews	
	Number of texts / bars	proportion /%
Positive mood	1043	76.58%
Neutral mood	138	10.13%
Negative emotions	181	13.29%

In the results of the sentiment analysis, the highest percentage of positive emotions was 76.58%; followed by negative emotions with 13.29%; and lastly, neutral emotions with 10.13%. In the segmentation statistics of positive sentiment, general positive sentiment accounted for 37.81%, moderate positive sentiment accounted for 26.07% and high positive sentiment accounted for 12.70%, with general positive sentiment accounting for the highest percentage, followed by moderate positive sentiment. On the one hand, this indicates that consumers in the *Siraitia grosvenorii* market have a relatively good feeling about it, but on the other hand, it indicates that there is still some room for improvement in the quality of *Siraitia grosvenorii*. Further analysis of the online review texts shows that positive sentiment is mainly reflected in size, freshness, packaging, and cough suppressant effect. In the segmentation statistics of negative emotions, 9.62% of the respondents were generally negative, 3.16% were moderately negative and 0.51% were highly negative. The analysis of the online review texts shows that the negative sentiment is mainly reflected in the aspects of "broken packaging, poor taste, partially broken fruit, small fruit", etc. Some consumers also reported that "over-dried fruit is burnt, unpleasant taste, does not match the photo, very poor quality and different sizes". This also reflects the lack of quality control, packaging and details of *Siraitia grosvenorii*.

5. Conclusion and suggestion

5.1. Conclusion

In this paper, Yongfu *Siraitia grosvenorii* is taken as the research object, and ROST-CM6 software is used to conduct high-frequency word analysis, emotion analysis and semantic network analysis on the online comment text that captures consumers' consumption perception of *Siraitia grosvenorii* products on Taobao shopping platform. It is believed that consumers pay high attention to the external packaging and taste of *Siraitia grosvenorii* products, which are reflected by high-frequency words such as "packaging", "taste" and other high-frequency words; "good", "sweet" and other high-frequency words to illustrate consumers' taste buds' feelings when consuming *Siraitia grosvenorii*; High-frequency words such as "throat", "cough" indicate that consumers buy *Siraitia grosvenorii* out of the need to moisten their throats and relieve cough. From the semantic network analysis, "packaging", "taste" and "texture" are in the core position, which further verifies that consumers will consider the external packaging design and the taste quality of the *Siraitia grosvenorii* products at the same time when buying *Siraitia grosvenorii* products. The sentiment analysis shows that consumers' overall perception and experience of *Siraitia grosvenorii* products are good, but there is also negative sentiment. Negative sentiment comes from three main sources:

First, *Siraitia grosvenorii* do not pay attention to packaging design. On the one hand, there is a homogenization of packaging, simplification, some directly sold in bulk, resulting in susceptible to moisture; On the other hand, there is the phenomenon of "excessive packaging" and "luxury packaging". Due to the high added value, it inhibits the consumption desire and affects the sales volume.

Secondly, the quality of *Siraitia grosvenorii* products on the market is different. Some *Siraitia grosvenorii* is not fresh, the flesh is moldy and black, the water from the bubble tastes bitter, which affects the taste of *Siraitia grosvenorii* products and reduces the sales volume of *Siraitia grosvenorii* products.

Thirdly, the logistics system is not perfect, consumers online shopping *Siraitia grosvenorii*, merchants delivery time is late, logistics speed is slow, received goods damaged.

5.2. Suggestions

5.2.1. Improve the packaging design of *Siraitia grosvenorii*

As a native product of Guangxi, the packaging design of *Siraitia grosvenorii* should first reflect the regional culture, improve consumers' intuitive feeling of the product and regional emotional experience, so as to realize the spread of culture and product sales. Secondly, the implementation of different brand design, highlight the brand characteristics, expand the *Siraitia grosvenorii* product publicity channels and brand awareness. Finally, to "green packaging" and "zero packaging" change. Reasonable grasp of different customer groups to different packaging design needs, to achieve packaging perfection and packaging cost balance.

5.2.2. Grasp the quality control of *Siraitia grosvenorii* products

On the one hand, we should control the tracking of *Siraitia grosvenorii* from picking to processing, product detection to storage, and sales quality, and timely recycle unqualified products. On the other hand, compared with the traditional charcoal roasting method, the low-temperature dehydration process can reduce the generation of broken fruit, maximize the nutritional value of the fruit, better color, and better improve the quality of the *Siraitia grosvenorii* products.

5.2.3. Create a modern storage and logistics system

First of all, the establishment of a modern logistics system is an important way to improve the quality and efficiency of the logistics industry, which also contributes to the free flow of products and resource elements across regions and industries, so as to optimize the allocation efficiency. Secondly, consider planning and building modern *Siraitia grosvenorii* cold chain storage center, *Siraitia grosvenorii* logistics center, distribution center, to ensure that fresh *Siraitia grosvenorii* can be kept fresh for a long time after picking, ready for processing. Finally, make full use of the advantages of local highways and railways to create a multi-linkage logistics system.

Acknowledgements

Thanks to Shan Peng, an undergraduate from the School of Economics and Management of Guangxi Normal University, for her participation in the review text analysis.

References

- [1] Qin Hongbo, Wang Xingui, Guo Lunfa, Jiang Xineng, Wei Wanqiang, Wu Lingqiong, Tao Xinglai. *Research on high-yield cultivation technology of Siraitia grosvenorii histoponic seedlings [J]. Fujian Journal of Agriculture*, 2019, 34(02):198-203.
- [2] Dai S., Wang H. L. *Research progress on the natural sweetener Siraitia grosvenorii sweet glucoside [J]. Chinese Patent Medicine*, 2023, 45(02):503-509.
- [3] Liao Xiansheng, Man Gaopeng, Wei Yanxi, Huang Shiyang, Xie Ziyang. *Research on the "Internet+" marketing model of Siraitia grosvenorii—taking Guilin Yongfu County Siraitia grosvenorii as an example [J]. Mall Modernization*, 2021(02):63-65.
- [4] Lu Yan, Qin Shengyun, Hu Tingting, Wei Zhongxun. *Building a ten billion dollar "sweet industry" in Guangxi [J]. Enterprise Science and Technology and Development*, 2022(02):25-27.
- [5] Lei Xiangyue. *Research on the development status and countermeasures of Siraitia grosvenorii industry in Yongfu County [D]. Guangxi University*, 2019.
- [6] Liang Wei. *Research on the development of rural special industries in Guangxi under the strategy of rural revitalization - taking Siraitia grosvenorii industry in Yongfu County, Guilin as an example [J]. Modern Marketing (Lower Journal)* 2022(08):129-131.
- [7] Yang Jiahang. *A study on the strategy of live cross-border e-commerce for Chinese special agricultural products based on RCEP--Guangxi Siraitia grosvenorii as an example [J]. Today Media*, 2022, 30(12):133-136.